ee on
House of

September 1987

# HEALTH RISK ANALYSIS

y in
es

GAO

**United States**
**General Accounting Office**
**Washington, D.C. 20548**

**Program Evaluation and**
**Methodology Division**

B-227612

September 30, 1987

The Honorable Robert A. Roe
Chairman, Committee on Science,
Space, and Technology
House of Representatives

Dear Mr. Chairman:

In a December 4, 1985, letter, your committee asked GAO to investigate the quality and scope of risk analysis activities conducted by selected federal agencies with responsibility for regulating environmental health and safety. This report is on work carried out in response to that request. It serves as a pilot study for possible future evaluations of federal health risk analyses.

Copies of this report will be sent to the secretary of Health and Human Services, the secretary of Labor, and the administrator of the Environmental Protection Agency and will be made available to other persons who request them.

Sincerely,

Eleanor Chelimsky
Director

# Executive Summary

## Purpose

Many regulatory programs in the federal government are designed to protect public health. They include efforts to ensure the safety of the food supply, eliminate unnecessary hazards from the workplace, and prevent environmental degradation. Many federal agencies have adopted risk analysis methods to strengthen their decisions on whether and how to regulate health hazards. Their risk analyses have a major effect on regulatory decisions regarding health. But little is known about the extent to which sound methodological practices are being implemented.

The House Committee on Science, Space, and Technology asked GAO to investigate the quality of the risk analysis activities of selected federal agencies that have responsibility for regulating health and safety. This report provides preliminary information on possible weaknesses and areas of strength in federal health risk analyses. In three pilot case studies, GAO evaluated the risk analysis work that supported regulatory actions at three federal agencies, each of which employed a different approach to risk management.

## Background

Risk analysis is the process of examining information concerning the level of risk posed by a hazard, the acceptability of that level, and possible actions to reduce the risk, if necessary. Risk analysis includes the elements of risk assessment, in which health effects from exposure to hazardous materials are defined, and risk management, in which policy alternatives are weighed and integrated with the results of risk assessment, in order to reach a decision on the most appropriate regulatory action. Risk management also includes monitoring and evaluation to determine if the regulatory action is achieving its intended effects in reducing risk.

At the Food and Drug Administration (FDA), GAO evaluated the risk assessment and risk management work that supported a proposed action on methylene chloride, which is used for decaffeinating coffee and as a flame retardant in hair spray and is regulated under the Federal Food, Drug, and Cosmetic Act. At the Occupational Safety and Health Administration (OSHA), GAO evaluated the risk assessment and risk management work supporting an action on inorganic arsenic, which is emitted into the air as a by-product of smelting activities and regulated under the Occupational Safety and Health Act of 1970. Finally, at the Environmental Protection Agency (EPA), GAO examined the risk management work supporting an action on volatile organic compounds, which are emitted from onshore natural gas-processing plants and are

regulated under the Clean Air Act; risk assessment work was not evaluated because none was conducted specifically for developing this standard.

GAO evaluated the extent to which accepted scientific and administrative practices were employed; GAO used criteria it developed from a search of the risk analysis literature. The final regulatory decision itself was not evaluated, however.

The main strength of this study is the comprehensiveness of the examination of each case. The study's main limitation, which derives from its being a pilot, is the small number of cases. Because GAO examined only one case at each agency, it is not possible to generalize about the quality of the risk analysis work conducted by the agencies or the federal risk analysis process as a whole.

## Results in Brief

The quality of the risk analysis work varied, fewer problems occurring in the risk assessment phases than in the risk management phases.

The risk assessment work was generally adequate. FDA and OSHA were able to meet basic expectations for sound scientific performance in identifying and estimating the risks posed by specific hazards. However, the use of unverified assumptions to fill largely unavoidable gaps in the data resulted in great uncertainty about the magnitude of the risks.

The risk management work in all three cases exhibited serious problems. The integration of policy options and risk assessment results was poorly documented, and the basis for regulatory decisions was unclear. Follow-up evaluations of the regulatory actions were not generally performed, largely because of cost or technical infeasibility. This means that the agencies cannot determine the risk reduction, if any, that is achieved by regulation.

## Principal Findings

### Risk Assessment

The risk assessment work GAO evaluated generally met the technical and scientific criteria developed from the literature. The identification, characterization, and estimation of risk were well conducted by FDA and

OSHA. They did a credible job of reviewing and evaluating the available evidence on a hazard. (See pages 22-31 and 44-51.)

GAO did find some problems with risk assessment that were primarily related to the availability of data. In some situations, gathering information was beyond current technical abilities; in others, resource constraints limited the agencies' abilities to acquire the data. These problems resulted in the agencies' using assumptions when data were not available, introducing some unknown level of uncertainty in all three cases. Other problems in the risk assessment work included the absence of systems for prioritizing potential hazards for analysis and the absence of formal risk assessment guidelines at FDA. (See pages 22-31, 44-51, and 68-69.)

## Risk Management

GAO found significant problems in the risk management work. In the FDA and EPA cases, the development and evaluation of risk management options and the actual decisionmaking process were poorly documented, so that it was difficult to determine the information the regulatory decisions had been based on. It was not always clear how regulatory options had been developed, how full the consideration of options had been, or how uncertainties associated with the options had been presented to decisionmakers. In the EPA case GAO examined, and possibly in the FDA case, the precision of estimates may not have been sufficient for choosing a final regulatory option from alternatives. (See pages 31-35, 51-55, and 65-77.)

The extent and quality of the guidelines for risk management varied greatly, between and within the agencies. EPA had extensive guidelines for the development and evaluation of regulatory options, but it had no guidelines for decisionmaking. In contrast, OSHA had strong guidelines for decisionmaking but no guidelines for developing and evaluating risk management options. FDA had no guidelines for either options development or decisionmaking. (See pages 32-34, 52-54, 66, and 73-76.)

None of the agencies had conducted follow-up evaluations of the regulations to determine if they were achieving the intended effects in reducing risk. An evaluation effort has been planned at EPA; an evaluation is not technically feasible for the type of case examined at FDA; and cost and other considerations prevent OSHA from planning such an effort. Thus, the efforts devoted to risk analysis on these three cases have resulted in regulations that are only presumed to be reducing adverse

effects on health. No knowledge of the actual health effects of these regulations is currently available. (See pages 35-36, 55, and 78.)

## Recommendations

GAO is making no recommendations because this report covers a pilot effort in which too few risk analysis cases were examined to permit generalization.

## Agency Comments

The Department of Health and Human Services and the Department of Labor indicated general support for GAO's effort to evaluate the risk analysis process and help improve regulatory decisionmaking. However, all three agencies expressed a number of general concerns. Some comments are critical of the cases GAO selected and suggest that the report holds the agencies accountable for the quality of scientific research that is beyond their control. Other comments assert that the report confuses the quality of the scientific work with the administrative process and that GAO's criteria are not universally applicable or valid.

After carefully reviewing these issues, GAO believes these criticisms reflect a lack of understanding of GAO's intent in performing this study. Further, prior to the start of the evaluation, each agency reviewed the list from which its case was selected and was asked to raise objections to cases that it thought were inappropriate. No objections to the cases were raised. GAO agrees that no single case can represent an entire agency's risk analysis efforts, and the report does not make such representations. Moreover, while GAO does point out flaws in the scientific research the agencies relied on in their risk analyses, GAO does not hold the agencies accountable for these flaws.

GAO disagrees that the report confuses the administrative process with the quality of the scientific work; on the contrary, the report purposely links the two dimensions. Process questions are important because without them, the quality of the research remains unknown; the documentary record is the prescribed means for demonstrating the quality of the research. Finally, GAO does not argue that the criteria it developed are universally applicable or that they exhaust the characteristics required for adequacy. Rather, they are appropriate for the cases examined and for other similar cases.

These agencies' comments are discussed in detail in chapter 5 and appendixes VI-VIII.

# Contents

## Tables

## Figures



### Abbreviations

| | |
|---|---|
| CFSAN | Center for Food Safety and Applied Nutrition |
| DOL | Department of Labor |
| EPA | Environmental Protection Agency |
| FDA | Food and Drug Administration |
| GAO | General Accounting Office |
| HHS | Department of Health and Human Services |
| NAS | National Academy of Sciences |
| OMB | Office of Management and Budget |
| OSHA | Occupational Safety and Health Administration |

# Introduction

One of the major responsibilities of the federal government is to protect public health. Many regulatory programs have been put in place to provide this protection. Assuring the safety of the food supply, eliminating unnecessary hazards from the workplace, and preventing environmental degradation are just a few of the objectives of these efforts. To assist in fulfilling this responsibility, many federal agencies have adopted and, in some cases, developed analytical methods that strengthen the basis for regulatory decisions. The use of these methods and the decisionmaking they support are referred to as the "risk analysis process."

Risk analysis is the process of examining information concerning the level of risk posed by a hazard source, the acceptability of that risk level, and possible actions to reduce the risk, if necessary. One particular area that has received considerable attention over the last decade is the use of risk analysis to develop policies to protect the public from health hazards. Three agencies with principal responsibilities for managing such hazards are the Food and Drug Administration (FDA), the Occupational Safety and Health Administration (OSHA), and the Environmental Protection Agency (EPA). Each of these agencies uses some form of risk analysis to reach decisions concerning whether a risk is acceptable and, if not, how best to reduce it. Millions of dollars are spent each year in implementing regulatory decisions to provide protection from health risks, and it is therefore very important to evaluate how well this work is conducted.

The House Committee on Science, Space, and Technology asked us to examine the quality and scope of federal risk analysis work. The request was initiated because the extent to which sound methodological practices are being implemented is largely unknown. This report presents our work in response to the congressional request, a pilot examination of three risk analysis cases, one each at FDA, OSHA, and EPA. We will use this initial work for determining the viability of the research approach and identifying the most important issues for a possible follow-on project.

## The Risk Analysis Process

The term "risk analysis" covers a wide range of analytical and management activities, but it has no consistent definition. For the purpose of our evaluation, we adopted a rather wide but not the most inclusive definition. We characterize risk analysis as a process in which technical methods—both quantitative and qualitative—are used to produce data to guide a decision on whether some action to reduce a risk is necessary. Risk analysis is the overall process through which hazards are identified, estimated, and evaluated and some decision is made as to whether

to reduce the level of risk posed by the hazard. This can include a decision to take no action.

Risk analysis encompasses both risk assessment and risk management, the two distinct elements upon which regulatory actions for public health, safety, and welfare are based. This is consistent with the terminology used by the National Academy of Sciences (NAS) in its 1983 landmark study, Risk Assessment in the Federal Government: Managing the Process. According to NAS,

"Risk assessment is the use of the factual base to define the health effects of exposure of individuals or populations to hazardous materials and situations. Risk management is 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."[1]

The phases of both elements of the process, as we have defined them in this report, are shown in figure 1.1. In our evaluation, we addressed each phase of the risk analysis process. It is described in more detail in appendix I, and technical terms are defined in the glossary.

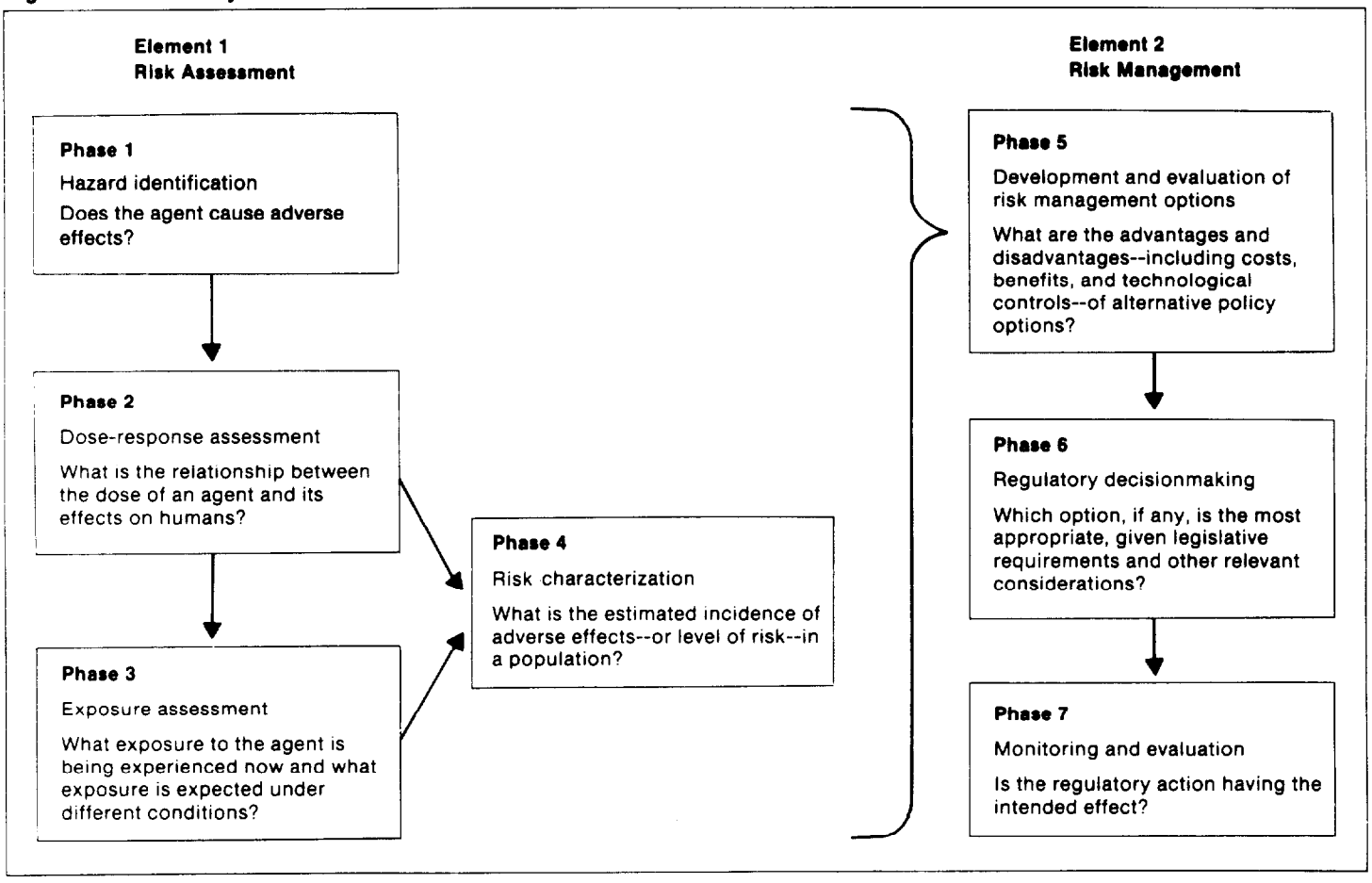
The phases of risk assessment are hazard identification, dose-response assessment, exposure assessment, and risk characterization. Hazard identification is the act of determining the specific risk sources that should be analyzed. The bases for making such decisions are often clinical and field observations of adverse health effects resulting from exposure to risk sources. Often, follow-up work to understand the associated exposure problems is conducted. This initial phase of the risk analysis process emphasizes the following activity: from an array of possible, suspected, or proven hazards, some are selected for in-depth analysis.

During the second phase, dose-response assessment, the magnitude of the risk associated with the hazard is estimated. Most often, the estimate is defined in terms of the probable occurrence of adverse effects on health. The emphasis during the dose-response assessment phase is on developing an estimate of human risk associated with a hazard.

The third phase, exposure assessment, is aimed at determining the extent of human exposure to the particular substance. This phase

[1]National Academy of Sciences, National Research Council, Risk Assessment in the Federal Government: Managing the Process (Washington, D.C.: National Academy Press, 1983), p. 3.

**Figure 1.1: Risk Analysis: Its Two Elements and Seven Phases**




Source: Adapted from National Academy of Sciences, National Research Council, Risk Assessment in the Federal Government: Managing The Process (Washington, D.C.: National Academy Press, 1983).

includes the characterization of the sources of exposure and an estimation of the level of exposure for different population groups. It may also include estimations of exposure under several different regulatory controls.

Risk characterization is the final phase of the risk assessment element of the risk analysis process. In this phase, the information accumulated from the previous phases is brought together to describe the nature of the human risk and estimate the magnitude of the public health problem. The uncertainties associated with the information available for each phase are considered, and attention is given to how to handle groups with different exposures or special sensitivities to the substance.

Completing these four phases completes the risk assessment part of the process. The information derived through risk assessment is then fed into the risk management element, which begins with the development and evaluation of options for controlling the identified risk and goes on to regulatory decisionmaking and the monitoring and evaluation of the regulations. The types of development and evaluation options available and the techniques for their development depend, to a large extent, on the legislation under which the substance is being regulated. In the next section, we discuss several general types of risk management approaches that control how options are generated.

The regulatory decisionmaking phase of risk management results in a decision concerning whether to regulate a risk source and, if so, the option to use. In this phase, it is important for the decisionmakers to consider the various options available, the degree of uncertainty associated with the options, and the possible effect of the regulations. If a regulation is deemed necessary, it is promulgated by the agency, first as a proposed and then as a final regulation.

The seventh and final phase, risk monitoring and evaluation, is one we have added to the NAS model. During this phase, the regulating agencies are responsible for ensuring that regulatory decisions are implemented as required and that the implemented regulations are, in fact, achieving their objectives. If they are, the process ends. If they are not, the development and evaluation of risk management options (phase 5) begins anew.

## Types of Risk Management Approaches

The categorization of risk management approaches is often dictated by legislation and is therefore important to our evaluation. The type of risk management approach that is used is usually controlled by the type of hazard being evaluated and, correspondingly, by a particular agency's legislative authority. The risk management approaches most generally used are

1. risk only,

2. risk balancing, and

3. technological control.

The risk-only approach characterizes analyses in which only the level of risk is considered in deciding whether a risk source should be reduced.

Risk balancing considers other factors in addition to risk level, such as the economic costs or benefits of regulation. The technological control approach emphasizes the application of the best technologies available to reduce either a hazard or exposure to it and, thus, to reduce the associated risk. Different statutes require the use of different approaches in developing and evaluating regulatory options.

## Issues in Risk Management

Much of the literature on risk analysis centers on how uncertainty arising in the risk assessment phases is dealt with in risk management—that is, how uncertainty about the level of risk is characterized, documented, and reflected in considerations leading to some decision on whether or not to control a risk source. A second issue concerns whether an estimate of risk is maintained as calculated or whether it is somehow revised without scientific basis.

Another issue is whether a full and adequate number of regulatory options for controlling a risk source is developed and how each option is evaluated. This is an especially critical concern, because the adequacy of the risk management response to an assessment is totally dependent on the options developed to control a risk source and the absolute and relative analysis of each option.

One other issue commonly addressed in the literature concerns whether information produced by a risk analysis is properly used. Every risk analysis that is conducted provides information that had not been previously available. Data developed on a particular risk source can be applied to similar risk sources; methods used within either the risk assessment or risk management element of a specific risk analysis can be applied to other analyses. Ensuring that such data and information are made available and used improves future work and advances the state of the art of risk analysis.

One important issue of concern that does not receive much attention in the risk analysis literature is whether evaluation follows analysis in order to ensure that an implemented regulation has the effect intended. The issue is not enforcement, in which the effort is to determine whether the sector being regulated actually implements the regulatory action. Rather, the question is whether the regulatory action, once implemented, actually works—whether the intended effects of the regulatory action are in fact occurring.

## Objectives, Scope, and Methodology

This project is a pilot study for a larger effort to address the adequacy of the risk analysis work supporting federal health regulation. The main objective of the pilot study is to provide some preliminary information on possible weaknesses and areas of strength in the federal risk analysis process. Through additional objectives that do not affect the present report, we will further refine our evaluation criteria, which can be used to assess the adequacy of a risk analysis and to gain experience in applying the methodology and determining how to streamline a possible follow-on project. Our review is intended not to evaluate the "correctness" of any final regulatory decisions but, rather, to examine how risk analyses are conducted.

Our focus in this report is on the adequacy of the general risk analysis process. In addition to risk assessment and risk management, the process includes the work that the regulatory agencies perform as well as others' scientific research that the agencies use. We define "adequacy" as including a technical or scientific dimension and a procedural or administrative dimension. The scientific dimension is the extent to which accepted scientific methods, procedures, and principles are employed. The administrative dimension is the extent to which accepted administrative practices are employed. Both risk assessment and risk management include scientific and administrative components.

To meet our objectives, we evaluated three cases, one each at FDA, OSHA, and EPA. The cases are of regulations that represent the three risk management approaches discussed above. The FDA case applied the risk-only approach to a proposed regulation for the control of methylene chloride under the Delaney anticancer clause (section 409(c)(3)(A)) of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 301 et seq.), which restricts the use of carcinogenic food additives. We examined a risk-balancing action, a final rule for inorganic arsenic exposure limits at OSHA, under the Occupational Safety and Health Act of 1970 (Public Law 91-596). At EPA, we investigated a final new source performance standard for the control of emissions of volatile organic compounds from onshore natural-gas processing plants under section 111 of the Clean Air Act (42 U.S.C. 7411(a)(1)(C)); the rule approached risk management through technological control. (Further information on our objectives, scope, and methodology is in appendix II.)

We selected the cases from regulatory actions published in the Federal Register from 1981 through 1985 for statutes we specifically selected to provide coverage of the three types of risk management approaches. For the three statutes, we drew up lists of potential cases that represented

the relevant risk management approach and that the agencies verified for accuracy. We eliminated final actions that were only minor modifications to previous regulations, so that our pilot cases would provide information about as much of the risk analysis process as possible. We selected our cases at random from the remaining actions on the verified lists.

In order to collect the information necessary for our evaluation of the risk analysis process supporting the three cases, we obtained all an agency's documents that constituted the written record of its risk analysis efforts. These materials included those contained in the public docket as well as relevant internal agency documents. Then we conducted interviews with the designated agency officials to fill in remaining gaps and answer any questions we had after examining the written material.

In order to assess the adequacy of the risk analysis work, we performed an analysis of the material we had collected. Our criteria for this analysis were those listed in appendix III that fell to the most specific level we deemed feasible. To develop these criteria, we reviewed related technical literature, guidelines, and general scientific information on the risk analysis process. Two expert panels reviewed the criteria, and we revised them in accordance with their comments (the members of the panels are listed in appendix IV).

Because our criteria rely on the published literature, within which there is no consensus, we do not argue that the criteria completely exhaust the characteristics required for adequacy or that all the criteria we employed are necessary to fully determine adequacy. However, they are representative of the criteria that have been recommended in the fairly extensive risk analysis literature. We will make final revisions to our criteria from our experience with the pilot cases before we implement any possible follow-on project.

The criteria are organized by phase of the risk analysis process. Each phase in figure 1.1 has criteria that address the administrative and scientific components of the general risk analysis process. The administrative area includes sets of related criteria on guidelines, internal and external expert reviews, and administrative review. The scientific components vary according to the work to be carried out for a particular phase. For example, the dose-response assessment phase includes sections on studies of animal and human responses to varying doses of an agent, and methods used to extrapolate from animal studies to humans, whereas the exposure assessment phase includes sections covering the

characterization of the sources of exposure, the populations exposed, and the characteristics of that exposure. Every phase also includes criteria for evaluating the documents and reports produced by an agency.

To rate the contents of the data we collected on each case against the standards of adequacy contained in our criteria, we employed conventional content analysis procedures.[2] First, we determined the criteria that were relevant to each portion of the documents and interviews. Next, we coded the information according to how well the relevant work had been performed for each criterion. Following the coding procedures established by Walter Lippmann and others, we assigned a rating to each applicable criterion, basing the ratings on our judgments of the adequacy of the relevant risk analysis work.[3] The ratings ranged from one, for work inadequate to meet a criterion, to five, for outstanding work. The other points on the scale were two for less than adequate work, three for adequate work, and four for more than adequate work. A zero was assigned if no work was performed. Rather than assigning a zero in instances in which no work could be expected, as when a criterion was not applicable to a particular case, we simply assigned no rating.

The literature on content analysis contains several approaches for ensuring reliable measurement—that is, little disagreement between analysts. One of the recommended procedures is to make the categories sufficiently specific to reduce coding from a judgmental task to clerical counting. To the extent feasible, we did this, but a judgmental component remained.

In addition to reviewing the literature from which our criteria were derived, each analyst received instructions on how to apply them. Finally, each case was coded independently by two staff analysts. After these analysts assessed each case, they compared their ratings and resolved differences greater than one point in the rating given to any criterion, arriving at a final rating for each criterion.

The reported ratings represent simple, unweighted averages across all applicable criteria for each set of criteria for the phase. Average ratings of 2.5 or less indicate less than adequate performance; average ratings

[2]Ole R. Holsti, Content Analysis for the Social Sciences and Humanities (Reading, Mass.: Addison-Wesley Publishing Company, 1969).

[3]W. Lippmann and C. Merz, "A Test of the News," The New Republic, special supplement, 23 (1920), 1-42.

above 3.5 are more than adequate. Because of averaging, adequate ratings may reflect outstanding as well as inadequate performance for specific criteria. Average ratings are reported on each component within a phase for each case study.

## Strengths and Limitations

The main strength of this study is the comprehensiveness of our examination of each case. Comprehensiveness was especially important, given that our study is a pilot. Our criteria were detailed and covered all phases of the general risk analysis process. We examined all available case-related documents and conducted interviews with agency personnel in order to fully examine the entire risk analysis process for each case.

A limitation of our study, which also derives from its being a pilot, is the small number of cases. Because we examined only one case at each agency, it is not possible to generalize about the quality of the risk analysis work conducted by the agencies or the federal risk analysis process as a whole. The study is limited to an examination of the quality of work performed for these three cases. In some instances, it is possible to make statements about general agency policies reflected in these cases, but this pilot evaluation is essentially oriented toward developing the most appropriate method for looking at risk analysis for a possible larger study to follow this one and toward examining the risk analysis work for these particular cases.

## The Organization of the Report

Each case has one chapter in this report, FDA first, OSHA next, and EPA last. The three chapters 2-4 provide general background information on the cases and present our findings, which are best understood in the context of the framework of evaluation criteria discussed above and contained in appendix III.

For each case, we present the results of our evaluation for each phase of risk analysis, giving the numeric ratings for that phase and following these with short descriptive statements and examples of the strengths and weaknesses of the risk analysis work conducted within the phase. The information we provide for each phase is oriented toward sets of related criteria and addresses areas that received more-than-adequate ratings (strengths) and areas that received less-than-adequate ratings (areas for improvement). Chapter 5 summarizes our findings, presents issues we identified across the three cases, and discusses major agency comments and our responses.

# FDA's Risk Analysis for Regulating Methylene Chloride

## Background

Methylene chloride, a hydrocarbon solvent, is produced in high volume for use in a variety of consumer and industrial products. It is used as a flammability depressant in aerosol products, including hair spray, and for extracting caffeine in food processing. It has been linked to cancer in animals. On December 18, 1985, FDA issued a proposed rule to ban the use of methylene chloride in aerosol cosmetics because the chemical had been found carcinogenic. However, the continued use of methylene chloride in the decaffeination of coffee was to be permitted. The rulemaking we reviewed represents FDA's reinterpretation of the Delaney clause of the Federal Food, Drug, and Cosmetic Act, and it reflects a risk-only approach to risk management.

The Federal Food, Drug, and Cosmetic Act authorizes the regulation of food and food additives, color additives, human and animal drugs, medical devices, and cosmetics. Methylene chloride is regulated under three sections of the act. Section 601(a) provides the authority to ban adulterated cosmetics, those that contain "any poisonous or deleterious substance which may render it injurious" under usual conditions of use. Section 409(c)(3)(A), the general safety clause, prohibits FDA from approving a food additive unless data are presented that establish that it is safe for a specific use. Section 409(c)(3)(A), the Delaney clause, specifies that "no additive shall be deemed to be safe if it is found to induce cancer when ingested by man or animal . . . ." Judicial review has determined that FDA is not required to "apply the strictly literal terms of the statute irrespective of public health and safety considerations . . ." in de minimis situations.[1]

The regulation of food and cosmetics is the responsibility of FDA's Center for Food Safety and Applied Nutrition (CFSAN). CFSAN is to ensure that foods are pure, wholesome, and safe; cosmetics are safe and made from appropriate ingredients; and both food and cosmetics are truthfully and informatively labeled. The regulation of food includes food additives, substances that may directly or indirectly become components of food. FDA must conclude that a food additive is safe and issue a regulation permitting its use before it is allowed into the market. Unlike the provisions concerning food additives, the act does not require a manufacturer to satisfy FDA that a new cosmetic product is safe before it is placed on the market. However, FDA does monitor cosmetics for safety and removes hazardous products from the market.

[1]Monsanto v. Kennedy, 613 F.2d 947 (D.C. Cir. 1979).

Methylene chloride is a direct food additive in its use for decaffeination. FDA's initial, premarketing approval of methylene chloride as a direct food additive for decaffeinating coffee was issued in 1967 in response to an industry petition under the general safety clause. In 1982, a National Toxicology Program study initiated concern at FDA over the potential carcinogenicity of methylene chloride in decaffeinated coffee and cosmetics. This study was the first to indicate positive evidence of carcinogenicity. The report was later withdrawn, because of apparent methodological problems, but other studies published between 1982 and 1985 reported evidence that methylene chloride was carcinogenic in animals.

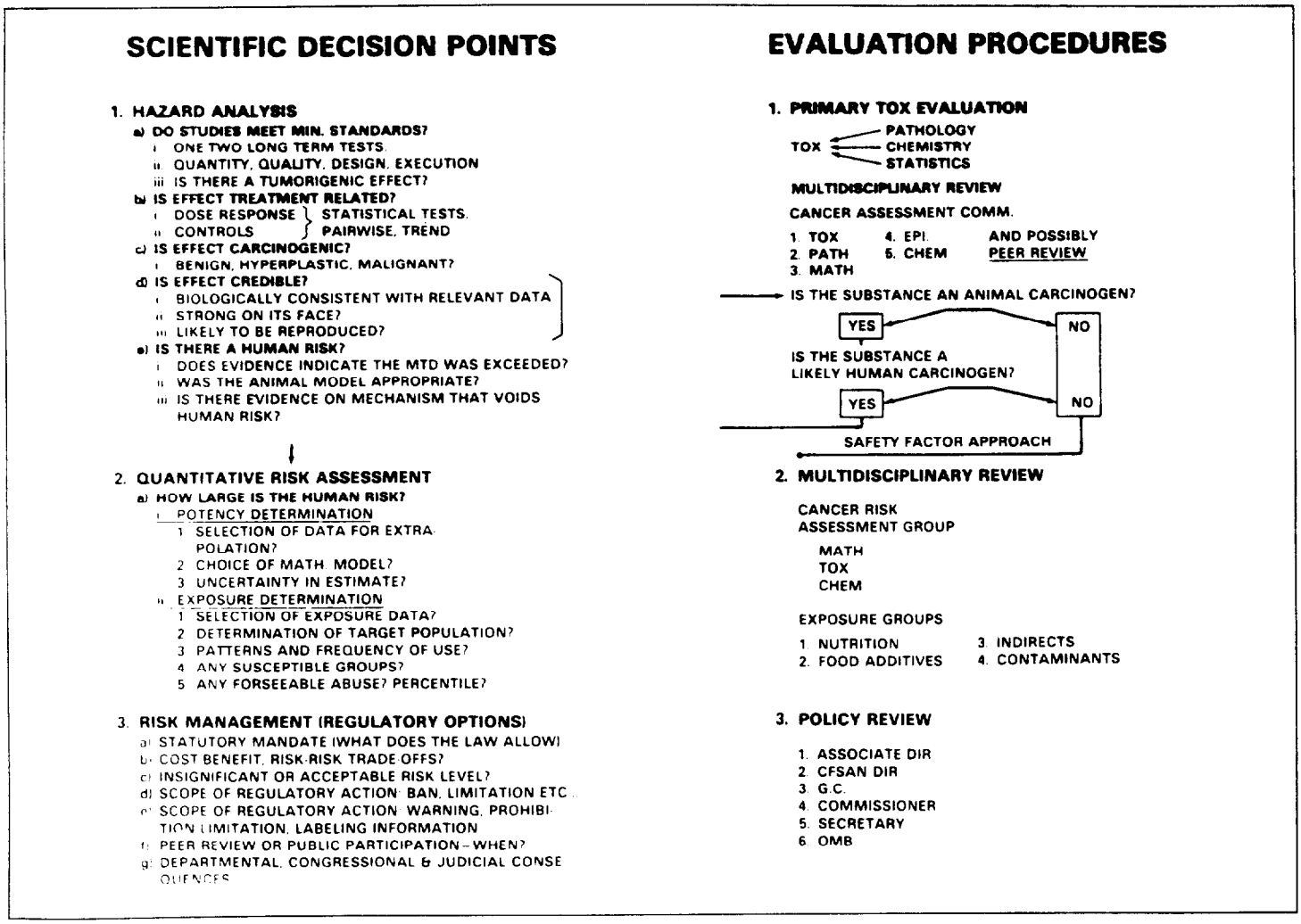
The rulemaking we examined is the result of risk analysis work FDA conducted from 1980 through 1985. The major portion of the work occurred in 1985, after the results of the studies showing strong evidence of carcinogenicity became available. The public comment period for the proposed rule was extended from February 18, 1986, to April 4, 1986, and reopened from December 5, 1986, to January 5, 1987, to allow discussion of the results of new studies provided to FDA in the fall of 1986.

In the proposed regulation, FDA tentatively determined that cosmetics containing methylene chloride are adulterated under section 601(a) and proposed to ban it totally in aerosol cosmetics but not in the decaffeination of coffee. FDA concluded that its use in the decaffeination of coffee poses a negligible, or de minimis, risk of cancer and that there would be no benefits from prohibiting it for this purpose. In place of a literal interpretation of the principle of the Delaney clause, FDA substituted the principle that some—that is, a negligible—risk is acceptable. FDA proposed that no change be made in the current limit of 10 parts per million residue of methylene chloride in decaffeinated coffee.

## FDA's Risk Analysis Process

FDA's risk analysis process differs according to the type of potential hazard examined. In this report, we limited our attention to direct food additives for which the Delaney clause is applicable. FDA's risk analysis process for direct food additives contrasts in some respects with our model. Comparing the two in detail is difficult, because FDA does not have guidelines that specify how it carries out the risk assessment element. Additionally, FDA officials told us that their risk assessment techniques evolve and change with each case they consider. Figure 2.1 depicts the phases of FDA's risk analysis process as it has been presented in FDA's briefing charts. In this chapter, we discuss only the items under "scientific decision points" in the figure.

**Figure 2.1: FDA's Risk Analysis Process[a]**




[a]Only the left side of this chart, "Scientific Decision Points," is discussed in the text. "MTD" under hazard analysis e(i) is "maximum tolerated dose," or the maximum dose that a test animal can tolerate for the duration of a bioassay without a significant effect on its health or average survival.

Source: FDA briefing charts, Food and Drug Administration briefing, Center for Food Safety and Applied Nutrition, 200 C St. S.W., Washington, D.C., March 10, 1986.

FDA's "hazard analysis" corresponds to what our model refers to as hazard identification. FDA combines in its "quantitative risk assessment" what our model refers to separately as dose-response assessment, exposure assessment, and risk characterization. In this part of its process, FDA discusses such things as uncertainties, exposure intensities and durations, and population groups affected by exposure, in order to

derive an estimate of overall human risk. These topics typify the items addressed under our criteria for exposure assessment and risk characterization. FDA also acknowledges that the same evidence is used to determine carcinogenicity and to assess the dose-response relationship.

FDA officials stated that the majority of FDA's or CFSAN's work involves premarketing approvals and is geared not toward doing primary research but, rather, toward reviewing the quality of research submitted to it. If information it needs is not available, the agency usually fills in data gaps with assumptions or awaits further outside research.

For the risk management phases, FDA has no overall coordinated policy. FDA officials stated that the three phases in our model—the development of options, regulatory decisionmaking, and monitoring and evaluation of regulations—do not occur as formal, separate actions at FDA. Groups and subgroups meet ad hoc as upper-level management needs additional information. When regulatory options are developed, they are presented in an action memorandum. Representatives of several FDA groups, including toxicologists, food-additive officials, regulation specialists, and attorneys, contribute to the memorandum, which is reviewed by the director of CFSAN, who then submits it to the commissioner of FDA for approval. Decisionmaking begins with staff recommendations that are presented in the action memorandums and are sometimes based on requests from the office of the commissioner of FDA or the office of the secretary of Health and Human Services to develop a regulation along certain lines.

# Evaluation Results for Methylene Chloride

## Hazard Identification

The groups in CFSAN involved most in the risk assessment phases are the cancer assessment committee and the quantitative risk assessment committee. Because these committees are composed of FDA scientists, the work conducted by the two committees is considered under "internal expert review." FDA's documented reviews of studies, memorandums, and other materials were primarily conducted by one or both of these committees. These reviews also fulfilled the major administrative review function within FDA. Therefore, we cannot discuss administrative

review separately for any of the risk assessment phases. The administrative review component is assigned a rating of 0 because no separate administrative review occurred. This should not be interpreted to mean that there was no consideration of administrative factors. Also, we could give no rating in the area of structure-activity relationship studies, since FDA used no such studies for methylene chloride. Our ratings for hazard identification are shown in table 2.1.

**Table 2.1: Criteria Ratings for FDA's Hazard Identification**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 2.3 |
| b. External expert review | [a] |
| c. Internal expert review | 3.3 |
| d. Administrative review | 0 |
| 2. Scientific | |
| a. Prioritization of potential hazards | 0 |
| b. Determination of hazard and weighting of evidence | 4.0 |
| c. Structure-activity studies | [a] |
| d. Short-term bioassays | 2.8 |
| e. Long-term bioassays | 2.9 |
| f. Epidemiological studies | 2.3 |
| g. Documentation and reporting | 3.8 |

[a]No rating assigned.

FDA relied primarily on two long-term animal bioassays sponsored by the National Toxicology Program that were validated by the program's board of scientific counselors. The board concluded that methylene chloride is carcinogenic in mice but that the evidence is equivocal in rats. Although FDA performed no additional outside expert review on the data used for hazard identification and dose-response assessment, we concluded that such a review would have served no useful purpose in light of the high quality of the program's review. Consequently, we considered the external expert review to be inapplicable for these phases of the process and assigned no rating.

## Strengths

We gave FDA a relatively high rating for the determination of the hazard and the weighting of evidence, even though FDA's law does not technically permit a weighting difference between studies on different species. However, FDA attaches greater importance to evidence from human studies than that from animal studies because, as experts on risk assessment

suggest, data on humans are more applicable. FDA sometimes uses data on humans, when the data are adequate, to modify an estimate of the risk of carcinogenicity when it believes that the animals may have been more sensitive than humans to the agent being tested.

FDA may yet do this in the methylene chloride case. It received new data in October 1986, including a human carcinogenicity study that may lead it to change its position on the use of methylene chloride in hair sprays. In this study, humans are shown to be at lower risk from inhaled methylene chloride than has been estimated from animal studies. This information may affect FDA's final decision on whether to ban methylene chloride in aerosol cosmetics.

FDA gives priority to the strongest data available. For example, the first animal study that FDA cites as showing positive evidence of carcinogenicity of methylene chloride is a study sponsored by the National Toxicology Program in which distinctly dose-related increases were found in the incidence of male and female mice developing both benign and malignant tumors of the lung. These results led FDA to conclude that methylene chloride is carcinogenic to male and female mice.

FDA's documentation and reporting were more than adequate. FDA characterized hazard-related uncertainties very well. For example, in a cancer assessment committee memorandum commenting on one study of mice, FDA described the unusual phenomenon of mice having convulsions with substantial frequency; investigators had tried to identify the reason without success. FDA expressed a lingering concern that the animals were unique or that the study was conducted unconventionally and reported this concern as a source of uncertainty. Assumptions were explained well and FDA's contentions were well supported by facts selected from studies.

## Areas for Improvement

FDA has adopted no formal guidelines for the risk assessment process. Agency officials told us that they generally subscribe to the recommendations of the 1983 National Academy of Sciences report, prepared under contract to FDA, and to the guidelines published by the Office of Science and Technology Policy. They emphasized that their risk assessment process has been evolutionary and that each case may require different techniques. They told us that FDA prefers to maintain the flexibility of not following formal written procedures, which in their view represent bureaucratic requirements. These officials stated that

the agency relies on many years of experience in cancer research to ensure the adequacy of risk assessment, including hazard identification.

However, by not having guidelines, FDA can give no assurance that important aspects of risk analysis will not be omitted. Guidelines could provide clear expectations and yet be structured to provide flexibility to meet the particular circumstances of each case. NAS and the HHS task force on health risk assessment recommended that federal regulatory agencies adopt guidelines for risk assessment, including hazard identification, to promote clarity, completeness, and consistency.

FDA does have guidelines on the testing requirements for determining the safety of additives. They are contained in a 1982 FDA publication entitled Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food. However, we were told that these guidelines are used only to provide information to industry on the testing requirements to be met for the premarketing approval of a new additive and that FDA does not use them for reviewing substances already approved. They did not play a role in the methylene chloride case.

According to several FDA officials, FDA also does not set priorities for the analysis of potentially hazardous substances, because it must spend considerable time on the petitions for the review of substances that come in through the regular premarketing petition-and-review process. The officials stated that regulations requiring FDA to handle cases within a specific time inhibit the agency from setting its own priorities. But this should not affect substances it has already approved, such as methylene chloride, which subsequent studies have shown may pose health hazards. The public needs some assurance that the potentially most dangerous risks are being analyzed first.

The 1982 guidelines mentioned above include provisions for a priority ranking scheme, but according to an FDA official, this scheme has not been implemented. Some approved substances have been examined to determine whether the data that supported their initial approval would meet current standards, but the guidelines have never been used for acting on approved substances. The reconsideration of methylene chloride was not initiated as a result of any priority system.

Our low rating of the available epidemiological studies concurs with FDA's own assessment. The epidemiological studies that FDA considered in the methylene chloride rulemaking are minimally adequate with

respect to our evaluation criteria. To the agency's credit, FDA did not rely extensively on these studies, noting in the proposed rulemaking that design limitations such as the small samples of workers and insufficient duration of exposure made it impossible to draw any definitive conclusions about methylene chloride's causing cancer in humans. We agree that these problems raise questions about the validity of the studies' findings, but this leaves the agency with a gap in the information it needs for its risk analysis process.

## Dose-Response Assessment

Because FDA used the same studies for hazard identification and dose-response assessment, many of our comments on hazard identification are applicable here. Our ratings for short-term bioassays, long-term bioassays, and epidemiological studies differ slightly because we applied one or two different or additional criteria in our evaluation of the dose-response assessment. Our ratings for dose-response assessment are in table 2.2.

**Table 2.2: Criteria Ratings for FDA's Dose-Response Assessment**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 2.0 |
| b. External expert review | 4.2 |
| c. Internal expert review | 3.3 |
| d. Administrative review | 0 |
| 2. Scientific | |
| a. Study selection for final estimation | 4.8 |
| b. Interspecies extrapolation models | 5.0 |
| c. Low-dose extrapolation models | 5.0 |
| d. Short-term bioassays | 2.7 |
| e. Long-term bioassays | 3.3 |
| f. Epidemiological studies | 2.6 |
| g. Documentation and reporting | 3.8 |

Study selection received a high rating, although little documented information was available on this topic. In order to substantiate our evaluation of FDA's procedures for study selection, we interviewed an FDA official who participates in decisionmaking in both the cancer assessment and quantitative risk assessment committees, the two groups in CFSAN involved most in risk assessment. We were told that a public comment concerning results in an Eastman Kodak epidemiological study was received after the proposed rule was published in December 1985 and

that it "flatly contradicts" the inferences about human risk derived from the National Toxicology Program study on rats and mice. FDA did not promulgate a final rule, partly because it was waiting for more information from the Eastman Kodak study, which it received in fall 1986. In doing so, FDA is demonstrating appropriate openness to new and potentially important epidemiological data.

The interspecies extrapolation work was very well done. FDA used a standardized scaling factor of body weight in milligrams per kilogram per day for its interspecies extrapolation for decaffeinated coffee. FDA was very concerned that although it has used this formula for many years, EPA has used a different formula, measuring body surface area, that predicted a much higher potency for methylene chloride's inducing abnormal tumors. After several discussions between the agencies, each decided to stay with the risk extrapolation method it had been using, arguing that there was no clear evidence for selecting one formula over the other.

FDA's low-dose extrapolation work was also well done. In addition to holding discussions with EPA on whether one interspecies extrapolation model may be better than another, FDA asked the Federation of American Societies for Experimental Biology to research evidence on interspecies and low-dose extrapolation models. The issue examined in low-dose extrapolation models was whether a linear or a quadratic model provided a better representation of human response. A linear model implies a steadily increasing response as a dose increases, while a quadratic model implies an exponentially increasing response.

Our ratings for the other components of dose-response assessment were adequate or less than adequate, but it should be noted that FDA has shown strength in several aspects of this work that are not reflected in the scores because it is offset by work that was less than adequate relevant to other criteria. For example, in the late 1970's, FDA promulgated good laboratory practices guidelines for organizations testing potentially unsafe products. Consequently, many of the studies we looked at showed strong evidence of good laboratory practices, and the ratings we gave them are high when measured against related criteria. Also, FDA has the scientific capacity to detect and quantify chemicals at levels down to about one part per billion. Additionally, through the use of extrapolation procedures, FDA is confident that it can estimate upper levels of risk even in noncarcinogenic additives that contain carcinogenic material. Further, FDA uses committees staffed by individuals of

different scientific expertise (the cancer assessment and the quantitative risk assessment committees), which helps build into the risk assessment process (at least for carcinogens) a comprehensive view of a potential hazard and the risk it could pose for humans.

## Exposure Assessment

### Strengths

Our ratings for exposure assessment are shown in table 2.3. Overall, source characterization was carried out more than adequately. Potential sources of exposure were examined, and the amount of emissions was examined for methylene chloride in hair sprays; this was not relevant for coffee. The production and distribution of methylene chloride was not examined for hair sprays but was considered for its use in decaffeinating coffee. One of the sources for this information was a description in a General Foods Corporation quality control manual.

**Table 2.3: Criteria Ratings for FDA's Exposure Assessment**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 0 |
| b. External expert review | 0 |
| c. Internal expert review | 3.3 |
| d. Administrative review | 0 |
| 2. Scientific | |
| a. Source characterization | 3.7 |
| b. Exposure routes and concentration | 3.9 |
| c. Populations at risk | 0 |
| d. Documentation and reporting | 3.2 |

Actual routes of exposure in natural settings were not examined in coffee and aerosol studies, but experiments were conducted by Dow Chemical Company with aerosols containing methylene chloride to determine its possible effect on humans. One study was performed in a beauty salon setting but under experimental rather than natural conditions, in that sprays and exposures were established as part of the experiment; conditions of human exposure were measured closely. Intensity of exposure for both hair sprays and coffee were also examined. FDA looked at the intensity of exposure to methylene chloride from decaffeinated coffee, using figures from General Foods Corporation on methylene chloride residue, and estimated consumption figures from several sources.

These data gave FDA a reasonable foundation for developing propositions on source characterization, and we considered the assessments technically strong.

When information was weak, FDA made reasonable efforts to strengthen it, including the use of conservative assumptions. In determining amounts of methylene chloride residue in decaffeinated coffee, FDA had figures available only for General Foods Corporation coffees. Because FDA did not know whether the residues of other manufacturers were as low as those for General Foods—100 percent of the samples tested had 0.10 parts per million or less—FDA assumed that all products would contain methylene chloride at a concentration equal to the current maximum limit of 10.0 parts per million. Most importantly, after proposing its rule, FDA independently tested samples of General Foods decaffeinated coffee.

## Areas for Improvement

FDA did not use an external expert review of the exposure assessment. An outside review may be especially important in exposure assessment when gaps in the data force an agency to make assumptions or when there are no data for entire areas, as for methylene chloride exposure. The lack of any external expert review here is a deficiency.

Populations at risk were not examined in the studies FDA used. In the two inhalation studies on hairsprays, volunteers were exposed to conditions in which aerosols were sprayed in a room, methylene chloride concentrations were measured, and blood was measured for elevated carboxyhemoglobin levels. These experiments were not designed to examine a "real" population of either hair care specialists or persons exposed to methylene chloride in home hair care products. FDA did address the risk level associated with inhalation for hair care specialists and home users, but it did so in the form of projections from experimental studies and assumptions rather than by examining a general population or occupational conditions.

Similarly for coffee, exposed groups were not examined, but estimates of exposure were computed for "average" consumers of decaffeinated coffee and for 90th-percentile coffee consumers, who consume twice as many cups per day as the "average" consumer. Separate estimates were derived for consumers of brewed and instant decaffeinated coffee. Extrapolation procedures were used to estimate different risks for these groups.

## Risk Characterization

For hair sprays, FDA derived a time-weighted average human exposure to methylene chloride for the consumer use of hair spray, basing it on an experimental human study. Using a linear model to extrapolate from the incidence of benign and malignant tumors in female mice exposed to aerosol cosmetics, FDA calculated that the upper boundary of human consumers' lifetime risk was 1 death from cancer beyond the expected number of deaths from cancer among 1,000 persons or 1 among 10,000 persons, depending upon whether the comparison of animal and human doses is based on the concentration in the air or on milligrams per kilogram of body weight per day. The occupational risk was estimated to be higher; that is, for hair care specialists, the upper boundary of risk was estimated at 1 in 100 or 1 in 1,000.

Similar techniques were used to arrive at overall risk estimates for drinkers of decaffeinated coffee. FDA calculated the risk for consumers of brewed decaffeinated coffee as 1 death from cancer beyond the number expected in 1 million persons who consume it throughout their lifetimes and for drinkers of instant decaffeinated coffee as 1 in 2.5 million.

FDA derived its information from a number of sources to develop overall risk estimates. Our rating of the agency's estimation procedures reflects this process. Our ratings for risk characterization are in table 2.4.

**Table 2.4: Criteria Ratings for FDA's Risk Characterization**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 0 |
| b. External expert review | 0 |
| c. Internal expert review | 3.3 |
| d. Administrative review | 0 |
| 2. Scientific | |
| a. Estimation procedures | 3.5 |
| b. Documentation and reporting | 2.6 |

## Strengths

Generally, FDA did a good job in the area of risk estimation procedures. The agency clearly laid out the steps it went through in developing overall health risk estimates. It also included and quantified uncertainties in the analysis; identified high-risk subgroups; defined hazards with respect to duration, frequency, and intensity of exposure; and accounted for the compounding effect of uncertainties.

## Areas for Improvement

Our observations about the administrative components for exposure assessment are the same here. Further, there is a potential problem with FDA's use of confidence intervals, or measures of uncertainty. FDA did not calculate confidence intervals for its risk estimates, so that any sampling error in the estimates—the extent to which a true value may vary from the estimated value—is not apparent. Because of the uncertainties involved in the agency's estimates and the assumptions that had to be made for such areas as exposure levels, the overall error (sampling and nonsampling) in the estimate is undetermined.

This is a concern particularly for the risk estimate for methylene chloride in decaffeinated coffee, because the agency used for the de minimis classification a threshold of 1 death from cancer beyond the expected deaths from cancer in 1 million persons consuming brewed decaffeinated coffee throughout their lifetimes. Without some measure of the uncertainties, questions may be raised as to whether the actual risk falls below this threshold. It should be noted that FDA used conservative assumptions when data were not available and that the agency believes that the actual risk is likely to be less. But since FDA's estimated risk for consumers of brewed decaffeinated coffee was 1 in 1 million, the same threshold for the de minimis classification, some attention to this issue seems warranted. In situations where the estimates of risk are at or near the threshold value being used, confidence intervals are particularly important.

## Development and Evaluation of Risk Management Options

FDA did not develop an option paper or action memorandum for the methylene chloride action. Options and their potential consequences were, however, discussed in an "environmental impact assessment" and in an "economic impact statement" entitled "Threshold Assessment of the Proposal to Prohibit the Use of Methylene Chloride in Aerosol Cosmetic Products." The latter was written to comply with executive order 12291, which requires agencies to prepare "regulatory impact analyses" if the economic effect of proposed rules would warrant it. FDA's statement for methylene chloride was prepared in order to determine whether a "regulatory impact analysis" would be required. Such statements are not actually part of risk assessment or of statements of options developed within risk management, but we have considered them in order not to rule out potentially relevant work. Our ratings for the development and evaluation of risk management options are in table 2.5.

**Table 2.5: Criteria Ratings for FDA's Development and Evaluation of Risk Management Options**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 0 |
| b. External expert review | 0 |
| c. Internal expert review | 1.5 |
| d. Administrative review | 0 |
| 2. Technical and scientific | |
| a. Development of risk management options | 1.7 |
| b. Documentation and reporting | 0 |

## Strengths

Neither the administrative nor the technical and scientific components of this phase were adequately performed as a whole, although the review of risk management options for practicality—one step in the development of options—was well conducted. The practicality of the decision to ban methylene chloride in hair sprays is not a specific point in FDA documentation, but the "environmental impact assessment" did say that

> "Available substitute solvents for methylene chloride in cosmetic aerosol hair spray products include ethanol-water systems, 1,1,1-trichlorethane and dimethyl ether. In addition, hair spray manufacturers can use alternate packaging forms such as spray pumps which preclude the need for methylene chloride. There are no essential uses of methylene chloride for which substitutes are not available."

## Areas for Improvement

The development and evaluation of risk management options and all risk management decisions proceeded without guidelines. When decisionmakers need information, these steps usually involve the FDA commissioner, deputy commissioner, associate commissioner, special assistants, and general counsel and the director of the Center for Food Safety and Applied Nutrition in an internal process that flows upward and is informal and iterative. Specific groups and subgroups meet to carry out risk management, but the meetings are ad hoc and not part of a formal process.

Experts within FDA apparently coordinate their comments, but they have no guidelines. Little information was available on their role in the development of options. Therefore, we considered the internal expert review inadequate. In addition, no guidelines or documentation were available to us to substantiate that an administrative review had been performed.

The formal development and description of options were not written down in the methylene chloride case. Thus, there was no true options development process to evaluate. The only relevant information available for us to evaluate in this step were the economic and environmental "impact" documents. The costs and benefits associated with FDA's actions, as well as those associated with possible alternative actions, were discussed in these documents, but they were inadequate for the methodological purposes of this step, since they were written in order to meet the requirements of the executive order. Documentation was inadequate for this step, because apparently no other documentation and reporting occurred.

## Regulatory Decisionmaking

### Strengths

Our ratings for regulatory decisionmaking are in table 2.6. Compliance with legislative authority was extensively discussed in this case. FDA took steps that effectively overturned more than a quarter century of understanding of the Delaney clause. For the general interpretation of the clause's ruling principle that no risk is to be accepted from a carcinogen deliberately added to the human food supply, FDA substituted the interpretation that some risk is acceptable provided it is minimal. In referring to legislative authority, FDA discussed the meaning behind the Delaney clause in the legislative history, and it cited court cases that permitted its use of a de minimis risk approach in different contexts. About one third of the discussion in the proposed rule proposes a use of the de minimis doctrine in a way that avoids applying the Delaney clause to the use of methylene chloride for decaffeinating coffee.

**Table 2.6: Criteria Ratings for FDA's Regulatory Decisionmaking**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 0 |
| b. Compliance with legislative authority | 5.0 |
| c. External regulatory review | 0 |
| d. Administrative review | 0 |
| 2. Technical and scientific | |
| a. Decisionmaking procedures | 2.8 |
| b. Documentation and reporting | 0 |

We have taken no position on the appropriateness of FDA's interpretation. We simply examined the justification the agency provided for public review. We believe FDA developed a logical and well-documented justification of its action relative to the legislative requirements.

## Areas for Improvement

Although FDA's position is that not having formal guidelines allows flexibility for tailoring each risk analysis to particular circumstances, we have observed that it also allows the possibility that important steps in a process may be forgotten or eliminated inadvertently. Further, the concern that guidelines inhibit flexibility is not universal at FDA. An FDA official who participates in risk management efforts stated that a standard for evaluating evidence is needed in order to ensure an understanding of the basis for decisions. The official additionally voiced concern that although FDA does a credible job of evaluating the science relating to risks, it needs to do a better job of separating science from policy issues: currently, the scientists are concerned with the law (such as interpretations of the Delaney clause), and the lawyers are concerned that the science may be affected by policy considerations. Thus, it may be that a substantial area of FDA's risk analysis process is not being adequately addressed.

There were no external reviews and no indications of how the administrative review was carried out. External reviews at this stage ensure that other agencies are given the opportunity to provide comments, that the agencies exchange information, and that inconsistencies in the way different agencies regulate the same substance can be studied. EPA had recently added methylene chloride to its list of carcinogens to be regulated, but FDA did not send its proposed regulation to EPA for review prior to publication in the Federal Register. Similarly, the proposal was not sent to the Consumer Product Safety Commission for review, although the EPA Federal Register notice mentioned work by the commission on methylene chloride in household products and the commission subsequently proposed a rule covering such uses.

FDA does not ostensibly consider "other factors," such as the effect of a rule on industry, as part of the actual rulemaking decision, but circumstances in this case made it appear as if FDA had, in fact, considered those factors. For example, FDA lacked industry information as to whether specific coffee producers could lower methylene chloride residues in decaffeinated coffee to a level below that permitted in current

FDA regulations. FDA realized that coffee producers other than the General Foods Corporation might have a difficult time meeting and measuring residues for a more stringent standard and might have to make major changes in the way they decaffeinated their coffee. FDA sought ultimately not to make the standard more stringent (although the staff proposal suggested doing so) but to calculate the human risk at current standard levels (10.0 parts per million). Doing this ultimately allowed an argument for setting the de minimis risk threshold at 1 death from cancer beyond expected deaths from cancer among 1 million lifetime consumers of brewed decaffeinated coffee, the "safe" threshold FDA normally uses, rather than at a lower level, such as 1 in 10 million or 100 million, which would be the threshold with more stringent standards.

Although we do not know that FDA gave any weight to economic effects in arriving at its proposed rule, the appearance is that it did, an appearance that seems to indicate that a purportedly risk-only approach could involve industry considerations, implying a risk-balancing approach. A decision resulting from risk-balancing considerations would not necessarily be unjust, but the public is entitled to know the basis on which decisions are made and whether statutory mandates are being followed. No documentation on the decisionmaking phase was available. Therefore, FDA's steps in reviewing the options and arriving at a decision about the standard are not clear and are not available for public review.

## Monitoring and Evaluation

We did not rate monitoring and evaluation, because monitoring is not applicable to a proposed regulation and evaluation is not feasible for food additive regulations. FDA has no formal program of evaluation for determining whether a regulated substance continues to pose a health hazard. FDA informally monitors research progress on substances it regulates through contacts with health risk assessors in such organizations as those of the National Toxicology Program, the Gordon conference, and the World Health Organization. When FDA obtains new information from them, it may be prompted to reconsider a substance, as it did with methylene chloride. FDA monitors compliance with its actions. It also has a system for monitoring adverse reactions to food additives voluntarily reported by medical personnel, researchers, and consumers. While this system can provide some specific indications of negative effects on health, it does not provide an evaluation of the effectiveness of food additive regulations.

The reason no evaluation studies are performed is that minute doses of food additives produce effects too small to measure, making epidemiological studies virtually impossible. The experts with whom we discussed this issue expressed considerable doubt that even the cumulative effects of human exposure to several food additives could be detected with epidemiological methods. It does not appear feasible to evaluate the effectiveness of food additive regulations, because of the small doses, the small effects of individual food additives, and the methodological difficulties of measuring either individual or cumulative effects. Therefore, reductions in the estimated 1-in-1-million lifetime risk from exposure to methylene chloride would not be possible to detect.

In order to place a 1-in-1-million lifetime risk in perspective, we calculated the estimated annual incidence of death from cancer from an individual risk of that level. Assuming a lifespan of 70 years and a population at risk of 200 million that grows 1 percent from one year to the next, the total annual incidence for the next year would be approximately 3. Epidemiological methods could not detect an effect this small, especially given that cancer is also produced by many other causes.

## Sources of Supporting Research

FDA's sources of supporting research on methylene chloride are enumerated in table 2.7. More than one third (11 of 24) of the studies FDA relied on were conducted by companies or organizations in industries affected by the proposed regulation.

**Table 2.7: Studies FDA Used in the Methylene Chloride Case by Sponsoring Source**

| | Sponsoring source | | | |
|---|---|---|---|---|
| Type of research | Other government agency | Government contractor | Academia | Industry |
| Short-term bioassay | 0 | 0 | 7 | 2 |
| Long-term bioassay | 0 | 2 | 0 | 5 |
| Epidemiological | 0 | 0 | 2 | 0 |
| Exposure | 1 | 1 | 0 | 4 |
| **Total** | **1** | **3** | **9** | **11** |

FDA officials told us that a typical study of rats or mice costs about $500,000, which does not include costs for reviewing the study or additional administrative costs. Additionally, FDA officials told us that the agency's facilities are inadequate for long-term testing but that a new facility soon to be completed should be adequate. However, FDA is not permitted to do such tests if they are an industry's responsibility under

the law, as for premarketing approval of the use of a substance. Because the majority of FDA's work involves premarketing approvals, the agency is not geared to doing its own research studies, and, therefore, when it finds deficiencies or gaps in the data available from bioassays, epidemiological studies, measurements of exposure levels, and the like, the agency must await further outside work or proceed on the basis of assumptions.

# Summary and Conclusions

## Strengths

The risk analysis process was consistently strong in three areas. FDA's general scientific work was good; the communications about risks, uncertainties, and assumptions were complete; and FDA consistently identified and relied upon the strongest scientific information available for its regulatory decisionmaking. All three strengths relate to work performed for the risk assessment phases of the risk analysis process.

FDA has shown strength in several ways in carrying out the risk assessment process. Many of the studies we reviewed showed strong evidence of good laboratory practices, and we gave them high ratings on our criteria. For internal expert review, FDA used committees staffed by persons of different scientific expertise. They helped build into the risk assessment work we evaluated a broad view of a potential hazard and the risk it could pose for humans.

FDA's communication of risks was another strength. Throughout the risk assessment phases, FDA was rigorously explicit about uncertainties and assumptions and about the consequences of those uncertainties and assumptions. This made clear FDA's logic supporting its findings and generated confidence in the integrity of FDA's risk assessment process.

Throughout the development of its proposal, FDA identified as important or not important all the studies it considered and described the limitations of the documents. That it did enhanced the credibility of FDA's risk assessment.

FDA also gave priority to the strongest information that it had. For example, in determining that methylene chloride is carcinogenic, FDA relied

heavily on a well-conducted, well-reviewed study of mice sponsored by the National Toxicology Program. When information was weak or when there were gaps in the data, FDA used conservative assumptions. For example, it assumed that all decaffeinated coffee contains methylene chloride residues at a concentration equal to the current maximum limit of 10 parts per million, even when some data indicated that specific brands might have lower residues. Further, FDA placed little reliance on related epidemiology studies that exhibited design limitations.

In the fall of 1986, FDA received the results of an epidemiology study conducted in England that may be more conclusive about how carcinogenic inhaled methylene chloride can be to humans than any information FDA had when it published its proposed rule. This information may affect FDA's final decision to ban the use of methylene chloride in aerosol cosmetic products. FDA extended the public comment period once from February 18, 1986, to April 4, 1986, and again when new study results were received, from December 5, 1986, to January 5, 1987. A final rule had not yet been promulgated when we sent a draft of this report to FDA for comment in March 1987.

These examples illustrate FDA's efforts to develop and use the best information possible.

## Areas for Improvement

The risk analysis work was problematic in four areas. FDA did not follow formal guidelines for any part of the risk analysis process; it did not systematically set priorities; documentation was often poor or nonexistent; and factors other than risk alone appear to have played a role in the final regulatory decision. These problems cut across both risk assessment and risk management but were the most prominent in the risk management phases.

FDA has virtually no guidelines, formal, written, or prescribed, on procedures for carrying out the phases of risk analysis. Its guidelines for toxicological testing are used only for the premarketing approval of new additives and cover only a limited part of the risk analysis process. FDA officials prefer to maintain the flexibility they believe is afforded by not following written procedures. However, not following guidelines may promote unpredictability and allow for the intrusion of regulatory policy into the science aspects of risk analysis, because their integration is not procedurally proscribed by guidelines. These problems may recur with each analysis and have negative implications for public confidence in the integrity and independence of FDA's risk analysis process.

FDA does not set priorities on substances to regulate. FDA has reasons for not setting priorities, including its resource commitments to the regular petition-and-review process. However, setting priorities would assure the public that the greatest hazards were being addressed first.

Some of FDA's internal decisionmaking memorandums were not well documented. We found citations incomplete, and studies alluded to should have been summarized more thoroughly to provide a better understanding of the decisions being made. Importantly, there was very little documentation of the risk management process with respect to administrative review, options development, or decisionmaking.

FDA does not ostensibly consider the effect of its regulatory decisions on industry, but circumstances in the methylene chloride case made it appear that it had considered it. FDA considered reducing the permissible level of methylene chloride residue in decaffeinated coffee to below 10 parts per million but chose not to do so. There were concerns that some coffee manufacturers might have to make major changes in how they decaffeinated their coffee if the standard were tightened. FDA's decision was not to tighten the standard (although the staff proposal suggested a more stringent standard) but to calculate the human risk at the current standard. Doing this ultimately allowed FDA an argument for the de minimis risk threshold of 1 death from cancer, potentially caused by the substance, beyond the expectation of deaths from cancer in 1 million people, the "safe" threshold FDA normally uses, rather than 1 in 10 million or 100 million under a more stringent standard. The use of thresholds may pose difficulties for classifying substances as de minimis when uncertainties associated with an estimate of risk mean that the actual risk may be much higher. Without some measure of the uncertainties, questions may be raised as to whether the actual risk falls below the threshold.

FDA has no formal program of evaluation for determining whether a regulated substance continues to pose a health hazard. FDA informally monitors research progress on substances it regulates and, when FDA obtains new information, it may be prompted to reconsider a substance. FDA also has a system for monitoring adverse reactions to food additives voluntarily reported by medical personnel, researchers, and consumers. The reason no evaluation studies are performed is that minute doses of food additives produce effects too small to measure, making epidemiological studies virtually impossible. Thus, there is no evaluation of the effectiveness of food additive regulations.

# OSHA's Risk Analysis for Regulating Inorganic Arsenic

## Background

On January 14, 1983, OSHA promulgated a final regulation that reduced the permissible exposure limit of inorganic arsenic in the workplace from 500 to 10 micrograms per cubic meter of air. Basing its decision on studies of respiratory cancer mortality among copper-smelter workers and chemical-production workers, OSHA determined that there is a significant risk of respiratory cancer at levels of arsenic exposure higher than 10 micrograms. The regulation reflects the risk-balancing approach to risk management that according to the Supreme Court is called for in the Occupational Safety and Health Act of 1970, which requires OSHA to set a 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" (section 6(b)(5)).

However, prior to a 1980 Supreme Court ruling on benzene, OSHA based its regulatory decisions on the technological control approach to risk management. The agency explicitly recognized the element of cost consideration inherent in technological control. After demonstrating that exposure to a substance was a hazard, OSHA controlled the substance to the extent that was technologically and economically feasible. The agency did not use quantitative risk assessments to help determine an appropriate level of exposure. For example, OSHA had reduced the arsenic standard to 10 micrograms per cubic meter of air on May 5, 1978. Shortly afterward, industry challenged the agency's decision in the courts, questioning not that inorganic arsenic is linked to respiratory cancer but, rather, OSHA's claim that this level is necessary to prevent material impairment.

In July 1980, while the arsenic standard was being reviewed by the Ninth Circuit Court of Appeals, the Supreme Court struck down OSHA's benzene standard, ruling that OSHA must in the future show that a proposed regulation is "reasonably necessary or appropriate to provide safe or healthful employment."[1] The agency was directed to make two threshold findings before issuing new or revised standards. First, the agency had to show the presence of a "significant risk"; second, it had to demonstrate that a new or revised standard would appreciably reduce or eliminate that risk. To make these determinations, a quantitative risk assessment was required. However, the court indicated that this was not to be taken as "a mathematical straitjacket" and that OSHA was not required to "support its finding that a significant risk exists with anything approaching scientific certainty," adding that

[1] Industrial Union Department, AFL-CIO v. American Petroleum Institute, 448 U.S. 607 at 639 (1980).

"a reviewing court [is] to give OSHA some leeway where its findings must be made on the frontiers of scientific knowledge [and] . . . the Agency is free to use conservative assumptions in interpreting the data with respect to carcinogens, risking error on the side of over-protection rather than under-protection."

In light of the benzene decision, the Ninth Circuit Court of Appeals remanded the arsenic regulation, upon OSHA's request, to the agency in April 1981 to allow public comment on the significance of risk associated with exposure to arsenic levels above 10 micrograms per cubic meter. OSHA then announced a limited reopening of the arsenic case, but 2 months later, on June 17, 1981, in American Textile Manufacturers Institute v. Donovan, the "cotton dust" decision, the Supreme Court upheld OSHA's standard, ruling that "cost-benefit analysis by OSHA is not required by the statute because feasibility analysis is."[2] Instead, the court ruled that once OSHA determines that significant risk will be substantially reduced, it must reduce exposures to the lowest feasible level.

OSHA published its final arsenic standard in the Federal Register on January 14, 1983. The publication presented the agency's conclusion that based on the evidence reviewed, a standard for arsenic of 10 micrograms per cubic meter of air would significantly reduce the risk of death from lung cancer among smelter workers. Neither the technological nor the economic feasibility of the regulation was at issue.

By requiring OSHA to demonstrate that the current state of affairs constituted a "significant risk," the benzene decision strengthened the balancing interpretation of the Occupational Safety and Health Act of 1970. As the Office of Technology Assessment observed (1981, p. 180), the courts have interpreted the use of qualifying terms such as "unreasonable" (or "significant," in this case) to describe "risk" as meaning that the risks from a substance are to be weighed against other factors in deciding whether and to what extent to control exposure to the substance.

Although the Supreme Court rulings clearly move OSHA's risk management in the direction of balancing, the exact nature of the balancing required remains somewhat unclear. A more explicit balancing interpretation would have required OSHA to formally balance costs and benefits. For the purposes of our study, we considered section 6(b)(5) of the Occupational Safety and Health Act of 1970 to present a balancing approach to risk management, since feasibility must be considered as well as a substantial reduction in risk. The requirement that OSHA show

[2]American Textile Manufacturers Institute v. Donovan, 452 U.S. 490 at 491 (1981).

an appreciable reduction of risk involves a direct comparison between the estimated cost of regulation and the estimated reduction of risk. Moreover, a technologically and economically feasible regulation might be challenged on the grounds that even though it is feasible, the exposure levels it set are lower than those that could be shown to result in an appreciable reduction of risk.

## OSHA's Risk Analysis Process

OSHA described its overall analytic approach for setting health standards for workers as a four-step process. In the first step, risk assessments are performed and used in determining whether a significant risk exists. In the second step, the agency considers which of the exposure levels under consideration will substantially reduce the risk. The third step involves determining a protective exposure level that is both technologically and economically feasible. In the final step, the cost-effectiveness of control technologies is assessed.

OSHA's regulation management system contains the agency's guidelines and procedures for implementing the analytic approach. Both the risk assessment and risk management activities involved in OSHA's standard-setting process are performed in the four steps of the system: initiation of action, action recommendation, notice of proposed rulemaking, and final rule. OSHA's system establishes policies governing the development of standards, assigns responsibilities for the work required to develop and review standards, and records the standards-development process through internal documents. The system is structured by the steps for standards development outlined in section 6(b) in the Occupational Safety and Health Act of 1970, which are somewhat different from the phases of the generic risk assessment process we described in figure 1.1. Each of the risk analysis activities is, however, included in the four steps.

### Initiation of Action

The first step is analogous to the hazard identification phase of the risk analysis process as we have defined it. A need for action is brought to the attention of any of the directors of OSHA's seven directorates through one or more of the following: a criteria document of the National Institute for Occupational Safety and Health; a petition or request from a company, labor group, or public interest group; an issue raised by the Congress; an issue identified by the agency's field staff. The need is reviewed by OSHA's regulation review committee and a recommendation is made to the assistant secretary for occupational safety and health.

## Action Recommendation

The majority of the risk assessment work is performed in the action recommendation step, which is taken in two parts. Part one consists of risk assessment and is referred to as "risk analysis." The second part, referred to as "regulatory analysis," consists of the development and evaluation of risk management options.

OSHA's "risk analysis" proceeds in the following sequence. First, the agency performs a critical review of the literature. Second, the agency selects from the literature appropriate studies for a quantitative risk assessment that includes dose-response assessment. This corresponds roughly to risk assessment in our model. Third and finally, OSHA determines the significance of risk by evaluating from both scientific and policy perspectives the information it has gained. According to our model, this is a risk management decision.

In determining the significance of risk for the arsenic decision, OSHA had, as required by the benzene decision, to establish that exposure to inorganic arsenic above a specific level would present a health risk to workers and it had to define the permissible exposure limit that would reduce or eliminate the risk. To make these two determinations, the agency had to obtain information on the smelters and identify the conditions under which smelter workers were exposed to arsenic. Finally, OSHA performed several quantitative risk assessments.

In the "regulatory analysis," OSHA considers the technological and economic feasibility of alternative regulatory responses designed to appreciably reduce or eliminate the significant risk. The regulatory analysis consists of establishing alternative courses of action and stating the cost effectiveness of each alternative. Standards are set to eliminate significant risks to the extent that is feasible.

## Notice of Proposed Rulemaking

After the assistant secretary reviews the action recommendation, the notice of proposed rulemaking is prepared. The notice presents the results of the "risk analysis" and the "regulatory analysis" performed in the action recommendation step. It may also include a number of analyses that are required by executive orders, such as a "regulatory impact analysis" or an "environmental impact statement." The notice is reviewed, it is published in the Federal Register, public comments are received by the agency, and these comments along with other relevant information obtained during public hearings are analyzed by agency officials.

## Final Rule

The last step in OSHA's four-step standard-setting process is the final rule. This step is analogous to the development of the notice of proposed rulemaking. After the agency evaluates the public comments and, if necessary, revises the work supporting its standard, the agency publishes the final standard in the Federal Register.

# Evaluation Results for Inorganic Arsenic

This section presents our review of OSHA's risk analysis process for the remanded arsenic standard.

## Hazard Identification

No hazard identification work was performed for this action. As we noted above, OSHA's rulemaking on arsenic resulted from the courts' limited remand of the case to the agency for the purpose of complying with the Supreme Court's benzene decision. But hazard identification was unnecessary during the period of remand, because it had been established as early as 1976 that arsenic is carcinogenic in humans, when the American Conference of Governmental Industrial Hygienists adopted threshold values for smelting and nonsmelting settings.

## Dose-Response Assessment

With respect to dose-response assessment, OSHA relied on 13 epidemiological studies, using them to determine the significance of risk below 500 micrograms per cubic meter. Direct evidence of carcinogenicity in humans cannot be ignored. The agency also reviewed the available animal studies, although most showed no evidence of arsenic's carcinogenicity. According to one agency official, epidemiological data are used when they are available rather than the less relevant animal data. Therefore, our evaluation of the dose-response work is based on OSHA's use of the 13 epidemiological studies, none of which it conducted or contracted for. Table 3.1 displays our results. In some instances, we assigned no rating because components dealing with animal studies were not applicable.

**Table 3.1: Criteria Ratings for OSHA's Dose-Response Assessment**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 4.0 |
| b. External expert review | 4.2 |
| c. Internal expert review | 4.6 |
| d. Administrative review | 4.8 |
| 2. Scientific | |
| a. Study selection for final estimation | 4.0 |
| b. Interspecies extrapolation models | [a] |
| c. Low-dose extrapolation model | 4.0 |
| d. Short-term bioassays | [a] |
| e. Long-term bioassays | [a] |
| f. Epidemiological studies | 3.4 |
| g. Documentation and reporting | 3.7 |

[a]No rating assigned.

## Strengths

OSHA has prepared written guidelines that specify the administrative, procedural, and review steps that must be taken in developing or revising a standard. Our review of the agency's documents showed that it followed these guidelines in the arsenic case.

In January 1980, OSHA published a policy document on cancer in the Federal Register, stating general policies and procedures for identifying, classifying, and regulating potential occupational carcinogens. These policies contain guidelines for weighing the soundness of epidemiological data. OSHA used only the epidemiological data guidelines in the arsenic case, because the other guidelines are for the identification and classification of substances causing cancer and inorganic arsenic had already been identified and classified as a carcinogen when the agency began its risk assessment work for the remanded standard.

Both internal and external expert review were more than adequate: the review process was documented, and the expert comments were addressed. In addition, OSHA has formal and written guidelines both for an internal advisory committee to advise the agency during the standards development process and for external technical review. OSHA followed these guidelines in its review of the inorganic arsenic standard.

In selecting the studies on which to base its analysis, OSHA reviewed the available epidemiological studies and consulted the opinions of experts

in epidemiology. All the studies showed a positive association between exposure to arsenic and death from lung cancer. The agency did consider the merits of an epidemiological study of copper-smelter workers employed by Anaconda Minerals Company that suggested that arsenic has a threshold below which it is not carcinogenic, but OSHA did not find this study convincing. Among its other weaknesses, the study had an inadequate sample size.

The agency examined the merits of the linear and the quadratic models in determining the appropriate choice of low-dose extrapolation models. This examination was necessary because the level of airborne arsenic observed in the smelters was higher than 10 micrograms per cubic meter. A linear model implies that the response to a dose is a function of the level of the dose, the response increasing directly as the dose increases. A quadratic model implies that the response increases exponentially, a response that would be appropriate for a substance with a threshold below which exposure is not associated with cancer. From statistical measures, the opinion of experts, and the work of other regulatory agencies, OSHA decided that the linear model was more appropriate than the quadratic model for characterizing the risk associated with arsenic.

The evidence of a dose-response relationship was strong. Lung-cancer mortality was significantly greater than expected for workers with high exposure to arsenic than for workers with low exposure.

The population at risk in the dose-response work was well defined and examined. All 13 epidemiological studies were of smelter workers, the primary target of the regulation. No assumptions about the applicability of the experience of other groups to smelter workers had to be made.

The techniques used to arrive at standardized mortality ratios were sound. In all but 3 of the 13 epidemiological studies, the general population in the states in which the smelters that were studied were located was used as the control group. The estimated duration of the smelter workers' exposure to arsenic in the 13 studies was based on strong evidence. In all but two studies, employment and mortality data were available for a total time span of 25 years or more, thus ensuring that workers who had been in the smelter environment for 20 years or more were included.

The examination of information on the workers' history was another strength. All but 2 of the 13 studies examined company records in order

to determine the departments in which employees worked. In addition to collecting employment history data from company records, 6 studies collected data from interviews with the relatives and friends of deceased workers, Social Security data files, or department of public health files.

Finally, OSHA's documentation and reporting were strong. Internal documents listed the contents of OSHA's public files for the arsenic regulation and summarized major studies and testimony. Biweekly status reports provided a chronological picture of the agency's review process.

## Areas for Improvement

With the exception of the general policies and procedures noted above, OSHA has no formal written guidelines for the scientific aspect of dose-response assessment.

Only 2 of the 13 epidemiological studies OSHA relied on reported measures of the strength of the association between death from lung cancer and exposure to arsenic, but these 2 studies showed a strong positive and statistically significant association. The lack of measures of association in the remaining 11 studies was a significant weakness.

Similarly, the examination of other factors that could account for the greater-than-expected rate of mortality from lung cancer among smelter workers was weak. Seven of the 13 studies attempted to consider confounding factors, but 6 of these 7 examined only the possible influence of smoking and the 1 other examined other potentially carcinogenic substances involved in the smelting process. In general, the attempts to measure confounding factors were not adequate. The data on smoking were often based on a family's or an employer's memories of deceased workers. One study of the 7 that attempted to consider confounding factors supplemented incomplete data on the workers' smoking histories with data on smokers from the general population.

Despite these limitations, the results were consistent. All but 1 of the 7 studies concluded that the high mortality rate associated with lung cancer among workers exposed to arsenic could not be entirely accounted for by smoking. The study that was the exception found significantly elevated rates of death from lung cancer only among smokers. This result was used to suggest that smoking and exposure to arsenic interact to produce a mortality rate in excess of that produced by smoking alone but that exposure to arsenic alone does not produce a significantly elevated rate of death from cancer. However, this study was limited by a small sample size and measurement problems.

One of the studies reported that it was not possible to control for all the possible confounding factors because workers exposed to high concentrations of arsenic also tended to be highly exposed to other substances. Because of this limitation, the study we refer to in the paragraph above stated that it was not possible to conclude that exposure to other substances could be discounted as responsible for the elevated mortality rates.

## Exposure Assessment

The exposure data OSHA used were taken from the 11 epidemiological studies that contained exposure data. Our ratings of OSHA's exposure assessment work are in table 3.2.

**Table 3.2: Criteria Ratings for OSHA's Exposure Assessment**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 4.0 |
| b. External expert review | 4.2 |
| c. Internal expert review | 4.6 |
| d. Administrative review | 4.8 |
| 2. Scientific | |
| a. Source characterization | 3.0 |
| b. Exposure routes and concentration | 1.9 |
| c. Populations at risk | 3.4 |
| d. Documentation and reporting | 3.3 |

### Strengths

OSHA has guidelines relevant to exposure assessment in its regulation management system, discussed above. OSHA applied its formal written guidelines for weighing the soundness of epidemiological studies to its assessment of the quality of the exposure data. The examination of the population at risk was strong in the studies OSHA relied on. The studies measured size and key characteristics of smelter workers—years of employment in a plant and location of employment within the plant—as well as company records would allow. Overall, OSHA's documentation and reporting were strong, and the epidemiological studies were cited in the Federal Register and available in OSHA's public files.

### Areas for Improvement

OSHA's general policies and procedures contain no guidelines for the technical or scientific aspect of exposure assessment. As for the studies

OSHA used, the exposure data reported in them were compiled from company records. Ten studies measured the intensity of exposure, and 11 measured duration. Of those that reported intensity, 3 based their estimates on urinary arsenic samples, 5 on airborne arsenic samples, and 2 on both urinary and airborne arsenic samples. This work was minimally adequate, for two reasons. First, no study measured individual exposure. Second, the measures of intensity did not adequately reflect the actual exposure of the workers.

The 10 studies that provided information on intensity of exposure estimated exposure levels from airborne or urinary arsenic samples that were taken sporadically and, on the whole, were probably not very representative of the actual exposure levels of the smelter workers in the various departments of the plants. For example, to estimate exposure levels for the entire study period, some studies used samples of airborne or urinary arsenic that had been collected years ago, when the levels of arsenic in the workplace were higher than they are now, thus overestimating total exposure. Other studies that used recent samples to estimate exposure early in the study periods underestimated total exposure.

The method used to classify workers into exposure categories was also weak. For example, 2 studies placed workers in the "heavy" exposure category on the basis of the highest level of arsenic to which they had been exposed for at least 1 year. Similarly, another study placed workers in the "heavy" exposure category on the basis of the highest level of arsenic to which they had been exposed for 30 days or more. This method of classifying plant workers into exposure groups is subject to potential biases. Some plant workers might have been assigned to work in a "heavy" exposure area within a plant for 1 year and then spent the rest of their work stay at that plant in a "light" or "medium" exposure area, but the method of assigning these workers to exposure categories described above would classify them in the "heavy" exposure group for their entire stay at the plant. This could potentially overload the "heavy" exposure category with workers exposed to less than "heavy" exposures, and the result would be a dilution of the actual rate of respiratory cancer for workers who worked in "heavy" exposure areas for all or most of their careers in the plant. Similar problems are possible with respect to the "medium" and "light" exposure categories.

## Risk Characterization

OSHA used nine quantitative risk assessments to establish that exposure to arsenic greater than 10 micrograms per cubic meter entails a significant risk and that the imposition of this level as an exposure limit would

reduce or eliminate the risk. We applied our evaluation criteria for risk characterization to these nine assessments, which were based on the 13 epidemiological studies. The nine assessments were performed by a variety of government and private sources and varied greatly in quality. Our ratings of OSHA's work are in table 3.3.

**Table 3.3: Criteria Ratings for OSHA's Risk Characterization**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 4.0 |
| b. External expert review | 4.2 |
| c. Internal expert review | 4.6 |
| d. Administrative review | 4.8 |
| 2. Scientific | |
| a. Estimation procedures | 3.0 |
| b. Documentation and reporting | 3.4 |

## Strengths

The adequate or fully adequate assessments clearly identified the risk group and estimated the degree of risk associated with exposure to arsenic at three or more levels—500, 50, and 10 micrograms—and expressed risk in terms of deaths from lung cancer beyond the expected number of deaths from cancer per 1,000 smelter workers for each level. OSHA's decisionmakers thus had a basis for comparing the estimated risks associated with the several exposure levels.

Two assessments compared linear and quadratic dose-response models and provided a coherent and complete rationale for preferring the risk estimates generated by the linear models. The rationale was based on measures of statistical fit as well as on the prior work of other regulatory bodies.

The adequate and better assessments spelled out uncertainties associated with the risk estimates, characterized the uncertainties in the underlying epidemiological studies and in the final risk estimates, and discussed the reliability of risk estimates, providing the reader with a sense of the degree of uncertainty associated with the risk characterization work. The assessments also clearly articulated necessary assumptions and reviewed the literature, giving special attention to the studies that had been used.

Yet another strong point of the adequate and better assessments was the reporting of best estimates and confidence intervals. For example, two assessments described a range of deaths beyond the expected number associated with different exposure levels.

With respect to documentation and reporting, the adequate and better of OSHA's nine quantitative risk assessments reported their results completely and comprehensively. They cited the sources referred to in their analyses and related the findings of other risk assessments to their own. For example, two assessments reported the risk estimates from three other assessments OSHA used in its review of the arsenic standard. This enabled the reader to interpret the risk characterization work.

## Areas for Improvement

One assessment was a two-page subsection of a lengthy and comprehensive report on the health and environmental effects of arsenic based on only one epidemiological study, and it did not fully discuss the uncertainties associated with the risk characterization work. The analysis did not report confidence intervals or a range of risk estimates, nor did it discuss the uncertainties associated with the assessment work.

Another risk assessment did not use regression analysis to estimate risks at low exposure levels. Its conclusions were based on a critique of the intensity of exposure reported in a selected number of epidemiological studies. It should be noted that OSHA did not find this analysis persuasive, because it contained a number of shortcomings. Chief among them was that its claim that there may be a threshold for risk from exposure to arsenic was based primarily on the findings of a pilot epidemiological study with a small sample size.

## Development and Evaluation of Risk Management Options

When the permissible exposure limit of 10 micrograms per cubic meter of air became a final standard in 1983, OSHA had already established cooperative agreements in which the agency, private industry, and labor unions agreed to work together to implement the technological controls necessary for complying with this standard. Eventually, OSHA signed a total of five cooperative agreements.

However, because the arsenic case was a limited reopening, as a result of the Supreme Court's ruling in the benzene decision, OSHA was required to show only that a "significant risk" existed at levels of exposure above 10 micrograms per cubic meter and that the proposed standard would appreciably reduce or eliminate that risk. Thus, the development

of risk management options work was not at issue in the limited remand However, on the basis of the risk assessment required by the remand, OSHA determined that the risk to workers exposed to 10 micrograms per cubic meter remained a significant risk. The agency did not propose to make the standard more stringent, because the previously conducted risk management options work showed that a more stringent standard was not feasible. Although the work was not conducted for the remand, it was used to support the required decisionmaking and, therefore, falls within the scope of our evaluation.

In order to examine OSHA's development of risk management options, we interviewed an agency official who had worked with the agency's contractors on the economic analysis of the arsenic standard in the mid-1970's. We also examined two major economic and technological feasibil ity studies the agency used for its 1978 arsenic standard, one commissioned by OSHA and the other by industry. The economic work of the mid-1970's had been reviewed by two of OSHA's economists as well as by the secretary of the Department of Labor and the secretary's assistant, both of whom were trained economists. We could not obtain specific information about this review because it was not documented. The official we interviewed had not been personally involved with the review and could not provide us with details. Table 3.4 displays our ratings of OSHA's development and evaluation of risk management options.

**Table 3.4: Criteria Ratings for OSHA's Development and Evaluation of Risk Management Options**

| Risk analysis component | Ratinç |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | |
| b. External expert review | |
| c. Internal expert review | |
| d. Administrative review | |
| 2. Technical and scientific | |
| a. Development of risk management options | 4.6 |
| b. Documentation and reporting | 4.6 |

## Strengths

The two economic and technological feasibility reports we reviewed provided a clear and comprehensive picture of the smelting industry, setting a context in which to interpret the analysis and the variables, such as the costs and benefits associated with alternative exposure levels and risk management options. The variables were clearly defined. For example, the dollar value of the engineering controls necessary to achieve

alternative exposure limits was specified, and the technological feasibility of complying with them was clearly presented. The definition of benefits was complete, in that epidemiological studies on smelter workers were reviewed, the population at risk was clearly defined, and the effects on health were identified.

The methods used to estimate the economic and inflationary effect of alternative exposure limits were clearly specified. For example, the assumptions that had been necessary in arriving at these estimates were clearly discussed. In addition, the uncertainties inherent in most estimation work and problems associated with estimating the costs and benefits of the proposed regulation were spelled out. For example, it was explained that the development of estimated benefits for alternative exposure limits had been limited because of deficient data on the extent and range of exposure levels.

In the documentation and reporting of the findings, one strong point was that uncertainties associated with each risk management option were characterized. Another strength was that the economic information used to estimate the costs of alternative exposure levels was well documented, the epidemiological data considered in estimating the health benefits were clearly cited, and the costs of the technologies evaluated with respect to feasibility were also well documented.

### Areas for Improvement

OSHA had no formal guidelines for carrying out its economic analyses in the mid-1970's, when work on the arsenic standard began, nor did the agency have formal and written guidelines for internal and external expert review or for administrative review. Therefore, our low ratings on the administrative components denote an area for improvement.

## Regulatory Decisionmaking

Our evaluation of OSHA's regulatory decisionmaking procedures was based on internal documents that OSHA provided us. Our ratings are shown in table 3.5.

**Table 3.5: Criteria Ratings for OSHA's Regulatory Decisionmaking**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 4.8 |
| b. Compliance with legislative authority | 5.0 |
| c. External regulatory review | 4.0 |
| d. Administrative review | 4.8 |
| 2. Technical and scientific | |
| a. Decisionmaking procedures | 2.5 |
| b. Documentation and reporting | 2.7 |

## Strengths

OSHA's internal documents clearly presented risk management options to the decisionmakers. They stated options, described their associated consequences, and summarized the background factors that led to the development of different options. In addition, the meetings between the decisionmakers and the staff working on the arsenic standard were well documented.

In devising its final standard, OSHA considered new data, even though they might have modified or delayed the standard. For example, an industry group submitted an epidemiological study that had been a pilot study of a small subsample of Anaconda Minerals Company copper-smelter workers that it believed suggested there might be a threshold of risk for arsenic. However, the sample size was small and the findings were not reliable, so industry had requested OSHA to postpone setting the final arsenic standard until a full study could be completed. OSHA weighed this request against its review of epidemiological data, expert testimony received during the comment and hearing period, and the data from the quantitative risk assessments and set the final arsenic standard only after considering all this information.

## Areas for Improvement

OSHA's decisionmakers were provided with three of the risk assessments and the notice of proposed rulemaking, all containing discussions of uncertainties, but briefing reports, summaries from the assistant secretary, and other internal agency documents did not mention the uncertainties associated with risk estimates, nor did they completely discuss the methodologies, data, or assumptions that were used in arriving at these estimates. There was no internal documentation indicating that the decisionmakers were provided with comprehensive information on the debate concerning the use of linear versus quadratic dose-response

models in the quantitative risk assessments. The internal documents did not cite the studies they had drawn on, so that there was no clear reference to the sources of information being presented, and they were excessively brief. They mentioned, but did not fully discuss, key issues.

## Monitoring and Evaluation

OSHA has not conducted any work aimed at evaluating the effectiveness of the inorganic arsenic standard in reducing negative effects on health. Our ratings of OSHA's monitoring and evaluation are in table 3.6.

**Table 3.6: Criteria Ratings for OSHA's Monitoring and Evaluation**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 0 |
| b. Communication and disclosure | 0 |
| c. Use of evaluation findings | 0 |
| d. External expert review | 0 |
| e. Internal expert review | 0 |
| f. Administrative review | 0 |
| 2. Scientific | |
| a. Evaluation formulation and negotiation | 0 |
| b. Evaluation structure and design | 0 |
| c. Data collection and preparation | 0 |
| d. Data analysis and interpretation | 0 |
| e. Documentation and reporting | 0 |

The agency's cooperative agreements with industry and the labor unions have provisions on a compliance component, but no evaluation of the standard's effect on health is planned. Under the cooperative agreements, OSHA monitors each plant to ensure that the engineering modifications and changes in work practices mutually agreed on by the agency and industry actually meet the standard of 10 micrograms per cubic meter, but OSHA does not collect information on the workers' health. The monitoring is aimed at enforcement considerations, not at determining the efficacy of the standard.

We consider this monitoring to be inadequate in meeting the purposes of the monitoring and evaluation step of our criteria. Epidemiological studies of effects on workers' health, of the sort used in setting the standard, are clearly possible, but the agency has no plans to sponsor such studies.

## Sources of Supporting Research

OSHA presented epidemiological data as its major source of evidence in making its final ruling on arsenic. Table 3.7 enumerates the studies by sponsoring source. As the table shows, the majority of the research came from academic sources.

**Table 3.7: Studies OSHA Used in the Inorganic Arsenic Case by Sponsoring Source**

| | Sponsoring source | | | |
|---|---|---|---|---|
| Type of research | OSHA or a contractor | Other government agency | Academia | Industry |
| Epidemiological | 0 | 0 | 13 | 0 |
| Quantitative risk assessment | 2 | 2 | 1 | 1 |
| Economic or technological feasibility | 1 | 0 | 0 | 1 |
| **Total** | **3** | **2** | **14** | **2** |

## Summary and Conclusions

One of the main strengths of the arsenic case is that the agency had sound epidemiological data specific to the population at risk—smelter workers. This eliminated the need to extrapolate from one population to another, thus decreasing the degree of uncertainty associated with the findings. Another strength was that a number of the epidemiological studies had sufficient documentation on exposure to allow for quantitative risk assessments.

The evidence of a dose-response relationship between exposure to arsenic and death from lung cancer was clear and strong. The low-dose extrapolation models used in the quantitative risk analyses were reasonable and sound. The agency examined the linear and quadratic models and determined from empirical evidence that the linear model was the more appropriate for characterizing risk.

We found the agency's written guidelines comprehensive and specific. They stipulated procedures for internal and external review as well as administrative review by agency officials.

There were, however, some weaknesses. We found that the data on exposure to arsenic did not accurately represent the experience of the workers analyzed in the epidemiological studies. The method some studies used to assign workers to exposure groups might have put workers exposed to low levels of arsenic in the same group with workers exposed to high levels. Two of the studies OSHA classified as risk assessments

were inadequate. One study did not use regression analysis and, therefore, no low-dose extrapolation was possible; the other study was excessively brief and incomplete.

There was no comprehensive discussion in the internal documents of the methods and data used in the risk assessments or the uncertainties associated with the risk estimates, and it is not clear what information the decisionmakers received about these uncertainties.

OSHA has not planned or conducted any work aimed at evaluating the effectiveness of the inorganic arsenic standard in reducing negative effects on health. OSHA monitors plants to ensure that they meet the standard, but OSHA does not collect information on the workers' health. The monitoring is aimed at enforcement considerations, not at determining the efficacy of the standard.

# EPA's Risk Analysis for Regulating Volatile Organic Compounds

## Background

Volatile organic compound emissions from onshore natural-gas processing plant equipment are regulated because they contribute to the development of ozone, a "criteria" air pollutant. The compounds are important as air pollutants almost entirely because they are precursors of harmful substances formed in the atmospheric photochemical system, not because they produce direct effects themselves. Ozone, a reactive form of oxygen, is the most prevalent photochemical oxidant. Oxidants, the primary constituents of photochemical smog, are products of atmospheric chemical reactions involving volatile organic compounds, nitrogen oxides, oxygen, and sunlight. Ozone and other photochemical oxidants are linked to coughing and wheezing, throat and eye irritation, headaches, and the aggravation of chronic respiratory conditions.

For regulatory purposes, EPA defines volatile organic compounds as "photochemically reactive compounds." In petroleum products, they include propane and heavier gases. On June 24, 1985, EPA promulgated new source performance standards for equipment leaks of volatile organic compounds from natural-gas processing plants. The standard covers compressors, valves, pumps, pressure relief devices, flanges and other connectors, and open-ended lines. It was projected that 220 gas plants would be affected in the first 5 years of the standard; beginning with the fifth year, emissions of the compounds would be 16,100 metric tons (1 metric ton equals 1.1 tons) less per year than without the standard at a cost of $1.5 million per year.

Several EPA officials, including a representative of EPA's office of the general counsel, confirmed that new source performance standards represent the technological control approach to risk management. EPA officials also stated that besides purely technological matters, cost effectiveness but not risk or risk reduction is considered in setting the standards. However, our evaluation indicates that the extent to which the standards are and should be based on technology is the subject of a major controversy within EPA. The competing view is that the technological basis is not logically coherent and that performance standards should be based firmly on balancing control costs against the benefits of reduced risk—that is, benefit-cost analysis. Case law establishes that EPA is not required to balance control costs against benefits when setting new source performance standards. However, case law does not preclude EPA from conducting a benefit-cost analysis and, in the case we evaluated, benefit-cost analysis contributed to the decisionmaking.

The Clean Air Act provides a complex strategy for reducing air pollution in which different risk management approaches are applied to different

types of air pollutants under varying circumstances. New source performance standards are only one part, so it is important to understand their role in the larger strategy. Air quality standards for criteria air pollutants, which are harmful and widely emitted, are to reflect the risk-only approach; in the development of national ambient air quality standards for these substances, case law has established that EPA may not consider cost or technological feasibility. Emission standards for hazardous air pollutants, which are particularly dangerous but not widely emitted, apply an approach based on technological control; according to EPA's formal guidelines, national emission standards for hazardous air pollutants require the best available technologies, unless the supporting risk assessment indicates the necessity of more stringent controls for protecting health. However, a recent court ruling has determined that EPA can use only health factors to determine safe emission levels for hazardous air pollutants.[1]

EPA applies different risk management approaches to criteria air pollutants under varying circumstances. For example, section 110 and sections 171-78, the key regulatory portions of the act, require EPA and the states to attain and maintain the air quality standards; neither cost nor technical feasibility may be used as a reason for failure to attain the standards. Section 111 of the act requires EPA to promulgate performance standards for new or modified stationary sources of air pollutants, including criteria pollutants. A major purpose of new source performance standards is to help prevent pollution by providing nationwide uniform minimum operational standards applicable to categories of sources, regardless of air quality. Section 111 requires EPA to set standards reflecting

> "the degree of emission limitation and the percentage reduction achievable through application of the best technological system of continuous emission reduction which (taking into consideration the cost of achieving such emission reduction, any nonair quality health or environmental impact and energy requirements) the Administrator determines has adequately been demonstrated."

We refer to a standard meeting this definition as the "best demonstrated technology." The statute dictates that the technological control approach to risk management be applied to new source performance standards, and EPA policy statements confirm that the standards are to be based primarily on technology.

[1]Natural Resources Defense Council v. U.S. Environmental Protection Agency, No. 85-1150, slip op. (D.C. Cir., July 28, 1987).

Some phases of the risk analysis process in our model are addressed under separate sections of the act. For example, the risk assessment work, including hazard identification, is performed according to section 108. Consequently, in terms of our model, the risk assessment work that supports a new source performance standard for a criteria air pollutant is performed according to section 108. Moreover, the logic of the technological control approach to risk management does not require all the phases shown in our model. The approach implies only the consideration of the feasibility of control technologies once a hazard such as ozone has been identified. Accordingly, EPA's formal risk analysis work specifically for new source performance standards generally begins with risk management, since the hazards controlled by the standards have already been identified by work performed according to other sections of the act.

EPA officials told us that a purely technological approach is not possible. More stringent techniques for reducing emissions can almost always be made available if society is willing to accept the required cost. Therefore, the technological control approach necessarily contains some inherent consideration of cost. Section 111 of the Clean Air Act explicitly requires EPA to select the best demonstrated technology, taking into consideration cost, "nonair quality health," environmental "impact," and energy requirements. EPA officials told us that because of the specific language of the act, EPA uses increasingly stringent levels of technological control in which cost becomes a progressively less important consideration. The best demonstrated technology represents the least stringent technological control. Where a major emitting facility is required to secure a permit for new construction or modification, it is subject to the "best available control technologies" and, in a "nonattainment" area, the "lowest achievable emission rates." The Clean Air Act states specifically that best available control technologies must be at least as stringent as the best demonstrated technology required by any applicable new source performance standards.

In order to examine the risk analysis process conducted to support an action representing the technological control approach to risk management, we wanted to select a case in which cost had not been a central consideration, but only new source performance standards that reflect best demonstrated technologies are promulgated as nationally applicable final rules. The more stringent levels of control are applied when permits are sought for constructing or modifying a major stationary source and may be required, for example, for individual sources in areas that have not attained the national ambient air quality standards. Thus

these levels of control are not applied nationally. Consequently, we eliminated these actions from our list of candidate cases. Therefore, when we randomly selected a case from section 111 final actions, we realized that the best demonstrated technology would involve some cost consideration.

Our objective was to evaluate the entire risk analysis process, including both risk assessment and risk management. Since volatile organic compounds were regulated in this case because they contribute to the formation of ozone, and because no additional risk assessment work is formally conducted to support the development of a new source performance standard, we looked at the risk assessment work that supported the most recent revision of the national ambient air quality standard for ozone, promulgated in 1979. However, a formal evaluation of this work did not prove useful for several reasons. First, the risk assessment work that contributed to the 1985 standard for onshore natural-gas processing plants was not the same work that was used to support the 1979 ozone standard. The ongoing risk assessment work that was used for the onshore standard has not been subjected to final rulemaking and is therefore beyond our scope. Second, hazard identification work would have been relevant, but none was performed for the 1979 revised standard because ozone had already been identified as a hazard. Third, the 1979 standard was controversial, resulted in litigation involving industry and environmentalists, and has been studied intensively, particularly with respect to the technical research and decisionmaking for the revised standard. Under these circumstances, it did not seem useful to reanalyze the case. In appendix V, we provide a brief description of the risk analysis for the 1979 ozone standard, in order to describe the larger process from which the onshore standard emerged.

# EPA's Risk Analysis Process

## The Phases of the Process

The new source performance standard we looked at was developed according to a priority rating of major categories of stationary sources of air pollution that was mandated by the 1977 amendments to the Clean Air Act. The crude oil and natural gas production category includes natural-gas processing plants and is ranked 29th on the list of 59 major source categories.

The development of a standard is the responsibility primarily of the emission standards and engineering division of EPA's office of air quality planning and standards. The actual work is performed under contract by independent firms that are responsible for virtually all aspects of the regulation, including the drafting of the action memos that request the administrator's approval and the standard itself. EPA personnel supervise the process and edit and approve the results. The process is designed to take approximately 4 years to complete.

EPA divides its risk analysis activities for all new source performance standards into three basic phases that occur prior to the appearance of the final standard in the Federal Register: the source category survey, the development of a "background information document" for the proposal, and the development of the standard. The source category survey includes a literature search and telephone survey, one or two initial plant visits, the preparation of a report, and a decision concerning pollutants and facilities for intensive examination in the next phase. The source category survey identifies specific emission sources within a source category and corresponds generally to what we refer to in our criteria as source characterization. Our model of the risk analysis process conceives of this work as part of the exposure assessment portion of risk assessment. However, the specific circumstances of the Clean Air Act have led us to discuss our criteria for this work under the development and evaluation of risk management options.

The development of the background information document for the proposed standard is roughly the same as the phase we call the development and evaluation of risk management options. It includes systematic plant surveys, approximately four on-site tests of the source of emissions, an industry review of the technical sections of a draft of the document, and the final drafting.

The technical research is limited to identifying and evaluating control techniques. It produces estimates of the "incremental," or marginal, cost per unit of reduced pollutant for each of the technologies considered. The result is a series of option packages ordered by abatement effectiveness and cost effectiveness.

Decisionmaking begins as the technical research for the development of risk management options nears completion. The decision represents EPA's judgment of reasonable incremental or marginal costs per unit of pollutant that is reduced. "Cost effectiveness guidelines," also referred to as "thresholds," "cut-offs," or "criteria," represent the maximum

control cost per unit of reduced pollutant that is deemed reasonable and are used, at least in part, to determine the level of control that represents the best demonstrated technology.

The first step in decisionmaking is the development of the proposed standard. This includes the preparation of the contractor's recommendation and the drafting of the proposed standard and a preamble to it. A staff working group, a steering committee, the National Air Pollution Control Techniques Advisory Committee, each assistant administrator, and the Office of Management and Budget (OMB) all review the proposal package. The proposed standard is announced and published in the Federal Register after the EPA administrator has approved it. EPA also releases the background information document for the proposal, which reports the results of the technical research, or what is known as the "technical basis of the standard."

Once the proposed standard has been published, EPA responds to the public comments, documenting whatever changes it makes in the regulation in the background information document for promulgation. Again, the package is reviewed by a working group, a steering committee, each assistant administrator, and OMB. With the administrator's approval, the final rule is published and distributed.

## Technical Research for New Source Performance Standards

Generally, the activities required in the development of the background information document for proposal correspond to what we have defined as the development of risk management options, and the development of the standard corresponds to regulatory decisionmaking. We do not have criteria for each specific stage of EPA's research process for developing risk management options because of the wide variation in the types of research that may be necessary and the lack of emphasis on risk management in the risk analysis literature from which we derived our criteria. For example, our criterion on the adequacy of research methods does not distinguish between EPA's methods for estimating emission reduction and its methods for estimating cost. In applying this criterion to the research as a whole, we found quite different strengths and areas for improvement across the several stages of the research, so it is important to understand EPA's research process for new source performance standards.

We describe the stages of the research as (1) the source category survey, (2) the estimation of "baseline emission factors," (3) the identification of available control techniques, (4) the estimation of the effectiveness of

control techniques, (5) the development of "model plants," (6) the estimation of overall emission reduction, (7) the estimation of the cost of the control technique, (8) the estimation of cost effectiveness, and (9) the assessment of economic "impact." The estimation procedures involved in the first seven stages are essential to the validity and usefulness of the cost effectiveness estimates in stage 8. According to the background information document for the proposed standard, the analysis in stage 9 is used to define reasonable control costs.

The source category survey consists of site visits to plants in order to identify specific emission sources, such as types of equipment exhibiting significant leak rates. Pollutants to be considered for control also are examined.

The estimation of baseline emission factors is one of the more important components of the technical research, because it is the foundation for the subsequent estimates. Baseline emission factors are the average amount of emissions per day from one unit of each type of source under consideration prior to regulation.

Technical feasibility and safety are considered in the identification of control techniques and the estimation of abatement effectiveness. Control efficiency is expressed as the percentage of uncontrolled emissions reduced by the controls.

The estimation of overall emission reduction requires estimating the number of components that will be affected. This is accomplished by first defining hypothetical types or sizes of plants, referred to as "model plants," each of which has an estimated number of sources. The result is an estimate of the number of units of each source type at each hypothetical plant.

Overall emission reduction is then calculated by multiplying the number of components of each type at each model plant first by the baseline emission factor and then by the percentage control efficiency. This gives an estimate of the amount of pollutant reduced for each regulatory option at each model plant. The estimation of control costs includes estimating direct capital and installation costs for each regulatory option and each model plant. Annual, annualized, and net annual costs are also calculated for each option at each model plant. Annual costs, in this case, included estimates of the cost of the leak-detection and repair program. Capital and other costs are annualized by multiplying them by appropriate factors based on estimated equipment life, administrative

and support costs, and interest rates. The net annual cost is calculated by subtracting the estimated value of any product recovered per year through control ("recovery credits") from the estimated annualized cost.

The most important result of the research process is the estimation of cost effectiveness, defined as dollars per unit (megagram or metric ton) of reduced pollutant. Incremental or marginal cost effectiveness is calculated by dividing the increase in net annual cost by the additional or incremental amount of reduced pollutant for each successively more stringent option or specific technique.

The effects of each option on industry and consumers are considered in the economic "impact" analysis. Estimates are developed for production, employment, and price fluctuations; plant closings, curtailments, or relocations; populations affected most by price increases; and the international economy. Industry growth figures, also examined in this stage, are used to project the number of facilities to be affected after 5 years, as well as the energy and environmental impacts.

## Evaluation Results for Volatile Organic Compounds

The standard for onshore natural-gas processing plants emerged from a program of research on crude oil and natural gas production that began in 1976. Regulations controlling the emission of volatile organic compounds and other pollutants in several related industries emerged from this research. Specific work for the standard began late in 1979. Since the final regulation was promulgated in June 1985, the risk analysis process lasted more than 5 years. Since no risk assessment work was conducted specifically for developing this new source performance standard, and we did not evaluate the risk assessment work for ozone, the following discussion considers risk management only.

### Development and Evaluation of Risk Management Options

Since our ratings for risk management options are averages for our criteria and the stages of EPA's research process, significant strengths and areas for improvement tend to offset one another in the ratings we provide in table 4.1.

**Table 4.1: Criteria Ratings for EPA's Development and Evaluation of Risk Management Options**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 3.9 |
| b. External expert review | 3.8 |
| c. Internal expert review | 0 |
| d. Administrative review | 3.7 |
| 2. Technical and scientific | |
| a. Development of risk management options | 3.3 |
| b. Documentation and reporting | 1.9 |

## Strengths

EPA has developed extensive formal written guidelines for developing risk management options. They are contained primarily in EPA's manual for contractors and cover each step in detail, including when to write memos documenting the completion of project milestones, to whom to address them, and what information to include. Specific data collection activities are also spelled out. For example, the guidelines require reviews of the technical literature and telephone surveys to gather information on control techniques. They also require "regulatory letters," under section 114 of the Clean Air Act, that request information from industry. Original data collection is required for estimating the emission factors. For the onshore standard, the guidelines were generally followed and their use was documented.

As our rating indicates, the external expert review, conducted by the National Air Pollution Control Techniques Advisory Committee, was more than adequate. All aspects of the technical basis of the standard were reviewed in a public forum. The administrative review also was more than adequate. The formal written guidelines require three rounds: a staff level "working group" review, a "steering committee" review in the office of the director, and a "red border" review by an assistant administrator. The guidelines were followed and the process was documented.

The technical research was extensive. It included three regulatory option packages, as well as a "no regulation" option that was used as the basis for evaluating the other options. Cost estimates were developed for each piece of equipment and work practice in each option. Worst-case assumptions were generally avoided. The examination of production and distribution and the comparison of the controls with

those in similar industries were strong. Finally, the analysis of the economic effects was generally strong. It was based on comprehensive Department of Energy and industry publications. EPA estimated that the increase in natural gas prices because of the regulation would be less than half of 1 percent and no plant closures or curtailments were expected. The background information document for proposal concluded that the effects on profitability, output, growth, employment, productivity, and international trade "will be negligible or zero due to the regulatory alternatives analyzed."

## Areas for Improvement

EPA's guidelines do not require probability sampling of test plants and other such scientific standards that ensure the precision of estimates and quantifiable sampling error or uncertainty. EPA recognizes that probability sampling of test plants is preferable but has concluded that it is too costly. Such a conclusion may be warranted under practical circumstances but it produces a significant weakness in the quality of the technical research. We did not assign a higher rating to the guidelines because of this shortcoming in them.

We did not give the external expert review the highest possible rating because the National Air Pollution Control Techniques Advisory Committee panel is not adequately balanced. It is composed almost entirely of industry engineers. One member represents an environmental group. There is no representation of independent engineers such as those from academic institutions. The lead engineers on the project stated that industry representatives routinely overestimate cost and underestimate feasibility and that the industries that supply control techniques underestimate cost and overestimate feasibility. Under these circumstances, a balanced panel would have to include independent engineers.

EPA has no distinct internal expert review for the development of new source performance standards, and therefore a 0 was assigned for the internal expert review component. This should not be interpreted to mean that there was no consideration of the technical basis. The technical basis for a standard is reviewed during the administrative review.

We did not rate the administrative review higher because the guidelines do not specify the aspects of the work that must be reviewed. The guidelines tend to emphasize the timing and form of the review rather than its substance.

Although the technical research for the development of risk management options was extensive and contained several strong elements, we found significant areas for improvement in eight (or 50 percent) of the matters covered by our criteria: (1) the treatment of uncertainties, (2) the use of assumptions, (3) research methods, (4) the examination of potential sources of emissions, (5) the practicality of control techniques, (6) the estimation of the amount of pollutant emissions, (7) the degree of abatement effectiveness through control, and (8) cost estimation, including control costs and their relationship to emission reduction and the economy. Each of these areas offset the strengths discussed above and resulted in an average rating that does not reflect the variation across the specific criteria.

1. Uncertainty was not adequately quantified or discussed. Quantified uncertainties were presented in the background information document only for baseline emission factors. Although the statistical confidence intervals that were reported applied only to the six plants actually tested, the figures were used as if they were representative of the industry. No qualitative discussion of uncertainties was included for the final cost effectiveness estimates.

EPA's guidelines require precision of only plus or minus 30 percent for the final cost effectiveness estimates, adding that precision of plus or minus 5 percent is not considered cost effective for developing new source performance standards. Uncertainty of plus or minus 30 percent implies that regulatory options differing by more than 60 percent in estimated cost effectiveness may not differ in actual cost effectiveness. The lead engineers stated that the actual precision was between plus or minus 30 and plus or minus 100 percent. They stated that decisionmakers do not want to see ranges of values; instead, "point estimates" are preferred. However, these point estimates of cost effectiveness were not the result of statistically valid estimation procedures; they were the result of a complex process with numerous inputs and numerous potential sources of unquantifiable error.

2. The use of assumptions without adequate rationales reduced the rating we assigned. Although the use of assumptions is often necessary when empirical data are lacking, each time untested assumptions are used, a degree of unquantifiable uncertainty is introduced. For example, in the background information document for the proposed regulation, EPA assumed that half the compressors in the facilities affected by the regulation would be reciprocating compressors and the remaining half would be centrifugal compressors, but it provided no rationale for this

assumption and cited no document as a source. Industry subsequently provided a survey showing that reciprocating compressors predominate in the natural gas industry. EPA, estimating that the enclosed seal devices required for controlling emissions would be much more expensive for reciprocating than for centrifugal compressors, then dropped reciprocating compressors from the final regulation. The revised estimate of the relative frequency of the two types of compressors had led to the judgment that the cost of the equipment was "unreasonable." The consequence of the use of untested assumptions is that the results tend to become an artifact of the research process rather than an empirical description of the actual conditions.

3. Problems with research methods also reduced the rating we assigned. For example, the model plant methodology was inadequate. A small, a medium, and a large plant were developed, but no representative data were used to estimate the number of components at each one. EPA stated that the estimates were mostly based on a sample of plants that was examined intensively, but it was not a probability sample and in some instances assumptions were used.

The number of components at each plant was "indexed" by calculating the ratio of components of each type to "vessels," which EPA stated was a more easily counted population. However, "vessel" was not defined. The lead engineer expressed dissatisfaction with the model-plants procedure in this case, because vessels could not be defined adequately. Moreover, the number of vessels was estimated according to the ratio of vessels to "process trains," which also was not defined. Finally, EPA based its entire characterization of the small model plant on the assumption that there are gas plants with configurations (ratios of components to vessels and vessels to process trains) similar to the plants EPA examined. EPA assumed that the numbers generated by this procedure were in some sense representative of the industry, but the errors thus introduced directly affected the estimates of overall emissions, emission reduction, and cost effectiveness.

4. Problems with EPA's identification of potential sources of emissions reduced our rating. EPA's source characterization concluded that many components leak, but attention was focused on several specific types of equipment without explaining why other equipment was not examined. The internal documentation stated that the data base for some of the sources not examined was "weak and from a regulatory standpoint unsupportable." However, the internal documentation did not address

all the source types originally considered, and it did not cite the data referred to in support of the conclusion.

5. We found an important weakness in the review of the practicality of control techniques. EPA's guidelines indicate that section 114 regulatory letters and telephone surveys are to be used, but no probability samples or standardized interview schedules were used in the surveys. Information concerning the practicality of potential control techniques was frequently based on limited telephone contacts. The lead engineers on the project stated that OMB restrictions limiting the number of external contacts to fewer than 10 without OMB's approval prevented a more systematic inquiry.

6. Problems with the estimation of the amount of baseline emissions also reduced our rating. EPA measured volatile organic compound leaks from a sample of equipment components at four natural-gas processing plants and two additional plants were tested in a study sponsored by the American Petroleum Institute. Probability samples of leaking components were tested at the four EPA plants: a large sample was tested with a handheld meter, an organic-vapor analyzer that measures volatile organic compounds in parts per million (ppm) ambient concentration in the vicinity of a component, and a smaller stratified subsample was tested with an expensive but more definitive gas-chromatography technique that measures the amount of volatile organic compounds emitted from a component during a given period of time.

In order to estimate what the definitive gas-chromatography scores would have been for the equipment components tested only with the handheld meter, EPA applied regression analysis to the relationship between the two sets of scores for the sample that received both tests. The correlation was weak, the confidence interval was wide, and, thus, a large band of uncertainty was indicated. Moreover, 37 percent of the equipment components that received both tests, the most heavily leaking components, were not used in the regression analysis, because the handheld meter scores for these components were above the maximum accurate reading of 100,000 ppm. Instead, the average gas-chromatography score for the equipment components excluded from the regression analysis was assigned to these components.

The gas-chromatography sample was stratified to ensure precise estimates of the small percentage of the total number of components that leak heavily. But the absence of the most heavily leaking components from the analysis seems to have largely nullified this effort. The lead

engineer agreed that including in the analysis the 37 percent of the components that were omitted would have resulted in a larger estimate of total emissions and, therefore, a smaller estimate of the control costs per unit of reduced pollutant, or cost effectiveness, for all control techniques. More stringent controls might then have been deemed reasonable.

7. Difficulties with the abatement effectiveness estimates also reduced the rating we assigned. Even though specific equipment was assumed to eliminate leaks entirely, costs for its repair and replacement were included in the cost analysis, indicating that it would eventually leak. Therefore, emission reduction from these devices was probably somewhat overestimated. Assumptions were also employed in estimating the amount of emission reduction that could be achieved from the leak detection and repair program.

8. Difficulties in the estimation of costs further reduced our rating. No representative empirical data supported the analysis of the costs of technological control. Frequently, estimates for individual pieces of equipment were based on a single telephone call to a local distributor. How EPA derived other estimates was unclear. For example, a price for pipe of a specific diameter was reported but the document EPA cited as a source does not list the same size pipe. Assumptions of the life of the equipment in years, annual interest rates of 10 percent, overhead of 40 percent, and so forth were based on telephone calls, internal memos, and personal communications.

Net annual costs were calculated by subtracting from cost figures the estimated value of the volatile organic compounds, based on 1980 natural gas prices, that would not be emitted because of the regulation, called "recovery credits." By the time the regulation was promulgated in 1985, the 1980 prices were out of date and low, thus underestimating the recovery credits and overestimating the net control costs. But EPA did not refigure its estimates, and it provided no data to support its opinion that the higher gas prices would not have supported more stringent controls because they would be offset by higher control costs.

The incremental or marginal cost effectiveness of each specific control technique contained in each risk management option was used in decisionmaking, but these cost effectiveness data were presented only for one of the model plants in an appendix in the background information

document for the proposal. The primary discussion of control costs presented only the average cost effectiveness of the risk management options, which obscured the role of incremental costs in decisionmaking.

Moreover, as we noted above, inadequate sampling, measurement, assumptions, and other problems of this kind resulted in cost effectiveness estimates with a precision of between plus or minus 30 and plus or minus 100 percent. Since it is not possible to quantify the uncertainty of estimates derived by these inadequate methods, it is not possible to determine the probability that the estimates are actually higher or lower than any cost effectiveness cut-off. However, EPA used the cost effectiveness estimates in conjunction with established cut-offs, even though they were not precise enough for this use.

Finally, EPA's analysis of control costs in relation to industry and the economy was generally strong, but the rating we assigned was reduced because the effect of the options on industry decisions to delay the construction of new plants was not directly discussed. EPA officials stated that this is often one of the most important effects because it bears on the continued operation of older, less efficient facilities that are not subject to the performance standards. The preamble to the proposed standard, however, stated that such effects had been considered and found negligible.

Our rating in table 4.1 shows that documentation and reporting were inadequate. References in the background information document for proposal were frequently incorrect or missing. For example, no references were cited in the discussion of the model-plants procedure that would have provided a more detailed explanation. The reference to the major report from which the baseline emission factor data had been taken was not correct. In one instance, a statement referred to a diagram that was not included. Finally, computational formulas were not fully reported, making it difficult or impossible to determine how some source documents were used to derive the figures that were reported.

## Regulatory Decisionmaking

### Strengths

As our rating in table 4.2 indicates, the external regulatory review was more than adequate. The assistant administrator's review package was

sent to a standard list of other agencies, including the Consumer Product Safety Commission, OSHA, and OMB. OMB is the only agency that makes comments routinely. EPA documented OMB comments in outline form when meetings were held, but, since OMB conveyed many of its comments by telephone, no record of much of the interaction between OMB and EPA exists. EPA made only minor changes to wording and definitions in response to OMB's comments, although OMB consistently challenged the cost effectiveness of the proposed requirements. The lead engineers on the project stated that OMB's substantive comments were not adequately supported by empirical data. OMB's comments were based on its own benefit-cost analysis, which used an estimate of approximately $200 per ton, according to the EPA project's lead engineers.

**Table 4.2: Criteria Ratings for EPA's Regulatory Decisionmaking**

| Risk analysis component | Rating |
|---|---|
| 1. Administrative | |
| a. Formal guidelines | 0 |
| b. Compliance with legislative authority | 1.0 |
| c. External regulatory review | 4.7 |
| d. Administrative review | 4.3 |
| 2. Technical and scientific | |
| a. Decisionmaking procedures | 2.2 |
| b. Documentation and reporting | 1.8 |

Our rating for the administrative review was also high. The three rounds of administrative review that are required in EPA's guidelines were followed and documented.

We rated EPA's decisionmaking procedures low, although we gave high ratings on two of our criteria. The comparative review of similar risk management decisions was strong. For example, the lead engineers on the project stated that the similarity of the technological controls to those required in related industries was one reason why the OMB review was not more extensive. The public comment period was also extended to permit the submission of new data.

## Areas for Improvement

EPA's written guidance for decisionmaking covers only the preparation, timing, and administrative review of standards. As is indicated by our rating, when the agency promulgated the onshore standard, it had no formal guidelines for deciding what constitutes reasonable control costs. As we noted at the beginning of this chapter, EPA was in conflict over

whether decisions on standards should be based on technology or on balancing. Some believed that the technological basis is not logically coherent and that standards should be based firmly on balancing control costs against the benefits of reduced risk—that is, on benefit-cost analysis.

The formal guidelines EPA normally follows are referred to as "cost effectiveness guidelines," "criteria," "thresholds," or "cut-offs." They represent the maximum cost per unit of reduced pollutant that is deemed reasonable—that is, the best demonstrated technology. In 1985, following lengthy debate within EPA, the deputy administrator ruled that the cut-off number should be based primarily on technology and instructed EPA officials to reach consensus on the specific figures. The memo expressing the response gives a criterion for ozone of $1,250 per metric ton, which may be exceeded under specified conditions. However, the consensus was not achieved until September 1985, nearly 3 months after the final onshore standard was issued, and, thus, only informal guidelines could have been in effect. This agreement apparently did not end the conflict. In August 1986, EPA officials were again directed to reach consensus concerning decisionmaking guidelines for new source performance standards for volatile organic compounds but had not yet done so by August 1987.

In the absence of formal guidelines, we attempted to determine how a "reasonable control cost" had been decided, but different EPA officials gave us different explanations. The lead engineers for the project said it was based on benefits analysis; the incremental or marginal costs of specific control techniques were compared to the threshold or cut-off value, which was based on an analysis of the damage caused by ozone. They also stated that before 1981, cut-off numbers were based on the principle of affordability derived from the analysis of economic effects—that is, control should be as stringent as could be implemented without significant economic effect. The chief of the economics analysis branch in EPA's office of air quality planning and standards told us that the statements of the engineers were incorrect and that the decision had been based on technology, indicating that affordability, engineering judgment, and a consideration of the level at which significant resistance to the "reasonable control cost" would begin were factors in decisionmaking. However, rather than clearly stating any technology-based principles, the branch chief pointed out several serious problems with the use of benefits-based cut-off numbers and stated that while cost effectiveness cut-offs are one input, decisions are not made on the basis of a single criterion.

Analysts we interviewed from EPA's office of policy, planning, and evaluation believe there are no coherent guidelines based on technology and that decisionmaking would be rational if it were based firmly on benefits analysis. They stated that personnel in the office of air quality planning and standards "get a warm feeling in their stomachs" when they identify the best available technology. Otherwise, technology-based decisions are, in the opinion of policy, planning, and evaluation officials, nonrational and influenced by political considerations. One of their internal documents stated that while the best demonstrated technology is perceived as reflecting the limit of affordability, it conceals numerous tradeoffs that are not based on engineering considerations or technical feasibility.

The office of air quality planning and standards has used a "traditional" figure of $2,000 per ton of reduced pollutant as a "rule of thumb" cut-off for reasonableness for all pollutants, not merely ozone. Although this informal guideline is not based on benefits, it is not clear what it is based on. One EPA official stated that the number emerged historically; another said that a cost of more than $2,000 per ton makes engineers feel uncomfortable. All six personnel we interviewed on this subject agreed that no written documents set forth and defend this figure.

The estimated marginal cost of the most expensive technology required for the standard we evaluated was lower than $2,000 per ton. The action memo for the proposed standard expressed a figure of $1,300 per ton; the action memo for the promulgated standard indicated that the figure was $1,500 per ton. Ten days before the standard was made final, a memo from the office of air and radiation opposed the $1,000 per ton cut-off favored by the office of policy, planning, and evaluation, on the grounds that it would jeopardize standards for ozone that require technologies with marginal costs of $1,500 per ton. Officials in the office of policy, planning, and evaluation stressed that benefits analysis influenced the final decision but did not determine it, because the final $1,500 figure was too high, adding that their benefits analysis would put the figure between $400 and $800 per ton.

Our discussion has indicated that both balancing principles and technological control principles influenced decisionmaking for the onshore standard. It has also indicated that the technological control basis for decisionmaking has not been clearly articulated. We found no evidence, however, that it would be impossible to prepare coherent and analytically defensible guidelines based primarily on technology. According to some officials we interviewed, EPA has used a principle of affordability

to determine reasonable costs; the lead engineers said that worst-case cost analysis was used prior to 1981, to ensure that any errors were in the direction of overestimating negative economic effects. It is possible that this approach could be defended as analytically sound as long as what constitutes an unreasonable economic impact was specified.

On our criteria for compliance with legislative authority, we rated EPA's statements as inadequate. The appropriate legislation, section 111 of the Clean Air Act, was cited, but the discussion of EPA's interpretation and implementation of the statutory authority was inadequate. The standardized "boilerplate" descriptions included in EPA's publications contain discussions of cost considerations that we found ambiguous. They do not state how costs are considered. The Federal Register publications stated only that EPA had made a determination about the reasonableness of control costs and that cost had been "carefully considered." One change in the final standard was the result of information provided during the comment period, in which industry argued that equipment costs would be higher than EPA's estimates. The background information document for promulgation states simply that the equipment costs had been reassessed and found unreasonable, but the lead engineers stated that the reason for the change was that the new estimate exceeded the cutoff. The chief of the economics analysis branch stated that the office of air quality planning and standards would in general consider who the author of the comment was as well as what it said before determining that the new estimate indicated unreasonable costs. No document we reviewed in the onshore natural-gas processing plants case states how cost was considered in order to determine reasonableness. This means that the public was unaware that any doubt existed about the basis of decisions for the standard.

Furthermore, EPA's failure to fully describe its interpretation of section 111 is a significant deficiency, because the actual basis of its decision was different from that suggested in the documents presented to the public. The Federal Register notices for both the proposed and the final standards state that the average price increase caused by the standard was expected to be less than .01 percent. In the preamble to the proposed standard, EPA stated that "no adverse economic impacts are anticipated." No other cost considerations were discussed. These statements indicate that achievability or affordability is the criterion for determining the best demonstrated technology. However, the previous discussion has shown that in this case, the best demonstrated technology was determined, at least in part, on the basis of explicit cost-effectiveness thresholds and benefit-cost analysis.

Although the administrative review for decisionmaking was generally strong, the guidance for it concerned only its timing and completion, and none of the decisionmaking issues we have discussed were presented for formal administrative review.

Aside from the reasons indicated above, we gave EPA's decisionmaking procedures a low rating because they were not adequately based on the results of the work on the development of risk management options. The decisionmaking was influenced by benefits analysis that was based on risk assessment work but was not included in the documentation of the technical basis of the standard. Because EPA's benefits analysis did not have an official relationship to its decisionmaking, we did not apply our criteria here. However, in our 1984 report entitled Cost-Benefit Analyses Can Be Useful in Assessing Environmental Regulations, Despite Limitations, we found that benefits analyses were useful and should be pursued but lacked sufficient precision to support decisionmaking. Statements made by several EPA officials we interviewed indicate that inadequate precision has remained a problem. The benefits estimates reported to us ranged from $200 per ton to $1,400 per ton. The differences are the result of differing judgments on how the available and relevant research should be interpreted.

Finally, EPA's decisionmaking was not adequately based on the development of risk management options inasmuch as evidence is lacking that the analysis of economic effects was used. As we noted above, the economic "impact" analysis concluded that none of the regulatory alternatives that were examined would have a significant economic effect, yet several of the technologies were deemed unreasonable. Since the absence of a significant economic effect is not sufficient to determine reasonableness, it is difficult to avoid the conclusion that the cost effectiveness cut-off levels were the deciding factor.

As for documentation and reporting, table 4.2 shows that the rating we assigned was low. The documents we reviewed contained little discussion of uncertainty. The action memo, preamble, and background information document for the final standard contain no discussion of empirical uncertainties. Nonempirical uncertainties, such as industry objections, dominate the information reviewed by the administrator. As we discussed above, we could find no documentation concerning how cost was considered in decisionmaking. The formal statements of the decisionmaking process were ambiguous and misleading. It was not possible to determine from the written record how the final decisions were actually made.

## Monitoring and Evaluation

The Clean Air Act requires EPA to review new source performance standards every 4 years. The agency's policy for conducting such reviews is to treat a revised standard as similar to a new rulemaking activity for which most procedures applicable to an initial standard are to be followed. Since the standard we examined is not due for evaluation until 1989, we did not rate it for monitoring and evaluation. And, since we evaluated only one standard, we did not attempt to determine whether the reviews of standards are generally timely, although the lead engineers indicated that delays have occurred.

## Sources of Supporting Research

EPA relies primarily on independent contractors for the technical research necessary to support its standards, although adequate research from industry and other sources is used, if it is available. Table 4.3 shows by stage of the research process the sponsorship of the major research studies that EPA cited. As can be seen, the primary source of data for the technical support for the standard was EPA-sponsored research.

**Table 4.3: Studies EPA Used in the Volatile Organic Compounds Case by Sponsoring Source**

| | Sponsoring source | | |
|---|---|---|---|
| **Type of research** | **EPA or a contractor** | **Other government agency** | **Industry or contractor** |
| Source characterization | 1 | 1 | 0 |
| Emission factor estimation | 4 | 1 | 1 |
| Control technique identification and abatement effectiveness | 5 | 0 | |
| Model-plants development | 1 | 0 | |
| Overall emission reduction | 1 | 0 | 0 |
| Control technique cost estimation | 3 | 0 | 0 |
| Economic impact assessment | 0 | 4 | 0 |
| **Total[a]** | **7** | **5** | **2** |

[a]Not all columns add up to the totals, because some documents were used in more than one stage of research.

## Summary and Conclusions

Our evaluation of the 1985 onshore standard suggests that EPA's greatest strength was in the administrative aspects of the development of risk management options. We gave high ratings to the guidelines, formal administrative review, and external expert review. However, even these contained significant areas for improvement. The guidelines did not

require probability sampling; the external expert review panel, composed primarily of industry representatives, was not balanced adequately; and the administrative review emphasized form rather than substance.

The research for the development of risk management options was also generally adequate, but we did note areas for improvement. The major problems we identified were lapses of systematic inquiry; uneven documentation of sources, which often made it difficult to determine how numbers had been derived; failure to include all computational formulas; frequent use of assumptions rather than empirical estimates; failure to use probability sampling and other scientific standards to ensure reasonable precision of empirical estimates; and failure to provide adequate quantification and discussion of uncertainty. The difficulties were compounded with each step as additional parameters with different sources of error were used to arrive at the final estimate of cost effectiveness.

Most of the problems in the development of risk management options were empirical. Some of these may have been technological, as in the inability of the handheld organic-vapor analyzer to record concentrations of ambient volatile organic compounds greater than 100,000 ppm. Other problems were related to research costs, but probability samples could nonetheless have been drawn in order to permit the quantification of sampling error for the industry in general, and gas chromatography could have been used more extensively to minimize measurement error. OMB's requirements limited the use of the regulatory letters and telephone surveys mandated in section 114 of the Clean Air Act to nonprobability samples. Finally, assumptions were used frequently but were supported with inadequate rationales. Although the use of assumptions is often necessary when empirical data are lacking, untested assumptions introduce a degree of unquantifiable uncertainty.

Documentation and reporting were inadequate. EPA's failure to accurately cite all support documentation and to report all computational formulas means that other researchers cannot determine how EPA derived the figures it reported.

Except for the external regulatory review, which was more than adequate, the decisionmaking process was inadequate. When the standard was being prepared, EPA had no consistent policy for determining reasonable control costs. The decisionmaking principles that were followed were not documented. Whether a standard is to be based on benefits or technology, it is reasonable to expect EPA to prepare and follow a clear

consistent, and accurate statement of its decision rules and procedures. Moreover, the expectation that the decision rules, together with a rationale for them that EPA is prepared to publicly acknowledge and analytically defend, should be clearly documented and included in the public record is also reasonable. If there is legitimate doubt concerning the appropriate basis of a decision, it should be made available for public debate. We also believe that principles for determining reasonable costs that are consistent with the technological control approach to risk management could and should be articulated and defended.

Finally, our evaluation suggests that EPA's cost effectiveness estimates were not precise enough for use in conjunction with thresholds or cut-offs, whether they were based on benefits or some other criterion. The low precision and nonstatistical character of EPA's cost effectiveness estimates in this case means that EPA could not always satisfactorily determine whether the estimated control costs were higher or lower than the cut-off.

According to officials in the office of policy, planning, and evaluation, benefits-based decision rules are preferred because they improve the logical coherence and rationality of decisionmaking. Benefit-cost analysis may be conceptually "cleaner" and admit of a more easily defended logical rationale, but it may also lack sufficient empirical precision to avoid increasing the level of ambiguity beyond what we found in this case. The substantial uncertainties inherent in the current benefits estimates actually may further interfere with rational decisionmaking. Under these conditions, the policy debate over the appropriate level of control easily becomes submerged in the ostensible scientific interpretation of the input studies. The result of benefits-based decision rules at this time could well be to further obscure the decisionmaking process.

# Observations Across the Cases, Agency Comments, and Our Response

Because we analyzed only one case from each of three agencies, we cannot generalize about the adequacy of the federal risk analysis process in those agencies. In a few instances, our information about standard agency practices allows some generalization but, for the most part, our statements in this chapter are confined to observations about how risk analysis was conducted for each specific case and to concerns the agencies raised about our statements in this report.

## The Appropriateness of Our Risk Analysis Model

Our model describes what we believe is a generic risk analysis process. In general, it has proved to be an accurate representation of the logical steps of the risk analysis process. However, because it is generic, merely outlining the general logic of the process, the model ignores many factors that impinge upon the routine activities of regulatory agencies. According to logic, technological control controls a substance to the extent feasible, once it has been identified as hazardous. When a statute requires or is interpreted to permit the technological control approach to risk management, large portions of risk assessment indicated in the model may not be required. Our model is not concerned with most of the regulatory history of individual substances. For example, inorganic arsenic had been identified as a hazard well before OSHA began its action, and ozone had been identified as a hazard in the nineteenth century. Consequently, hazard identification was not required of either OSHA or EPA.

Our model is not intended to describe precisely how risk analysis is always implemented. Implementation varies because of differences in organizational structure, authorizing statutes, and their legal interpretations. The logical distinctions and sequential order in the model were not fully reproduced in any of the cases we examined.

For example, FDA combined dose-response assessment, exposure assessment, and risk characterization under the term "quantitative risk assessment." At OSHA, the implementation of the risk analysis process was deeply embedded in how administrative standards were set. Most of the technical risk assessment and options development work was conducted in the "action recommendation" step of the regulation management system. Within this step, OSHA used terms different from ours, such as "regulatory analysis," with which OSHA refers to the development and evaluation of risk management options. At EPA, the implementation of risk analysis is even more complex. The office of air quality planning and standards conducted only risk management activities for

the development of new source performance standards, the hazard identification and risk assessment work having already been performed by other offices pursuant to different sections of the Clean Air Act for somewhat different regulatory purposes. The development and evaluation of risk management options, which includes the source characterization portions of exposure assessment, constituted the major part of EPA's technical research, so that in the EPA case we considered portions of the exposure assessment phase of risk assessment under risk management.

The latter was the only adjustment to the structure of our model required in order to accurately reflect the work performed by the agencies. Despite the many differences and complexities in implementation, we believe it remains informative to retain the distinctions made in the model in terms of the inherent logic of the risk analysis process. It seems to us appropriate to use the model to evaluate the process across agencies.

## Risk Assessment

At OSHA and FDA, the technical and scientific risk assessment work was generally conducted well. The agencies did a credible job of reviewing the scientific evidence to determine levels of risk. The hazard identification, dose-response, and exposure assessment tasks were conducted in general accordance with current standards for risk assessments. The risk characterizations derived from the findings of these procedures were also well done. This is important, because it indicates that the basic expectations for sound scientific performance were met despite resource constraints.

In both cases, however, we found two significant problems. The first is that in some instances the limitations of equipment or measurement techniques did not allow the collection of necessary information. This occurred at OSHA, where accurate measurements of individual exposure and personal habits were not generally available for the type of analyses OSHA used. At FDA, the extent to which the animal bioassays it relied on can accurately predict human responses is not known.

The other, more pervasive problem was high research cost. At FDA, standard rodent studies cost about $500,000, excluding administrative costs, and even these may be constrained to test fewer than the optimal number of dosage levels. The epidemiological studies particularly relied on by OSHA are costly and their time requirements often exceed the deadlines that are typical in the regulatory process.

We identified problems in these cases in four other areas. First, FDA selected substances for study not according to a priority-setting system but on the grounds of available information, interrelated regulatory requirements, and actions taken by other agencies. While these reasons are valid, they are not sufficient. (The selection of emission sources to be regulated by EPA is the result of a priority-setting system mandated by the Clean Air Act.)

Second, the guidelines for risk assessment were mixed in quality. FDA officials believe that guidelines do not permit the flexibility necessary for coping with the varied circumstances under which risk analyses must be conducted, and in the methylene chloride case, they did not follow guidelines on how to conduct any risk assessment work. While OSHA's guidelines were generally adequate, the agency had no guidelines for assessing the dose-response relationship from epidemiological data. We believe this may result in a lack of consistency in how work is conducted from one risk analysis to the next.

Third, we found problems in the exposure assessment in both the OSHA and FDA cases. The epidemiological studies used by OSHA lacked adequate measures of the ambient concentrations of inorganic arsenic during the years workers were exposed to it. FDA was not able to adequately estimate actual exposure to methylene chloride.

Finally, when state-of-the-art limitations or cost considerations made specific data unavailable, both agencies resorted to untested assumptions. The use of assumptions may appear reasonable but it also introduces unquantifiable uncertainty and may lead to misleading results.

## Risk Management

As stated above, the development and evaluation of risk management options, which includes the source characterization portions of exposure assessment, constituted the major part of EPA's technical research, so that in the EPA case we considered portions of the exposure assessment phase of risk assessment under risk management. Some of the same problems we found in the risk assessment work in the FDA and OSHA cases occurred in this portion of the work for the EPA case. The limitations of equipment or measurement techniques did not allow the collection of important information. This occurred in EPA's source characterization work, in the use of the handheld organic-vapor analyzers. EPA also experienced the problem of high research cost. It affected

the number of plants EPA examined and the kind of sampling it performed. A larger and more rigorous study was judged prohibitively expensive. Finally, when state-of-the-art limitations or cost considerations made specific data unavailable, EPA also resorted to untested assumptions, which introduces unquantifiable uncertainty and may lead to misleading results.

While we found the risk assessment work generally adequate, we found that the risk management work at all three agencies had serious difficulties. The most widespread problem was poor documentation. In both the FDA and EPA cases, the decisionmaking process was poorly documented. It was not always clear whether or not regulatory options were developed, how they were developed, how full the consideration of other options was, or in what form the options were presented to the decisionmakers. It was also unclear how, or if, the uncertainties associated with the risk characterization or with the different options under consideration were presented to decisionmakers. Consequently, we do not know whether the information on which the decisions were based was complete.

The reviews by decisionmakers were also inadequately documented in the FDA and EPA cases. It is unclear what options were reviewed, what steps were taken, who was involved in them, and what the bases were for the final regulatory decisions. Thus we do not know what factors were considered or how the factors that were considered affected the final decisions. These crucial policymaking items were not documented. For example, cut-off levels based partly on benefit analyses were used in the decisionmaking at EPA but not documented.

Because of the poor documentation of the risk management, we could not examine in detail several issues we raised about the use of information from the risk assessment phases in the risk management phases. It was not clear how, or if, uncertainty arising in risk assessment was dealt with in risk management, and we could not tell how the results of risk assessment were used in any analysis of risk management options.

A further problem was especially apparent in the EPA case. The precision of the cost effectiveness estimates for the options considered did not appear to be sufficient to permit their use in conjunction with cost effectiveness thresholds. This may also be a problem for FDA's use of a de minimis risk threshold.

The extent and quality of the guidelines for risk management varied greatly, between and within the agencies. EPA had extensive guidelines for the development and evaluation of regulatory options, but they placed more emphasis on the form than the substance of the process. EPA had no guidelines for decisionmaking. In contrast, OSHA had strong guidelines for decisionmaking but no guidelines for developing and evaluating risk management options. FDA had no guidelines for either options development or decisionmaking. Guidelines for decisionmaking would seem to be particularly important, given that the agencies did not clearly document how the various factors were considered and united in decisionmaking for these three cases.

No follow-up evaluation has been conducted for any of the three cases. The Clean Air Act requires EPA to perform a full regulatory review every 4 years after the promulgation of a new source performance standard, a date not yet reached for this standard. EPA officials state that such reviews have, at times, been delayed. The guidelines treat the revision of standards as similar to original rulemaking efforts and require many of the same procedures. However, its evaluations are limited to determining whether the best demonstrated technology required is successful in reducing emissions as much as anticipated.

FDA and OSHA monitor compliance, and FDA collects reports on adverse reactions, but neither agency routinely evaluates the effectiveness of regulatory actions for reducing the types of risks we examined. Conducting evaluations to determine the effectiveness of risk reduction is complicated by cost and questions of technical feasibility. Because exposure to food-additive substances is low, measuring any adverse effect on health is beyond the ability of epidemiology. Two or three fatalities beyond the number expected per year, the number expected by a 1-in-1-million lifetime risk if the entire United States population were at risk, are not detectable, especially when many other sources pose the same risk of increased mortality. Technically, the reduction of mortality from some health risks could be assessed for at least some regulations, since some occupational exposures are high enough to cause effects detectable by epidemiology. But the cost of such studies would be high. Therefore, it is difficult or impossible to determine the effects of regulating for risk reduction in these areas.

A final area for discussion is the extent to which all the risk management activities appeared to contain elements of balancing. That is, there seems to have been some collapsing of the risk-only and technological control approaches into risk balancing. For example, a consideration of

the economic effect of a lower standard for methylene chloride residues in decaffeinated coffee seems to have influenced decisionmaking in the risk-only case at FDA. The Clean Air Act requires the technological control approach to risk management. Although it expressly allows the administrator to take cost in consideration when setting new source performance standards, some officials at EPA believe that standards should be based more stringently on benefit-cost analysis. While it might be argued that the risk-balancing approach should be preferred, our point is that some of the current laws still call for other risk management strategies.

## Agency Comments and Our Response

The Department of Health and Human Services and the Department of Labor indicated general support for our effort to evaluate the risk analysis process and help improve regulatory decisionmaking. However, all three agencies expressed a number of general concerns. The first, expressed by all the agencies, was that the cases we selected were inappropriate and not representative of the agencies' risk analysis efforts. A second general concern was that the report inappropriately holds the agencies accountable for conditions beyond their control, such as the quality of research they did not conduct or had performed under contract. A third concern was that our approach confuses the quality of the scientific research with the administrative process. The agencies commented also that our report does not sufficiently recognize the many factors that must be considered in exercising decisionmaking judgment. And a final, general comment concerns the validity and applicability of the criteria we developed for this study. We believe these agencies' concerns reflect a lack of understanding of our intent in performing this study. Each of the general comments is discussed in greater detail below.

### Case Selection

Each specific risk analysis case is unique in certain respects, and by itself it is not completely representative of an agency's risk analysis effort or of the risk analysis process as a whole. This is so because any specific case may be too old, may occur within a complex legislative context that in some sense is not standard, may employ an approach to risk management that does not require all possible aspects of the risk analysis process, or may be too new and unusual to be representative. For these reasons, we have emphasized in the report that no generalizations could be made except in limited areas concerning some general procedures.

The purposes of the report were to (1) provide some preliminary information on possible weaknesses and areas of strength in the federal risk analysis process that might be explored later in more depth and (2) refine our evaluation criteria and methodology. To achieve these purposes, we elected to apply our criteria to three specific cases that would provide the opportunity to review each major type of risk management strategy that has been enacted in law. We defined a risk analysis case in terms of the work required to support a specific agency action. In order to implement this approach, we provided each agency with an explanation of our study design and a list of actions published in the Federal Register that we thought represented the relevant risk management approach. The agencies validated the accuracy of our lists and their relevance to the specified risk management approach. All the concerns they expressed at that time with our design were resolved before we finally selected our cases. Once the lists were validated, a case was randomly selected from each list.

## Agency Accountability

We recognize that regulatory agencies must rely on research and other information that is beyond their control. The scientific research undertaken and the decisions our society makes to cope with risks are not areas in which one could legitimately conclude that all existing problems are the result of a regulatory agency's action. But this does not make the problems any less troublesome. Consequently, our approach was not to rate each agency in terms of its implementation of the risk analysis process. Instead, our approach was to examine and rate cases of the risk analysis process as a whole, which includes work the agencies conducted as well as work the agencies used that was conducted by others. Some of the problems we observed, if they occur consistently, might be addressed by the agencies; some might require the modification of institutional arrangements; others may prove intractable. While we certainly discuss and rate aspects of the risk analysis process that are beyond the control of the agencies, we have not held the agencies accountable for them. Moreover, since for the most part we cannot conclude that the problems we observed in these limited cases are representative, we did not attempt to determine how they might be addressed.

## Science Versus Process

In order to evaluate the risk analysis process as a whole, we believe it is essential to examine two analytically independent dimensions. The first may be referred to as the scientific and technical dimension, or the extent to which accepted scientific methods and procedures are employed. The second may be referred to as the administrative and

managerial dimension, or the extent to which accepted administrative practices are employed. These dimensions of adequacy are not identical to the risk assessment and risk management phases of risk analysis; they cut across both phases of any actual case of the risk analysis process.

While the scientific and administrative dimensions are analytically distinct, in practice they overlap. One reason for the concern the agencies expressed seems to be their view that process questions—such as whether research activities are sufficiently documented, computational formulas are presented, and so forth—are separate from questions about the quality of the research. In our view, process questions are important because without them the quality of the research remains unknown. Moreover, the documentary record is the prescribed means for demonstrating the quality of the research. In this sense, the two dimensions are inseparably linked.

A second reason for the agencies' concern that we confuse science and process seems to be their view that we criticize the use of less-than-the-best research techniques, even though we acknowledge that their use is often justified because of costs or other considerations related to process. We have stated above that we do not necessarily criticize the agencies in this respect. Our ratings are based on observed problems in three specific cases of the general risk analysis process. In our view, if we had allowed cost or other process factors to influence our evaluation of the quality of the scientific research, we would have been confusing science and process. Additionally, in order to emphasize the constraints under which the agencies operate, we discuss the institutional context in which agencies must make such decisions. Our purpose is to avoid the implication that the agencies are always free to use perfect research techniques.

## Factors Considered in Decisionmaking

A considerable portion of our report is devoted to describing the factors the agencies consider in decisionmaking. We also recognize that the authorizing statutes permit discretion in decisionmaking and that guidance in these matters has been provided by judicial review. However, as we state in several places, our view is that the agencies, in these specific cases, did not always clearly articulate the factors they considered, how they considered them, or how their consideration of the several factors was integrated in decisionmaking. If the decision rules according to which a judgment is exercised are not or cannot be articulated clearly, legitimate doubt will remain about the appropriateness of the judgment.

## Validity and Applicability of the Criteria

We have stated in this report that our procedures were to (1) derive criteria from guidelines proposed in the literature for conducting adequate risk analyses and (2) review them with expert consultants. The reason we adopted this approach is that we felt it would not be appropriate for us to rely solely on our own opinion of what constitutes an adequate risk analysis process. This means that we were limited to some extent by the literature. For example, because our criteria reflect the literature, they do not address every type of research the agencies may use to assess some types of risk sources, and some of our criteria presuppose that the risk source under consideration is a suspected carcinogen. More specific criteria would vary with the specific types of research required. As a result, we explicitly do not argue that the criteria we developed are universal, that they are necessarily exhaustive, or that all our criteria are required for an adequate risk analysis. Moreover, our practice in instances in which specific criteria were inapplicable was to give no rating. However, we do believe our general model and many of our criteria are applicable to noncarcinogens such as criteria air pollutants and prescription drugs.

In addition to these general comments, each agency provided more specific comments that were helpful in revising the report. The agencies' formal comments are printed in appendixes VI-VIII, along with our responses to each point. In addition to EPA's formal comments, EPA provided two sets of informal comments. We have not reproduced these in appendix VIII, but they were valuable in revising the report.

# The Risk Analysis Process

Risk analysis is the overall process through which a risk is identified, estimated, and evaluated with some resulting decision as to whether to reduce the level of risk. This can include a decision to take no action. Some experts in the area have suggested that a formal definition of risk analysis would be based upon the development of quantitative, probabilistic risk estimates for some sources through epidemiological studies, laboratory tests, and the like. Once the probabilistic risk estimate had been developed, it would be applied to a population at risk, or the population exposed to the hazard, and the number of expected insults to health would be calculated. Under this formal definition, only when a sophisticated analysis has been conducted is a risk analysis at hand. Some experts believe that risk analysis is evident when any data or information on a risk is used within the context of decisionmaking that results in the aversion of risk. In this definition, the data or information can be qualitative or quantitative and the decisionmaking informal or "automatic," as when certain physical criteria, if met, result in the automatic imposition of regulatory controls.

Our definition, while not the broadest one possible, is that risk analysis happens when the use of both quantitative and qualitative data results in a decision on whether some action to reduce the risk associated with a particular hazard is necessary. We decided that for our evaluation, if the decisionmaking apparatus is not invoked (regardless of the regulatory outcome), then a risk analysis has not been conducted. Thus, by our definition, if regulation is "automatic" when specific criteria are met, a risk analysis has not been conducted.

## The Phases of Risk Analysis

The term "risk analysis" for regulatory actions encompasses both risk assessment and risk management. The phases of risk assessment are (1) hazard identification, (2) dose-response assessment, (3) exposure assessment, and (4) risk characterization. The phases of risk management are (5) the development and evaluation of risk management options, (6) regulatory decisionmaking, and (7) monitoring and evaluation. The seven steps in these two phases constitute risk analysis in our model.

In hazard identification, the risk source to be analyzed is decided. Such decisions are often based on clinical and field observations of ill health resulting from exposure to risk sources. Examples are epidemiological studies, short-term and long-term bioassays, and comparisons of a substance's molecular structure with known carcinogens.

Dose-response assessment estimates the magnitude of the risk associated with the hazard, often in probabilistic terms. Dose-response assessment also relies on epidemiological data, bioassays, and interspecies extrapolations, in which human responses to a substance are extrapolated from responses observed in studies of animals. Since most studies involve relatively high doses, extrapolations must be performed to estimate risks at lower doses.

Exposure assessment characterizes the sources of exposure, the routes and concentrations of exposure, the level of exposure for different population groups, and sometimes exposure under different possible regulatory controls.

In risk characterization, the information from hazard identification, dose-response assessments, and exposure assessment is brought together to describe the risk to public health and its magnitude. Uncertainties as well as groups with different exposures or special sensitivities are considered and weighed. This information, in turn, is fed into the risk management process.

Risk management begins with the development and evaluation of options for controlling the risk, which depend largely on the legislation pertaining to the substance identified as a source of risk. Regulatory decisionmaking ends in the decision of whether to regulate the risk source and under what option. Once a final regulation has been issued, risk monitoring and evaluation help ensure that the regulation achieves its objectives. The focus of this phase is not enforcement but, rather, an evaluation of the postregulatory state of affairs with respect to the risk.

## Three Risk Management Approaches

The three risk management approaches that are most generally used are (1) risk only, ( 2) risk balancing, and ( 3) technological control. A fourth approach, marketplace management of risks, is not frequently employed by the federal government, and we excluded it from this report. The type of risk management approach used is dictated by the type of hazard being evaluated and the agency's legislative authority.

### Risk Only

Risk-only management considers only the level of risk: the source is to be controlled if the level exceeds one that is deemed acceptable. A clear example is the legislatively mandated control of food additives under section 409(c)(3)(A) of the Federal Food, Drug, and Cosmetic Act, which prohibits the use of carcinogenic food additives. This example is an

extreme type of risk-only approach—it seeks "zero risk." Some other risk-only regulations more flexibly allow the existence of a risk source rather than simply prohibiting it.

## Risk Balancing

Risk balancing considers not only risk but also economic or social effects or the like. It is the most commonly used risk management approach. For example, the Occupational Safety and Health Act of 1970 requires that risk analyses balance both the risks and the costs of controlling hazards. For risk analyses conducted under this act's authority, technologies used to control a level of risk fall into the category of "best practical" technology, the term "practical" implying that other, balancing factors were considered before deciding on a control approach.

One technique that is used in risk balancing is benefit-cost analysis, in which one weighs the costs of control, explicitly and directly, against monetized benefits such as the avoidance of disease, reduction of soiling and damage, and other social goods. When benefit-cost analysis is not appropriate, other related techniques are used, such as risk-benefit analysis, which evaluates health hazards and compares them to their benefits, such as the usefulness of the hazardous substance in a given circumstance. Risk-risk analysis compares the risks of different technological alternatives for accomplishing a given objective in order to determine the alternative with the lowest risk. Cost-effectiveness analysis looks for the least-cost path to achieving a particular control action.

## Technological Control

In technological control, the decisions about whether and how to deal with human exposure to a hazardous substance are focused on the opportunity for the use of technology rather than on other means, such as banning use of the substance. What makes this approach different from others that apply technologies to control exposure is that the emphasis, frequently a legal requirement, is often on requiring the "best available technology" to reduce the exposure down to a targeted level. Risk management in these circumstances includes determining what technologies are "available" and determining which among those that are available is "best," in terms of controlling hazardous chemical emissions, for example. Certain sections of the Clean Air Act exemplify federal legislation requiring technological control in managing risk.

# Objectives, Scope, and Methodology

## Objectives

This report is a product of a pilot study for a possible larger evaluation of the adequacy of the risk analysis work supporting federal health regulation. Our purpose is to provide some preliminary information on possible weaknesses and areas of strength. To describe and evaluate the risk analysis work being conducted by the federal agencies that have primary responsibility for managing health hazards, we formulated 19 evaluation questions covering risk analysis policies and practices.

For risk assessment, we asked questions 1-5:

1. What method is used for reviewing potential risks?

2. How sound is this method?

3. How justified and appropriate is the decision that a risk warrants a complete analysis?

4. What studies are done in the phases of the risk assessment process?

5. How persuasive and methodologically sound are these studies?

For risk management, we asked questions 6-19:

6. What review is performed to support the decision that a risk is excessive?

7. How complete and systematic is this review?

8. How are risk management options developed?

9. How adequate is this development process?

10. Are the options evaluated in a systematic way?

11. What data and information are accessible to the decisionmakers?

12. How complete and sound is this material?

13. Who reviews the risk analysis work?

14. How careful is this review?

15. How well supported is the decision to proceed or not proceed with a regulatory action?

16. Are evaluations being performed of the effectiveness of regulatory actions in reducing the risks they are intended to reduce?

17. How rigorous are these evaluations?

18. Who has access to them?

19. What use has been made of their findings?

## Scope

Each of the three cases we evaluated at FDA, OSHA, and EPA represented one each of the three major risk management approaches. The case from FDA was a risk-only rulemaking for a food additive regulated under the Federal Food, Drug, and Cosmetic Act and related to the Delaney clause in the act. At OSHA, we examined a risk balancing regulation under the Occupational Safety and Health Act of 1970. At EPA, we investigated a new source performance standard that employed a technological control approach to risk management under section 111 of the Clean Air Act. We conducted our field work from March through September 1986.

For the purposes of this project, a case is an event in which a hazard is or has been identified as a source of risk to public health and in which the risk analysis process is applied in order to reach a decision concerning how the hazard should be regulated. The universe of cases available to us consisted of events that occurred from 1981 through 1985 under the authority of the three acts named above and that were published in the Federal Register. We selected these statutes because they covered the three main risk management approaches. Under our definition, a hazard could have been identified and the risk assessment work for it could have been conducted prior to 1981, as long as the decision concerning its regulation was made between January 1981 and December 1985. Although we selected the cases randomly in order to mitigate bias problems, the sample size is not large enough to permit generalization.

We asked the agencies to verify lists of regulatory actions published in the Federal Register from 1981 through 1985, and we eliminated actions that, according to the agencies, did not fall under normal agency procedures. We did this so that the pilot cases would provide information about standard risk analyses at these agencies. We excluded actions taken under emergency procedures and actions carried out prior to

major changes in an agency's risk analysis procedures. From the remaining actions on the list for each agency, we selected one case at random. In the end, we selected the methylene chloride case at FDA from a list of 2 actions; the inorganic arsenic case at OSHA from a list of 2 actions; and the case on volatile organic compounds from onshore natural-gas processing plants at EPA from a list of 21 actions.

## Methodology

For our case studies, we developed criteria for evaluating the adequacy of the work in each step of the risk analysis process. First, we collected guidelines and general scientific information on risk analysis. The criteria we derived from this literature were consistently emphasized in it as important. Then we convened two panels of outside experts in risk analysis to critique the criteria, revising them after each panel's comments. Our final criteria for the case studies are in appendix III.

Our criteria for this pilot study do not perfectly reflect adequacy, and not all the criteria are necessary to fully determine adequacy. They simply represent the criteria that have been recommended in the fairly extensive risk analysis literature. We intend to revise the criteria before we implement a possible follow-on project and to base our revisions on our experience from the pilot cases.

We developed our criteria in a way that would provide detailed and comprehensive coverage of our 19 evaluation questions. We organized them by phases of the risk analysis process. The relationships between the evaluation questions, the phase of analysis, and the criteria are shown in table II.1. We answered the questions for which the criteria were descriptive from the background information we collected; we answered the questions for which the criteria were normative by assigning ratings to the quality of an agency's work in each of the seven phases of risk analysis.

We applied the evaluation criteria to the information we collected for the three cases. Some of the information was simple descriptive information, answering questions on the basic conditions of each case. Other descriptive information provided an understanding of the technical and organizational characteristics of the cases. The data sources we reviewed included the risk analysis case files, documents in the publications of the regulations, and technical reports.

**Table II.1: The Relationship of Our Evaluation Questions to Our Criteria**

| Question | Risk analysis phase | Criterion type |
|---|---|---|
| 1 | Hazard identification | Descriptive |
| 2 | Hazard identification | Normative |
| 3 | Hazard identification | Normative |
| 4 | Dose-response assessment, exposure assessment, risk characterization | Descriptive |
| 5 | Dose-response assessment, exposure assessment, risk characterization | Normative |
| 6 | Development of risk management options | Descriptive |
| 7 | Development of risk management options | Normative |
| 8 | Development of risk management options | Descriptive |
| 9 | Development of risk management options | Normative |
| 10 | Development of risk management options | Normative |
| 11 | Regulatory decisionmaking | Descriptive |
| 12 | Regulatory decisionmaking | Normative |
| 13 | Regulatory decisionmaking | Descriptive |
| 14 | Regulatory decisionmaking | Normative |
| 15 | Regulatory decisionmaking | Normative |
| 16 | Monitoring and evaluation | Descriptive |
| 17 | Monitoring and evaluation | Normative |
| 18 | Monitoring and evaluation | Descriptive |
| 19 | Monitoring and evaluation | Descriptive |

When we applied our criteria to answer the evaluation questions related to the quality of the risk analysis work, we rated each criterion on a scale from 1 to 5, 1 indicating that the agency had performed very inadequately and 5 indicating complete fulfillment of the criterion's objectives. The other points on the scale were 2 for less-than-adequate work, 3 for adequate work, and 4 for more-than-adequate work. We assigned a zero when no work had been performed. We wrote a justification for each rating, including our analysis and citations of the relevant source materials. All the data we obtained for each case were rated in this way for all the criteria that applied. If some criteria were inapplicable, this was indicated in the coding. We assigned no rating for criteria that were inapplicable, as for questions addressing whether guidelines were followed when the agency had no guidelines.

After we completed our ratings, we interviewed agency officials to gather more information concerning criteria that appeared not to have been met and to identify agency decisions and policies. We directed our questions at the officials who were responsible for conducting the risk

analysis under review. We incorporated this information into the ratings. Two of our staff members then examined each case independently, assessed the adequacy of the risk analysis, compared their ratings, and resolved any differences greater than one point.

We gave all criteria equal weight, except those which were not applicable to a case, which were weighted by zero. Ratings for related criteria for work in particular areas, such as source characterization, were assigned by averaging the ratings for all applicable criteria contained under that heading. These are the ratings that we report in chapters 2-4. Average scores of 2.5 or below represent less-than-adequate performance; average scores above 3.5 are more than adequate.

# Criteria for Evaluating the Adequacy of the Federal Risk Analysis Process

## Risk Assessment

### Hazard Identification

1. Administrative Components

a. Formal guidelines

(1) Did the agency have formal, comprehensive guidelines for conducting hazard identification?

(2) Were they written?

(3) Were they followed in this case?

(4) Was the use of guidelines documented?

b. External expert review

(1) Were there agency guidelines for external review?

(2) Were the guidelines written?

(3) Were they followed in this case?

(4) Were experts' concerns addressed?

(5) Was the review process documented?

c. Internal expert review

(1) Were there agency guidelines for internal review?

(2) Were the guidelines written?

(3) Were they followed?

(4) Were experts' concerns addressed?

(5) Was the process documented?

d. Administrative review

(1) Were uncertainties presented?

(2) Were assumptions presented?

(3) Were the studies used or conducted documented?

(4) Were there formal guidelines for administrative review?

(5) Were they written?

(6) Were they followed?

(7) Was the administrative review documented?

2. Scientific Components

a. Prioritization of potential hazards

(1) Was the magnitude of production of the substance examined?

(2) Was the intended use of the substance examined?

(3) Was the chemical structure examined?

(4) Was available toxicity information examined?

b. Determination of hazard and weighting of evidence

(1) Was the evidence from human studies characterized separately from that of animal studies?

(2) Were the two characterizations combined for an overall weight?

(3) Was other available supportive information reviewed to determine whether the weight should be modified?

(4) Did the weighting process use specific criteria for when to conduct quantitative risk assessments?

(5) Was each study that was reviewed weighted separately?

(6) Were areas in which additional research was needed identified?

c. Structure-activity relationship studies

(1) Were chemical properties examined?

(2) Were exposure pathways examined?

(3) Were structure-activity correlations examined?

(4) Were metabolic and pharmacokinetic properties examined?

(5) Were toxicological effects other than carcinogenesis examined?

d. Short-term bioassays

(1) Was the specific test selected validated for known animal carcinogens and noncarcinogens?

(2) Was a mechanism of action deduced from experimental evidence?

(3) Was the substance tested in a formal, planned battery of short-term tests?

(4) Were data presented on the purity of the tested suspect substance?

(5) Were relevant dose-response data available?

(6) Were there multiple positive results (at least 3) in multiple test systems (at least 2) measuring DNA damage, mutagenicity, or chromosomal change (which are multiple end points)?

(7) Were there multiple positive results for the same genetic effect from test systems of different biological complexity?

(8) Were structurally related carcinogens and noncarcinogens tested simultaneously with the suspect substance?

(9) Were procedures used to ensure the presence of enzymes that metabolize chemically unreactive carcinogens in mammals into reactive electrophiles?

e. Long-term bioassays

(1) Was the probability of false positives low?

(2) Were positive results obtained for both sexes of multiple species or strains?

(3) Were positive results obtained for multiple experiments?

(4) Were positive results obtained for multiple routes of administration?

(5) Were positive results obtained for multiple dose levels?

(6) Were there positive results for unusual tumor types as opposed to only common types with high spontaneous occurrence?

(7) Was the highest dose tolerated high enough to produce only minimal toxicity without reducing longevity?

(8) Was the follow-up period extended until low-dose survivors were reduced to 20-25 percent?

(9) Was exposure duration at least 18 months for mice, 24 months for rats?

(10) Was the possibly contaminating effect of low survival rates examined?

(11) Did each test group contain at least 50 rodents?

(12) Did postmortem examinations, or necropsies, include both gross and microscopic examinations?

(13) Were test animals randomly assigned?

(14) Was there evidence of adherence to good laboratory practices?

(15) Did each treatment group have a control group?

(16) Was there a rationale for the weight assigned to the significance level of each study in light of other information?

(17) Were both trend and pairwise statistical tests performed, with appropriate corrections for variable survival rates?

f. Epidemiological studies

(1) Were confounding variables such as smoking controlled?

(2) Was the probability of false positives low?

(3) Were positive results obtained for multiple studies?

(4) Were the results interpreted in terms of biological plausibility?

(5) Was there evidence of a dose-response relationship?

(6) Was there evidence that a reduction of exposure was followed by decreased cancer incidence?

(7) Were uncertainties involving the number of individuals needed for observation examined?

(8) Were uncertainties introduced by study duration examined?

(9) Was individual exposure measured?

(10) Were data on workers' history examined?

(11) Were treatment and control groups matched?

(12) Was the degree of association strong?

g. Documentation and reporting

(1) Were uncertainties characterized for each step?

(2) Were the data used for each step documented?

(3) Were study results reported completely?

(4) Were assumptions specified for each step?

## Dose-Response Assessment

1. Administrative Components

a. Formal guidelines

(1) Did the agency have formal, comprehensive guidelines for preparing dose-response assessments?

(2) Were they written?

(3) Were they followed in this case?

(4) Was the use of guidelines documented?

b. External expert review

(1) Were there agency guidelines?

(2) Were they written?

(3) Were they followed?

(4) Were expert comments addressed?

(5) Was the process documented?

c. Internal expert review

(1) Were there agency guidelines?

(2) Were they written?

(3) Were they followed?

(4) Were expert comments addressed?

(5) Was the process documented?

d. Administrative review

(1) Were uncertainties presented?

(2) Were assumptions presented?

(3) Were the data and techniques used documented?

(4) Were there formal guidelines for administrative review?

(5) Were they written?

(6) Were they followed?

(7) Was the administrative review documented?

2. Scientific Components

a. Study selection for final estimation

(1) Were available epidemiological studies, including those showing negative results, considered in preparing the final estimate?

(2) If long-term animal test data were selected, were issues associated with biological sensitivity and similarity to human responses addressed?

b. Interspecies extrapolation models

(1) Were standardized scaling factors (such as mg/kg/day or mg/kg/lifetime) used?

(2) Were the following differences between human and animal characteristics examined: (a) body weight, (b) life span, (c) body size, (d) genetic variability, (e) population homogeneity, (f) existence of concurrent disease, (g) metabolic patterns, (h) excretion patterns, (i) exposure regimen?

c. Low-dose extrapolation model: Was the model selected examined for consistency with biological plausibility and other available information?

d. Short-term bioassays

(1) Was the specific test selected validated for known animal carcinogens and noncarcinogens?

(2) Was a mechanism of action deduced from experimental evidence?

(3) Was the substance tested in a formal, planned battery of short-term tests?

(4) Were data presented on the purity of the tested suspect substance?

(5) Were the tests used in conjunction with other evidence?

(6) Were there multiple positive results (at least 3) in multiple test systems (at least 2) measuring DNA damage, mutagenicity, or chromosomal change (which are multiple end points)?

(7) Were there multiple positive results for the same genetic effect from test systems of different biological complexity?

(8) Were structurally related carcinogens and noncarcinogens tested simultaneously with the suspect substance?

(9) Were procedures used to ensure the presence of enzymes that metabolize chemically unreactive carcinogens in mammals into reactive electrophiles?

e. Long-term bioassays

(1) Was the probability of false positives low?

(2) Were positive results obtained for both sexes of multiple species or strains?

(3) Were positive results obtained for different experiments?

(4) Were positive results obtained for different routes of administration?

(5) Were positive results obtained for different dose levels?

(6) Were at least three dosage-level groups suitable for analysis tested?

(7) Were there positive results for unusual tumor types as opposed to common types with high spontaneous occurrence?

(8) Was the maximum tolerated dose high enough to produce only minimal toxicity without reducing longevity?

(9) Was the follow-up period extended until low-dose survivors were reduced to 20-25 percent?

(10) Was exposure duration at least 18 months for mice, 24 months for rats?

(11) Was the possibly contaminating effect of low survival rates examined?

(12) Did each test group contain at least 50 rodents?

(13) Did postmortem examinations, or necropsies, include both gross and microscopic examinations?

(14) Were test animals randomly assigned?

(15) Was there evidence of adherence to good laboratory practices?

(16) Did each treatment group have a control group?

(17) Was there a rationale for the weight assigned to the significance level of each study in light of other information?

(18) Were both trend and pairwise statistical tests performed, with appropriate corrections for variable survival rates?

f. Epidemiological studies

(1) Were confounding variables such as smoking controlled?

(2) Was the probability of false positives low?

(3) Were uncertainties involving the number of individuals needed for observation examined?

(4) Were uncertainties introduced by study duration examined?

(5) Were the results interpreted in terms of biological plausibility?

(6) Was individual exposure measured?

(7) Were data on workers' history examined?

(8) Were treatment and control groups matched?

(9) Was the degree of association strong?

(10) Was the response curve (whether linear or other) consistent with biological knowledge?

(11) Was there evidence that a reduction of exposure was followed by decreased cancer incidence?

(12) Were uncertainties involving exposure duration examined?

(13) Were positive results obtained for multiple studies?

g. Documentation and reporting

(1) Were uncertainties characterized for each step?

(2) Were the data used for each step documented?

(3) Were study results reported completely?

(4) Were assumptions specified for each step?

## Exposure Assessment

1. Administrative Components

a. Formal guidelines

(1) Did the agency have formal, comprehensive guidelines for preparing exposure assessments?

(2) Were they written?

(3) Were they followed in this case?

(4) Was the use of guidelines documented?

b. External expert review

(1) Were there agency guidelines?

(2) Were they written?

(3) Were they followed?

(4) Were expert comments addressed?

(5) Was the process documented?

c. Internal expert review

(1) Were there agency guidelines?

(2) Were they written?

(3) Were they followed?

(4) Were expert comments addressed?

(5) Was the process documented?

d. Administrative review

(1) Were uncertainties presented?

(2) Were assumptions presented?

(3) Were the data and techniques used documented?

(4) Were there formal guidelines for administrative review?

(5) Were they written?

(6) Were they followed?

(7) Was the administrative review documented?

2. Scientific Components

a. Source characterization

(1) Were production and distribution of the substance examined?

(2) Were the uses that create potential sources of exposure examined?

(3) Was disposal of the substance after use examined?

(4) Was the amount of emissions examined?

b. Exposure routes and concentration

(1) Were the routes of exposure, including movement from one medium to another, examined?

(2) Were the conditions of exposure (that is, agent transformation and human activity) examined?

(3) Were exposure routes outside the agency's regulatory jurisdiction examined?

(4) Was duration of exposure (including a description of major methods and their strengths and weaknesses) examined?

(5) Was the frequency of exposure (including description of major methods) examined?

(6) Was the intensity of exposure examined? (describe major methods)

c. Populations at risk

(1) Were the size and characteristics of exposed groups examined?

(2) Were high-risk groups (those who experience high exposure or high sensitivity) examined?

d. Documentation and reporting

(1) Were uncertainties characterized for each step?

(2) Was a range of exposure values presented?

(3) Were the data used for each step documented?

(4) Were study results reported completely?

(5) Was an integrated exposure assessment presented?

(6) Were assumptions specified for each step?

## Risk Characterization

### 1. Administrative Components

a. Formal guidelines

(1) Did the agency have formal, comprehensive guidelines for risk characterization?

(2) Were they written?

(3) Were they followed in this case?

(4) Was the use of guidelines documented?

b. External expert review

(1) Were there agency guidelines?

(2) Were they written?

(3) Were they followed?

(4) Were expert comments addressed?

(5) Was the process documented?

c. Internal expert review

(1) Were there agency guidelines?

(2) Were they written?

(3) Were they followed?

(4) Were expert comments addressed?

(5) Was the process documented?

d. Administrative review

(1) Were uncertainties presented?

(2) Were assumptions presented?

(3) Were the data and techniques used documented?

(4) Were there formal guidelines for administrative review?

(5) Were they written?

(6) Were they followed?

(7) Was the administrative review documented?

2. Scientific Components

a. Estimation procedures

(1) Was the information derived from hazard identification analysis, exposure assessment work, and dose-response assessment reviewed and evaluated in order to arrive at an overall health risk estimate?

(2) Were the uncertainties associated with the exposure assessment described (and, if possible, quantified) and explicitly included in the risk characterization analysis?

(3) Were the uncertainties associated with the dose-response work described (and, if possible, quantified) and explicitly included in the risk characterization analysis?

(4) Were high-risk subgroups identified and the degree of risk they faced examined and accounted for in the risk characterization?

(5) Were hazard sources defined with respect to duration, frequency, and intensity?

(6) Were the compounding effects of uncertainties associated with the hazard assessment, dose-response assessment, and exposure assessment accounted for in the risk characterization?

(7) Were the ranges of values for population parameters' known probabilities, or confidence intervals (in addition to best estimates), calculated for the risk characterization?

(8) Were both population and individual risk estimates calculated?

b. Documentation and reporting

(1) Were uncertainties characterized for each step?

(2) Were the data used for each step documented?

(3) Were confidence limits or best estimates reported for the risk characterization?

(4) Were both population and individual risk estimates reported?

(5) Was zero risk discussed?

(6) Were study results reported completely?

# Risk Management

## Development and Evaluation of Risk Management Options

1. Administrative Components

a. Formal guidelines

(1) Did the agency have formal, comprehensive guidelines for the development of management options?

(2) Were they written?

(3) Were they followed in this case?

(4) Was the use of guidelines documented?

b. External expert review

(1) Were there agency guidelines?

(2) Were they written?

(3) Were they followed?

(4) Were expert comments addressed?

(5) Was the process documented?

c. Internal expert review

(1) Were there agency guidelines?

(2) Were they written?

(3) Were they followed?

(4) Were expert comments addressed?

(5) Was the process documented?

d. Administrative review

(1) Were uncertainties presented?

(2) Were assumptions presented?

(3) Were the data and techniques used documented?

(4) Were there formal guidelines for administrative review?

(5) Were they written?

(6) Were they followed?

(7) Was the administrative review documented?

2. Technical and Scientific Components

a. Development of risk management options

(1) Were the variables or factors such as costs and benefits associated with each option specified?

(2) Were the methods and assumptions used in the development of such variables as costs and benefits specified?

(3) Were value judgments for each risk management option specified?

(4) Were uncertainties associated with the development of each risk management option specified?

(5) Did the agency use an analytical approach other than, or in addition to, worst-case analysis?

(6) Were the risk management options compared to earlier risk management options for similar hazards for the purpose of validation check?

(7) Was the development of risk management options independent of the risk assessment work?

(8) Were risk management options reviewed with respect to practicality?

(9) Was the achievable risk reduction estimated for each option?

(10) Were both population and individual risk indicators examined?

(11) Was the relationship between risk reduction and cost examined for each option?

(12) Was the "no regulation" option examined?

b. Documentation and reporting

(1) Were uncertainties characterized for each options?

(2) Were the data used in developing options documented?

(3) Was the development of risk management options reported completely?

(4) Were assumptions specified for each option?

## Regulatory Decisionmaking

### 1. Administrative Components

a. Formal guidelines

(1) Did the agency have formal, comprehensive guidelines for regulatory decisionmaking?

(2) Were they written?

(3) Were they followed in this case?

(4) Was the use of guidelines documented?

b. Compliance with legislative authority

(1) Was there clear reference to the appropriate legislative authority allowing the agency to regulate in the manner proposed?

(2) Did the regulatory decision include documentation of any deviations from previous regulatory practices for similar cases?

c. External regulatory review

(1) Were the comments made by external reviewers documented?

(2) Was each comment addressed by the agency?

(3) Was a rationale provided for making changes pursuant to the external review?

d. Administrative review

(1) Were there formal guidelines for administrative review?

(2) Were they written?

(3) Were they followed?

(4) Was the administrative review documented?

(5) Was a rationale provided for changes made during administrative review?

2. Technical and Scientific Components

a. Decisionmaking procedures

(1) Were the decisionmakers provided with information regarding the uncertainties associated with each step of the risk analysis?

(2) Were the underlying assumptions, methodologies, and statistical procedures used at each step of the risk analysis presented?

(3) Were the risk management options and their likely consequences presented?

(4) Were the risk management options ranked according to acceptability?

(5) Was comparative risk information presented?

(6) Were data for high-risk subgroups presented?

(7) Was a review of the practicality and feasibility of implementing the policy options presented?

(8) Was there any consideration of whether new data may be available shortly that could revise the regulatory response?

(9) Were the decisions made as part of the regulatory response clearly based on the information generated by the risk assessment work?

(10) Were the decisions made as part of the regulatory response based on the information generated by the evaluation of risk management options?

(11) Were the decisions made as part of the regulatory response based upon a comparative review of either agency or other risk response decisions?

(12) If a benefit-cost analysis was performed, was there evidence that the full range of risk management options was included?

b. Documentation and reporting

(1) Were uncertainties characterized for the risk management option selected?

(2) Were the data on which the decision was based documented?

(3) Were the decisions made as part of the regulatory response documented?

(4) Was the decision process reported completely?

## Monitoring and Evaluation

1. Administrative Components

a. Formal guidelines

(1) Did the agency have formal, comprehensive guidelines for conducting follow-up evaluations?

(2) Were they written?

(3) Were they followed in this case?

(4) Was the use of guidelines documented?

b. Communication and disclosure

(1) Were there formal guidelines for disclosure of evaluation results?

(2) Were they written?

(3) Were they followed?

(4) Were the data made accessible to other researchers outside the agency for reanalysis?

c. Use of evaluation findings

(1) Were the findings used to evaluate effectiveness?

(2) Were the findings brought to the attention of appropriate decisionmakers?

(3) If implied by the findings, were the results used to modify the regulation?

(4) If implied by the findings, were the results used to modify risk assessment procedures or models?

d. External expert review

(1) Were there agency guidelines?

(2) Were they written?

(3) Were they followed?

(4) Were expert comments addressed?

(5) Was the process documented?

e. Internal expert review

(1) Were there agency guidelines?

(2) Were they written?

(3) Were they followed?

(4) Were expert comments addressed?

(5) Was the process documented?

f. Administrative review

(1) Were uncertainties presented?

(2) Were assumptions specified?

(3) Were the data and techniques used documented?

(4) Were there formal guidelines for administrative review?

(5) Were they written?

(6) Were they followed?

(7) Was the administrative review documented?

2. Scientific Components

a. Evaluation formulation and negotiation

(1) Were the goals of the evaluation clearly defined?

(2) Was the regulatory action reviewed to determine its evaluability?

(3) Were any shortcomings pertaining to the evaluability of the regulatory action identified and managed?

b. Evaluation structure and design

(1) Was the methodology appropriately developed for evaluation of the regulation and its effect?

(2) Were the methodological weaknesses identified and managed?

(3) Were the measurement methods and instruments specified and described, and were their reliability and validity estimated?

(4) If sampling was used, was the approach described and supported?

c. Data collection and preparation

(1) Was there a data collection plan?

(2) Were the data checked for missing or inconsistent information?

(3) Were threats to the integrity of the data identified and handled?

d. Data analysis and interpretation

(1) Was the rationale for the selection of the statistical methods or other analytic techniques described?

(2) Were the weaknesses of the methods acknowledged?

(3) Were the assumptions underlying the methods used appropriate to the data to which they were applied?

(4) Was the unit of analysis appropriate?

(5) Were cause-and-effect interpretations bolstered by recognition and elimination of plausible, rival explanations?

(6) Were indications provided of both the statistical and practical significance of the findings?

e. Documentation and reporting

(1) Was the evaluation plan documented?

(2) Were the study results reported completely?

(3) In the reporting of the results of the evaluation, was an attempt made to link the results with the regulatory action being examined?

(4) In the reporting of the results of the evaluation, were the implications of the results discussed?

(5) Were uncertainties characterized?

# Expert Panels

## Membership of Panel 1

John F. Ahearne
Resources for the Future

Vincent T. Covello
National Science Foundation

Michael E. Kraft
University of Wisconsin-Green Bay

Lester B. Lave
Carnegie-Mellon University

Joseph V. Rodricks
ENVIRON Corporation

## Membership of Panel 2

Elizabeth L. Anderson
ICF-Clement, Inc.

Michael Gough
ENVIRON Corporation

Steven M. Swanson
American Petroleum Institute

# Risk Analysis for EPA's 1979 Ozone Standard

In 1979, EPA relaxed the ozone standard first promulgated in 1971. The 1971 standard for photochemical oxidants was set at an hourly average level of 0.08 parts per million ambient concentration. The 1979 revision relaxed the standard by 50 percent to 0.12 ppm.

Several events contributed to the change. While EPA was required by the Clean Air Amendments of 1970 to set a standard by 1971, it acknowledged that the initial criteria document on ozone contained considerable uncertainty. EPA acknowledged having misinterpreted a pivotal study concerning the effects of ozone on asthma victims. The American Petroleum Institute and the city of Houston petitioned EPA to revise the standard in 1976 and 1977, respectively. American Petroleum's petition alleged that studies, new at the time, demonstrated no significant adverse health effects at or below the 0.25 level. According to the director of EPA's office of air quality planning and standards at that time, regional office and state agency staffs also believed that exposure to two to five times the standard was not a public health problem that required an urgent response. Finally, a National Academy of Sciences study, commissioned by the Congress in 1973, found that "The technical data base for the oxidant standard was inadequate at the time the standard was set . . . ."

After the revised standard was issued, EPA was sued by the American Petroleum Institute and the Natural Resources Defense Council. The American Petroleum suit alleged that the evidence did not support such a stringent standard, while the National Resources Defense Council suit alleged that the evidence did not support a relaxation of the standard. Both suits alleged that procedural deficiencies had influenced the final decision in ways not supported by the evidence. The rulemaking procedures developed by the courts in the 1970's, in part to ensure that EPA based its actions on scientific evidence rather than political pressure, failed to produce a scientific consensus on the threshold for the adverse effect of ozone and revealed enough empirical uncertainty that the agency had a wide array of potential standards from which to choose. The agency's lawyers determined that they could defend any of the alternatives under consideration: 0.08 ppm, 0.10 ppm, and 0.12 ppm. Thus, the decisionmakers were forced to confront the difficult issue of decisionmaking under conditions of uncertainty.

The court found that "where the Administrator bases his conclusion as to an adequate margin of safety on a reasoned analysis and evidence of risk, the court will not reverse," even though uncertainty may remain. The court also determined that EPA may err on the side of overprotection

by setting a fully adequate margin of safety. In the remainder of this appendix, we briefly describe the phases of the risk analysis process that produced the controversy and current developments at EPA.

## Hazard Identification

Some of the acutely adverse effects ozone produces on health were identified in 1851, and it is still generally agreed that exposure to ozone constitutes a health hazard. Recent efforts have concentrated on establishing the dose-response relationship and the level of exposure that provides an adequate margin of safety.

Ozone is currently regulated only as an acute health hazard, although animal studies indicate that ozone may be carcinogenic. The project managers for the current revision told us that they have not reached consensus on a long-term primary standard and that uncertainty, including the lack of an adequately validated interspecies extrapolation model, is one reason for the lack of consensus. A draft staff paper published in 1986 stated that insufficient quantitative data were available to support a long-term primary standard but that these data should be used in developing a margin of safety.

## Dose-Response Assessment

The intense controversy surrounding the 1979 revision involved the dose-response relationship. Much of the key data were from clinical experimental studies on humans that were designed to demonstrate short-term decreases in lung function at response threshold levels but that were criticized for a number of reasons. For example, industry argued that the changes in pulmonary function observed after exposure to 0.15 ppm ozone may not be adverse health effects. There was little disagreement that adverse effects in some people are demonstrable above 0.25 ppm. It was levels below this that were in question.

Animal studies showed chronic effects, such as emphysema and impairment of immunological systems, from prolonged exposure to low levels of ozone. However, the extent to which these results are applicable to humans is uncertain. EPA officials stated that the results of animal studies carry greater weight when the effect is carcinogenic. More recent animal studies also suggest lung structure damage and increased susceptibility to respiratory infection following long-term exposure to ozone. A draft staff paper on ozone published in 1986 stated that further discussion with the clean air science advisory committee concerning the definitions of emphysema, preemphysematous lesions, and lung fibrosis may help clarify some of the uncertainty in these studies.

EPA has recently developed formal guidelines for risk assessment of car cinogens that employ conservative assumptions (but not the most conservative assumptions) to compensate for such uncertainties as those involving interspecies extrapolation. These guidelines are only indirect applicable, because ozone is regulated as a systemic toxin, not as a carcinogen. The basic difference is in the sound basis for believing that sy temic toxins have exposure thresholds for adverse effects on health, although they may be below ambient levels; methods for low-dose extrapolation and interspecies extrapolation are therefore different. However, such thresholds may be quite difficult to establish empiricall An EPA working group is studying possible guidelines for risk assessments of systemic toxins, but EPA has no guidelines currently in effect. Some EPA officials we interviewed stated that existing guidelines for ca cinogens were applicable but not being used, because the conservative assumptions would result in an ozone standard whose level would be lower than natural background levels. Although other EPA officials stated that the risk assessment guidelines for carcinogens do not apply to ozone, interspecies extrapolation for systemic toxins entails conserv tive assumptions just as it does for carcinogens. Thus, it appears that the effect of the animal studies on the standard is sensitive to the conservatism of the assumptions employed in extrapolating the results to humans.

## Exposure Assessment

Estimates of actual exposure to ozone are based on measurements of th ambient ozone level taken at monitoring stations in urban areas. For th 1979 revision, concentrations of ozone in the ambient air were those measured in 1974 at some 340 monitoring stations operated by state a local control agencies. Exposure assessment is complicated by the nee to relate the amounts of ozone precursors, such as volatile organic com pounds, emitted through human activities to the observed ozone levels and to predict the amount of emission reduction necessary to achieve the air quality standard. This portion of exposure assessment is especially difficult and controversial. It involves determining the amount o emissions of the substances that are transformed into ozone, analyzing the process of transformation itself, and estimating the dispersion of ozone and its precursors over the short, medium, and long distances du ing which the chemical reactions may occur.

The interaction of volatile organic compounds, nitrogen oxides, and su light under a variety of meteorological conditions to produce specific ozone levels is described by several types of models. These "air quality models" are highly controversial, because it is not feasible to achieve t

ideal requirements for validity and utility. For example, dispersion over distances longer than 1,000 miles while the photochemical reactions are occurring has not been adequately modeled. Improvements have been made since the 1979 revision, but the models remain limited to moderate dispersion distances and to the northeast corridor.

## Risk Characterization

The risks of neither individuals nor populations were estimated with quantitative risk assessment techniques. Instead, estimates of the human response threshold level of exposure were made with judgmental techniques, including subjective probability encoding. According to a major report in EPA's public record on ozone, subjective probability encoding was designed to "estimate" the probability that each of several exposure levels is the threshold for sensitive groups by averaging expert judgments.

EPA's techniques were controversial, as were the conditions under which they were performed. Some experts agreed to participate in the subjective probability encoding study only with the understanding that it was experimental and would not be used to support the standard, but the results were included in the publication of the rule, suggesting that they were used, whereas the Federal Register announcement was equivocal, stating that the technique was not used but was considered in setting the standard. EPA maintains that the notice simply explained the basis of the decision and EPA is currently developing a refined probabilistic risk assessment approach that it believes addresses many of these criticisms.

## Risk Management

Decisionmaking was made especially difficult because EPA's scientific staff could not agree on the level at which adverse effects on health were demonstrated. Staff from the ambient standards branch told us that science takes one as far as it can, which is only to a "window of uncertainty" or a range of values. The selection of a given value within that range is a policy decision. They also stated that the wider the window of uncertainty, the more controversial the final decision will be.

In the face of empirical uncertainty, it is likely that factors other than risk may be considered in deciding on a final standard. For the revised ozone standard, an action memo issued prior to the proposal discussed the economic effects of the alternative levels under consideration. Internal communications also noted that many cities could comply with a standard of 0.10 ppm by 1987. Although the EPA administrator at the time stated that EPA did not consider economic factors, these documents

create the appearance that cost and technical feasibility were considered.

EPA was pressured by other offices of the executive branch to relax the standard above the proposed 0.10 ppm. The Council of Economic Advisors, the Office of Science and Technology Policy, the Regulatory Analysis Review Group, and the Council on Wage and Price Stability criticized the proposed standard of 0.10 because it would have high marginal costs and it was based on inconclusive and flawed studies. When EPA proposed to relax the standard to 0.12 ppm, the Office of the President asked EPA to raise it further to the equivalent of 0.14 ppm by permitting that the standard could be exceeded an additional number of times annually. When EPA declined, the Council of Economic Advisors considered a last-minute appeal to the president, but similar appeals in another case were being litigated and the standard of 0.12 ppm was issued.

The National Resources Defense Council suspected that political pressures led EPA to reassess its interpretation of the evidence. A more fundamental issue, raised by EPA in the preamble of the final standard, is the absence of a demonstrable threshold for effects on health, which suggests that the scientific research did not provide an undisputed threshold. The risk analysis literature agrees that in the absence of demonstrated thresholds, some factors, whether economic or political, are likely to be balanced against others in the attempt to define an acceptable level of risk. Without clear decisionmaking principles that specify how uncertainties are to be considered, some legitimate doubt will remain about the correctness and propriety of the decisions.

# Comments From the Department of Health and Human Services

Note: GAO comments supplementing those in the report text appear at the end of this appendix.

**DEPARTMENT OF HEALTH & HUMAN SERVICES** Office of Inspector General

Washington, D.C. 20201

APR [illegible] 7

Mr. Richard L. Fogel
Assistant Comptroller General
U.S. General Accounting Office
Washington, D.C. 20548

Dear Mr. Fogel:

The Secretary asked that I respond to your request for the Department's comments on your draft report, "Health Risk Analysis: Technical Adequacy in Three Selected Cases." The enclosed comments represent the tentative position of the Department and are subject to reevaluation when the final version of this report is received.

We appreciate the opportunity to comment on this draft report before its publication.

Sincerely yours,

Richard P. Kusserow
Inspector General

Enclosure

COMMENTS ON THE GENERAL ACCOUNTING OFFICE
DRAFT REPORT ENTITLED: "HEALTH RISK ANALYSIS: TECHNICAL ADEQUACY IN THREE SELECTED CASES" MARCH 1987

We appreciate the opportunity to review and comment on the draft report, especially because this study was a pilot for the more extensive audit the General Accounting Office (GAO) has been requested to do. GAO is to be commended for their serious effort directed at a very difficult issue. The draft report comes to the Food and Drug Administration for review at a particularly appropriate time, as the Commissioner of Food and Drugs has initiated an Agency-wide program to strengthen risk analysis processes within the Agency. In accordance with the Agency's Action Plan, a review of the risk management process was undertaken in 1986 and a report to the Commissioner has been made.

FDA is keenly interested in the approach used by GAO for evaluating current and future risk assessment/management activities carried out by the Agency as well as those of other regulatory agencies of the Federal government. We believe it will be in the best interest of the agencies and the general public to assure risk assessments/management decisions made by regulatory agencies are of a high quality. We also believe that evaluations of such activities by those charged with oversight responsibilities should be fair and accurate. We have, therefore, critiqued the draft report with the intent of providing constructive feedback to the auditors to help them refine the evaluation model.

We have reviewed the draft from two perspectives:

1) Are the criteria and evaluation approach valid and applicable to other risk assessment situations under the purview of the Food and Drug Administration?

2) Are the specific findings relative to methylene chloride accurate, valid and appropriate to the regulatory milieu of food additives?

We also have the following general observations with regard to this particular evaluation and the risk assessment/management practices of the Agency.

First, we believe the approach used by the auditors of clearly separating risk assessment activities from risk management activities to be appropriate, even though there is clearly a significant area of overlap between the two in actual practice. The risk assessment activities of regulatory agencies should be well grounded scientifically. We were pleased that by-and-large GAO found FDA's risk assessment of methylene chloride to have been so grounded.

Risk management, on the other hand, must take cognizance of a number of non-scientific parameters, such as applicable statutes, expediency of action required by the identified risk, possible mechanisms for taking appropriate action to minimize or eliminate the risk (i.e., seizure, recall, product banning, labeling revisions, prosecutions, etc.), cost to the agency - and the public - of any proposed action and the likely effectiveness of alternatives under consideration.

See comment 1.

Secondly, the issue of documentation raises some questions about the validity of any set of criteria applied in a checklist mode for determining the quality of a complex process such as risk assessment/management. It would appear that documentation often serves merely as an indication that a particular activity was done, but does not address the quality of the activity, or the resultant decisions. Nevertheless, FDA has taken a look at this issue and found some weaknesses similar to those identified by GAO. The Commissioner of Food and Drugs has the recommendations of an internal risk management report under advisement.

See comment 2.

Thirdly, the report places heavy emphasis on the use of guidelines in both the risk assessment and the risk management processes and faults FDA for not having developed guidelines specifically for situations under its purview. We believe this is a misunderstanding of FDA's policy and practice. The agency uses the Office of Science and Technology (OSTP) publication entitled "Chemical Carcinogens; A Review of the Science and its Associated Principles," March 4, 1985, as a guideline for addressing cancer risk assessment. Other relevant guidelines are also used. As stated frequently to the auditors, FDA

See comment 3.

believes that guidelines should be flexible enough to allow scientific and managerial judgement to be applied to each individual case. The state of scientific knowledge is changing rapidly and there are frequently gaps in knowledge when decisions must be made.

See comment 4.

As to risk management, the options available to FDA are generally those in the applicable statutes, regulations, and in court decisions. These "guidelines" are followed in reaching risk management decisions.

See comment 5.

To develop guidelines for all risk assessment cases regulated by FDA would not only be very costly in light of the small public health advantage realized, but they would be so general as to be of little value. There are significant differences between the products we regulate (foods, drugs, cosmetics, medical devices, radiation emitting devices such as microwave ovens, etc.) and the risks are so diverse that appropriate risk assessment approaches for one product are not appropriate for another. There are also many Agency decisions that do not require a formal decision-making process or a full risk assessment.

See comment 6.

Fourth, we found it somewhat disturbing that all the criteria listed in Appendix III were apparently given the same weight. Some factors are significantly more important for determining risk than are others. We believe the model used by GAO for its evaluation would be more valid if the criteria were weighted to reflect their relative importance. It should be recognized, however, that weighting appropriate for one situation may not be for other situations.

See comment 7.

We also had difficulty understanding how GAO arrived at the ratings given the various components of risk assessment/management identified in the report. More explanation of the process and identification of the factors that produced a specific rating would be beneficial and would facilitate a "peer review" of the quality of the evaluation and it's relevance to risk analysis. We further question the validity of the approach used to resolve differences in ratings between the GAO reviewers.

Fifth, as the questions posed above indicate, we reviewed the criteria for their applicability to other FDA regulatory matters where either a less comprehensive risk assessment is appropriate or where different factors need to be considered. We do not believe the criteria are universally applicable, even within FDA. There are many products regulated by FDA for which risk assessments must be made, but for which formal risk assessments of the nature GAO apparently has in mind are

See comment 8.

not required. In some other instances the risk/benefit balance is so entwined (drugs, devices) that the assessment of risks alone would lack credibility in an overall evaluation. The criteria seem to be most

See comment 9.

applicable to cancer risk assessment cases. However, there are many more risks that the agency must address than just the risk of cancer. Also, many of the products regulated by FDA require a benefit/risk assessment rather than just a risk assessment (drugs, electronic medical devices, biologicals, animal drugs, etc.). We, therefore, have

See comment 10.

reservations about the validity of attempts to extrapolate from a single instance (no matter how randomly selected) to more generalized statements of program effectiveness or efficiency.

Finally, we are concerned about statements to the effect that FDA "appears to consider "other factors" in making its decision. A consumer agency, such as FDA, will always be subject to inquiries and pressures from regulated firms, associations and members of Congress representing them. Within the context of not setting any requirement or issuing any regulation that exceeds what is necessary to assure

See comment 11.

public health and safety, we believe one should weigh carefully the economic impact of our actions. The law does not preclude such consideration, but only defines what is determining. Public health and safety are the ultimate determinants of such decisions.

More specific comments about the criteria and the methylene chloride case follow.

1. Are the criteria and evaluation approach applicable to other risk assessment situations under the purview of the Food and Drug Administration?

See comment 12.

We believe GAO's use of the definition of risk analysis to determine what cases to evaluate has been too broadly applied.

Page 4 - Comments on GAO Report

As discussed below, the methylene chloride case is not a typical FDA risk analysis. To fairly evaluate the Agency's performance, we suggest that GAO's future evaluations be of more typical cases.

See comment 13.

- Measurement of agency performance versus factors outside the control of the agencies. We believe an evaluation of the risk assessment activities of an agency should focus on those aspects of risk analysis that are within the agency's control. If an absolute measure of overall quality of risk analysis including that of data submitted to the agency is desired by GAO, a distinction should be made between those factors that are within an agency's control and those that are not. The current scientific inadequacies in risk assessment, limitations in scientific procedure, the inadequacy of the data base an agency is often forced to accept, the constraints of the statute and limited resources all can adversely affect risk analysis/management. As measured against an ideal standard, an agency's risk analysis can be imprecise despite its own outstanding efforts in dealing with the information at its disposal. While the report commends FDA for dealing appropriately with incomplete and insufficient data, GAO should clearly define the objectives of the report in this respect.

See comment 14.

See comment 15.

- Timeliness and Consideration of the imperatives of the regulatory process. For any agency to address all the criteria listed in Appendix III as a model checklist for ensuring adequate risk analysis for every decision would frequently delay decision making considerably. The more serious the risk, the greater is the urgency to arrive at a decision. Often decisions must be made quickly to protect the public from acute problems. In those instances, the risk analysis phase may be drastically truncated. Any risk analysis scheme that would satisfy GAO's criteria would be lengthy and generally only useful in evaluating relatively low level risks. If these criteria were used, it could have the effect of faulting an agency for acting promptly in the interest of the public when either a less comprehensive risk assessment is appropriate or different factors need to be considered.

See comment 16.

- Unnecessarily detailed criteria. Application of these very detailed criteria (349) to evaluate specific risk cases would be very resource intensive and likely produce a high subjective and lengthy audit report.

See comment 1.

See comment 6.

See comment 17.

See comment 18.

Page 5 - Comments on GAO Report

- Form evaluated rather than substance. The use of such detailed criteria is likely to become little more than the application of a checklist to see that each element was considered, not a qualitative review to determine whether the analysis was appropriate to the degree of risk, the analysis done was of an acceptable caliber, etc. We believe a less extensive "checklist" that takes scientific judgment into consideration would be more useful doing a qualitative evaluation, because risk analysis is ultimately dependent upon good scientific judgment.

- Values assigned each criterion. As applied in the methylene chloride case, the criteria were all apparently given equal weight regardless of the relative value of the information to the analytical process. We believe this is an inappropriate application of the criteria, that resulted in a distorted evaluation of the methylene chloride decision. We recognize the difficulty of assigning weights to the voluminous list of criteria, but believe the effort to do so by GAO coupled with reducing the number of items would enhance the quality of their evaluation.

- Validity of determining risk assessment quality using only a documentation review. A major element in risk assessment is the quality of the agency's scientific analysis and judgment, which cannot be determined from a review of a documentation trail for any risk assessment/management case. The criteria indicate some but not all of the issues that are important to producing quality scientific analyses, nor do they indicate what an adequate scientific consideration is or how it might be assessed. Reviewing documentation on the basis of this or any listing of criteria will not address the quality of the science involved. It is conceivable to have documentation indicating that each and every point has been addressed and still have the science done poorly or inappropriate decisions. Conversely, it is equally conceivable to have omitted some of the phases GAO evaluated and still have done high quality science/decision-making. Not all or even most of an agency's decision-making is documented. Some of it is handled through routine administrative procedures, some through frequent, informal meetings. Auditing decision-making is difficult and requires a good deal of time and first hand observation. If an agency risk analysis process is to be judged against a set of criteria, they need to be very thoughtfully considered, and augmented with a thorough understanding of the institutional decision and review process.

See comment 19.

Page 6 - Comments on GAO Report

2. Are the specific findings relative to methylene chlroide accurate, valid and appropriate to the regulatory milieu of food additives?

In reviewing the draft relative to the above question, FDA took note of the stated intent of the auditors to study cases of actions published in the Federal Register between 1981-1985 that fall under normal agency procedures that use standard risk analyses. Since methylene chloride does not meet these criteria, we believe the usefulness of the audit approach for evaluating FDA's risk assessment activities is limited. The unique aspects of the methylene chloride case include:

- Methylene chloride is both a food additive (used in decaffeinating coffee) and an ingredient in some aerosal cosmetics (used in a propellant). The safety of both cosmetic ingredients and direct food additives is such that only rarely has the agency had occasion to propose corrective action for substances with a long history of use such as methylene chloride has. This would automatically make this risk assessment non-routine.

- The methylene chloride case has not yet been completed. A proposed course of action was published in the Federal Register for the purpose of eliciting comments concerning the merits and demerits of the case, both from a scientific perspective and from that of the general population. Such comments form an integral part of the final risk assessment done by the agency and thus of the final disposition of the case. In this specific case, FDA received information that is valuable to several aspects of the risk analysis. The audit, therefore, evaluates an ongoing risk analysis process and makes recommendations on the basis of an incomplete record.

- The proposed action as publised in the Federal Register is precedent setting. It represents the first - and only - situation where FDA has proposed that the safety of a carcinogenic substance added intentionally to food could be evaluated by a risk assessment process. Until recently, the only course of action open to the Agency relative to an intentionally added food substance that has been determined to be carcinogen - regardless of the amount or the potency - was to ban its use in food. Recent court decisions have opened the way for FDA to consider whether or not the risks from such substances are so small as to be insignificant (de minimis) and, therefore, whether continued use of the substance is permissible. The methylene chloride case is the first direct food additive "de minimis" case to be considered by FDA.

FDA evaluates many food additive petitions each year that could be considered a risk assessment. Had GAO selected any one of these petitions for its pilot study, we believe the results would have been a more valid test of FDA's risk assessment activities and of the evaluation model being used.

Other areas of specific concern raised by GAO's application of the criteria to methylene chloride include:

See comment 2.

- The report faults FDA for having no written guidelines to follow in assessing the risks of methylene chloride. In conducting the methylene chloride analysis, however, FDA followed the OSTP guidelines mentioned above.

See comment 7.

- The report does not provide sufficient guidance as to how the criteria were applied to allow us to determine how the specific ratings were obtained. The ratings illustrate, however, some of the problems inherent in evaluating a complex process such as risk assessment by means of a criteria checklist. The following examples illustrate the problems resulting from such arbitrary decisions.

  See comment 20.

  First, the report states that FDA combines the hazard identification and dose response assessment phases because the question, "is effect treatment related?" is a dose-response question, not a hazard identification question. To the contrary, however, this question is not used to determine the response relative to dose (as would be done for a dose response assessment) but rather is used to determine whether there is any potential hazard, the basic question of hazard identification. By merely listing this question as a dose-response issue alone, the report introduces an invalid assumption.

  See comment 21.

  Secondly, the statement that "No systematic follow-up occurs to determine if the implemented regulations are having their intended risk reduction effects" is erroneous not only because the case has not yet been concluded, but also in its assumptions of what such a follow-up could accomplish. Risks regulated by FDA in the food and cosmetic area are generally far too small to be measured in the human population. This is clearly the case for a chronic effect such as cancer. The Federal Food, Drug, and Cosmetic Act sets a very high standard of safety, offering protection from harmful effects that may occur, not just those that can be shown to have occurred in humans.

  The methylene chloride case represents FDA's tentative conclusion that long-term use as a cosmetic ingredient poses an unacceptably high risk (although probably still too small to measure directly in humans) although the risk from its food use is trivially small. The effect of FDA's proposed rule, if made final, will eliminate

the risk from the cosmetic use of methylene chloride. The protection offered is from an effect that could occur only after years of use. Nevertheless, FDA does follow up to assure that the edicts of its final rules are carried out in the market place. While this type of follow up may not be what GAO has in mind, nor is it a perfect indicator that desired results have been met, it does provide a measure of the effectiveness of FDA regulatory actions.

See comment 22.

- The rating given on prioritization is misleading and appears to be based on invalid premises: 1) it implies that there is a list of known hazards from cosmetic ingredients or food additives that should be addressed in a priority order. 2) It fails to recognize the priority order established in the statute, which gives persons/firms the right to petition the agency and sets timeframes for responses by the agency, thus establishing a chronological priority system. 3) When the agency becomes aware of data indicating a substance used in foods or cosmetics may be hazardous, a prompt risk analysis is undertaken. When two or more such situations arise in roughly the same timeframe, the one concluded first is often dictated by such practical considerations as the availability of pertinent data or the need to resolve difficult inconsistencies, not by a priority scheme.

See comment 23.

- Omission of a rating for toxicological studies indicates a misunderstanding of the nature of the studies evaluated by FDA. All of the biological studies evaluated are toxicological, including the carcinogenicity studies, which are one type of toxicological studies.

See comment 24.

- The rating given epidemiological studies raises the questions of the intent of this GAO evaluation. Is it to evaluate how well FDA does risk assessment or to evaluate the sufficiency of data submitted to FDA? If it is the latter, data submitted to FDA are outside the Agency's control and should not be considered in evaluating the Agency's performance. An agency should consider all information that might have a bearing on an issue when doing a risk analysis. The fact that not all information analyzed is useful is not a fault of the process but is merely an indicator of the thoroughness of review. The GAO report both gives credit to FDA for not relying on epidemiological studies and faults the Agency for having an information gap.

See comment 25.

- The ratings on external expert review raises a question of the internal consistency in the chapter and a second question of the desirability of such expert reviews for every case.

  FDA's analysis of the hazard, dose-response, exposure, and risk characterizaton were all published in the Federal Register with an opportunity for outside comment. In none of the four cases was FDA's analysis submitted for outside review before publication. Yet, GAO rated the first two steps as more than adequate but gave a zero rating for the latter two steps. Apparently, FDA's risk analysis was given credit for standard operating procedures in National Toxicology Program's (NTP) bioassay program that took place before FDA received the data. This confuses the conduct of research outside FDA with the quality of FDA's risk analysis.

See comment 26.

The rating also implies that outside peer review of each step would be desirable. We know of no standing outside group with particular expertise in exposure to both food additives and cosmetic ingredients that would have helped at this early stage. GAO appears to be requiring a new procedural step for analysis that is not required by law, that may be unnecessary, and that could delay completion of an analysis and add to its cost.

See comment 27.
Now pages 38-39

- The summary of Areas for Improvement (pp 2-44 to 2-47) again gives the impression that food and cosmetic products regulated by FDA are of sufficient hazard that an active monitoring program would be able to assess negative health effects and concentrate on the "greatest" hazards. We believe that this is not the case. Also, any monitoring program for the use of food additives and their possible effects in the general population would be prohibitively costly, virtually impossible and would produce results of questionable value. As stated above, FDA does have follow-up programs appropriate to determining whether its regulations are being observed. Since methylene chloride is still under consideration within the agency, follow-up of any nature would be premature.

See comment 28.

- While we agree with GAO that FDA's documentation of the risk management decision process is somewhat weak, we believe the inference drawn from the lack of documentation is out of keeping with the actual results of the risk management decisions reached by FDA in the majority of cases. FDA has been dealing successfully with increasingly visible and pervasive public health risks for the past 80 years. These risks have been associated with an enormous variety of products and have represented varying degrees of hazard and public exposure. Because of the exigencies inherent in dealing with certain risks, it is often necessary to resort to verbal rather than written discussions of the facts and exercise prompt decision-making. Being responsive to public health emergencies sometimes results in gaps in the documentation of the deliberative process. This is not to say that we have not made a careful and reasoned analysis of the risk and possible options. There are functional decision-making systems within FDA, and while procedures for decision-making vary with a situation, Agency components with expertise and responsibility are involved in the process.

In sum, FDA believes Chapter 2 illustrates a confusion between evaluation of process and evaluation of data, does not properly distinguish who is responsible for generating data, does not recognize the utility of a proposal for encouraging generation of good data, and does not distinguish between those data that are necessary for a valid decision and those that may simply provide more knowledge (at considerable cost) but that could not affect a decision. As discussed previously, we believe the checklist rating system as applied to methylene chloride frequently led to misleading and inaccurate conclusions.

## GAO Comments

1. This comment seems to reflect the general issue of the distinction between science and process, which we address in chapter 5. It should be recognized that our criteria are more than a checklist; they also call for an assessment of how well each item was performed.

2. We state in the report that FDA informally uses the guidelines of the Office of Science and Technology Policy. Our rating reflects the fact that FDA has not adopted any formal guidelines. We also report FDA's reasoning for not adopting formal guidelines. However, as we indicate in the report, formal published guidelines are important because, as Milton Russell, former EPA assistant administrator, and Michael Gruber, an EPA official, stated recently, such guidelines "foster a consistent approach across [cases], . . . establish a standard for quality of work and comparison of studies, and help inform the public about how scientific judgments and assumptions have been incorporated into risk assessments." We concur with their conclusion that

> "making decisions about risk in the absence of guidelines may lead to idiosyncratic decisions that cannot easily be explained or defended and that are subject both to accusations of capriciousness and to real or perceived manipulation in the service of political expediency."[1]

Moreover, our criterion for the use of formal guidelines was based in part on a recommendation by the Department of Health and Human Service's (HHS's) task force on health risk assessment that "each PHS [Public Health Service] agency should develop and implement guidelines for the conduct of the health risk assessments it undertakes as a basis for regulatory, service, or educational risk management strategies."[2]

3. We do not state, as HHS's comment implies, that guidelines should be inflexible. In our view, guidelines or criteria should be treated as working documents, subject to change as knowledge grows, that provisionally resolve open scientific issues for the purpose of decisionmaking.

4. The applicable statutes permit considerable discretion. The purpose of decisionmaking guidelines is to publicly articulate how discretion is exercised. Whether decisions can withstand court challenges is not the

[1]Milton Russell and Michael Gruber, "Risk Assessment in Environmental Policy-Making," Science, April 17, 1987, p. 287.

[2]U.S. Department of Health and Human Services, Task Force on Health Risk Assessment, Determining Risks to Health: Federal Policy and Practice (Dover, Mass.: Auburn House Publishing Company, 1986), p. 306.

only or even by itself necessarily the best test of decisionmaking procedures. According to a recent article by Lester Lave, "there is good reason to inform the affected parties and the public of the basis for a decision" because "there is no other process likely to secure public confidence and consent."[3]

5. As HHS notes, the development cost, areas of applicability, and level of specificity of guidelines are important considerations. However, we do not believe it is self-evident that these considerations override the need for guidelines. The general issue concerning the validity and applicability of the criteria we developed is addressed in chapter 5. As we state in the report in chapters 1 and 5, our specific criteria were derived from suggested guidelines published in the risk analysis literature. We make no claim that they are universally applicable; we claim only that they represent the criteria published in the literature. We do believe, however, that the general model of the risk analysis process applies to the examples HHS cites in this comment. Specific criteria would, of course, vary with the types of research required. We also do not claim that our approach applies to all FDA decisions.

6. It is true that a criterion may vary in importance from case to case. We thought this was an important issue and discussed it extensively with our second external advisory panel of risk analysis experts. The consensus of this group was that one could not assign weights for specific criteria or phases of the process for the general case. The panel recommended against assigning variable weights in either general or specific cases. We implicitly assigned a weight of zero by giving no rating for specific criteria or phases of the process that were inapplicable to specific cases. If agencies want to set weights in their assessment process, we believe this would be reasonable.

7. Our procedures for applying the criteria and resolving differences between analysts are commonly recognized in content analysis. We recognize that content analysis is not the most objective of all approaches to measurement, although we believe it is the most objective approach possible in this context. Content analysis is far more systematic than the other methods that have been used to evaluate actual instances of the risk analysis process. We have expanded the discussion of these procedures in response to this comment. The reference is to page 17.

[3]Lester B. Lave, "Health and Safety Risk Analyses: Information for Better Decisions," Science, April 17, 1987, p. 294.

8. We do not suggest that risk is all that is or should be assessed in the risk analysis process. However, balancing risk against benefits or any other factor is a risk management strategy and is thus analytically distinct from risk assessment. Consequently, our model includes the assessment of benefits as part of the research necessary for risk management. Additionally, it is true that fewer and less-specific criteria in our evaluation framework are applicable to assessing benefits than to assessing risks. This is a reflection of the material published in the risk analysis literature, and it is one specific reason we explicitly do not argue that the criteria we used are necessarily exhaustive.

9. The general issue concerning the validity and applicability of the criteria we developed is addressed in chapter 5. It is true that some of the specific criteria presume that the risk source under consideration is a suspected carcinogen. This is also a reflection of the literature. However, as with all the specific criteria, those presuming a carcinogenic risk would be given a weight of zero if they were not applicable and, thus, no rating would be assigned.

10. We state explicitly that with the exception of some generally applicable procedures, we do not generalize from the limited number of cases we examined. The purpose of random selection in the pilot study was to avoid the implication that we had handpicked a "worst-case" example.

11. The statement in this comment that "we believe one should weigh carefully the economic impact of our actions," appears to acknowledge the accuracy of our statement that the agency considered factors besides the risk to public health and safety. But the meaning of the subsequent statement that "Public health and safety are the ultimate determinants of such decisions" is ambiguous. Such ambiguity is the source of our concern that the way the various factors are united in decisionmaking is not clearly articulated.

12. The general issue of the appropriateness of the cases we selected is addressed in chapter 5. FDA's development of the proposed action on methylene chloride includes both risk assessment and risk management activities, which is how we define the risk analysis process. The criticism that we should have selected a more typical case is without merit; FDA objected to the first case we selected (an indirect food additive) because it was atypical, and we subsequently eliminated all cases of that type from the list of candidate cases. FDA specifically included the methylene chloride case on our list. Once we had placed it there, we then

randomly selected it in accordance with the selection procedures we describe in the report. See also comment 19.

13. The general issue of agency accountability is discussed in chapter 5. We recognize that regulatory agencies must rely on scientific research and other data that are beyond their control. Consequently, our approach was not to rate each agency in terms of its implementation of the risk analysis process. Instead, we examined and rated specific cases of the risk analysis process. While we certainly did rate aspects of this process that are beyond the control of the agencies, we did not hold the agencies accountable for them.

14. We do not suggest that FDA should apply our criteria or any others as a checklist. It is reasonable to expect, however, that guidelines that include the equivalent of criteria would be incorporated into standard agency procedures. It is not self-evident that the use of guidelines would inhibit FDA's responsiveness. Rather, guidelines would establish a base for understanding the minimum requirements for conducting adequate or acceptable risk analyses. No matter what the exigency, the violation of basic guidelines or standards for defining acceptable analytical work will result only in questionable conclusions.

15. HHS's comment that the use of guidelines applies only to relatively low-level risks is not well founded. We are referring here to new or modified rulemaking actions in which a risk assessment forms the basis for agency decisionmaking. We know of no FDA rule that was made with such urgency as to prohibit the use of guidelines. The acute problems HHS mentions in this comment apparently refer to enforcement actions in which the hazard is well understood and no risk assessment is necessary. When risk assessments are used to estimate the risk to public health in order to determine whether and how much regulation is necessary, it is the risk assessment that reveals the relative severity of the hazard, and, thus, guidelines for conducting adequate risk analyses apply equally to relatively high and low risks.

16. This comment refers to the succeeding studies we may conduct. As we have noted in our report, one of the purposes of this pilot study was to gain experience in applying the criteria. We are assessing our experience and the forms that a further evaluation might take. We disagree that detailed evaluation criteria lead to subjective results. In our view, the more specific the criteria, the less room there is for unexplained subjective judgments to bias the results. For this reason, we developed the most specific criteria feasible.

17. This comment refers to the general issues, discussed in chapter 5, of the distinction between science and process and of our consideration of the factors agencies must consider in decisionmaking. HHS's frequent allusion to scientific judgment is perplexing. All scientific judgments are exercised according to rules that, in principle, can be articulated. In the interest of open decisionmaking, it is preferable that rules be explicitly articulated and included in the written record.

It is true that not all the issues some analysts might consider important to the quality of scientific work are addressed in our criteria. As we state in the report, we defined adequacy in terms of the guidelines suggested in the risk analysis literature, and we explicitly state that they are not necessarily exhaustive. No firm consensus has been reached concerning the specific requirements for adequacy. We provisionally resolved this issue through recourse to the literature. In addition, since the literature does not include explicit attempts to define adequate science, we did not attempt a definition. However, we do recognize the importance of this comment. In response, we have included a tentative definition, based on the literature, of what we mean by "adequate scientific work," and, as noted above, we have expanded the discussion of our methodological approach to the assessment of the adequacy of scientific work. The reference is to pages 15-17.

18. This comment seems to refer to the general issues of the distinction between science and process and the validity and applicability of our criteria, which are discussed in chapter 5. We did not rely solely on a review of the documentation of the process, although the written record is the prescribed means for demonstrating the quality of scientific research. Where we had questions concerning the quality of the scientific work after reviewing the research reports, we interviewed the responsible agency officials. However, a scientific research report should be sufficient to show not only that accepted methods were used but also whether they were used correctly. As we state in the report, the role of our criteria was more than a checklist; our design called for an assessment of how well each item was performed.

For example, although confidence intervals may be calculated for any sample data, their correct use requires that the empirical population to which they refer is the same as the actual population of concern. An evaluation that merely noted that confidence intervals were calculated or reported and did not determine that they actually reflected the degree of uncertainty of the estimates would be of little value. We established our criteria at the most specific and detailed level feasible in

order to reduce as much as possible the judgment inherent in applying standards of adequacy. The judgments that necessarily remained were based on the discussions of the relevant issues and the rationale for proposed guidelines in the risk analysis and general scientific literature.

We do not agree that any phase of the risk analysis process applicable to a given risk management approach could be omitted without serious consequences. For example, in cases in which the level of risk is balanced against other factors, omitting either dose-response assessment or exposure assessment would invalidate the entire risk assessment. Similarly, even though evaluations of the effectiveness of a regulatory action may not always be technically feasible, their absence means that there is no way to determine whether risk management has reduced a risk.

HHS's comment also confuses our use of criteria to evaluate the quality of risk assessment work with an agency's use of guidelines to direct risk analyses. Agency guidelines alone cannot ensure that high-quality scientific work is actually performed. The purpose of guidelines is to establish a standard for quality, ensure consistency, and articulate the use of scientific policy and assumptions.

HHS's statement that most of FDA's decisionmaking is not documented is especially alarming. We know of no valid reason for not documenting formal regulatory decisions or their procedures or decision rules. We agree that evaluating a decisionmaking process is difficult, even under the best of circumstances, but as we state in the report, the purpose of establishing decisionmaking guidelines, which FDA does not now have, is to carefully consider the institutional review and decisionmaking process, establish procedures including documentation, and ensure the separation of scientific and policy issues.

19. We discuss the general issue of the appropriateness of the cases we selected in chapter 5. We state explicitly that we do not generalize about the agency's implementation of the risk analysis process because, as this comment observes, each case has unique aspects that prevent it from being fully representative of an agency's activities. Accordingly, every action could be shown to be nonstandard in some sense.

All the unique elements HHS mentions in this comment are discussed in the report, and we make no recommendations. Moreover, the criticism that we should have selected one of the "many" premarketing food additive petitions FDA reviews each year is without merit. Initially, we

did select such a case from a list of candidate cases reviewed by FDA. But FDA strenuously objected that the indirect food additive we selected was not appropriate because the Delaney clause did not apply to it and its risk assessment reviews were cursory and, thus, not standard. FDA requested that we consider only direct food additives and specifically asked us to place the proposed methylene chloride action on our list of candidates for evaluation.

FDA also objected to our consideration of cases completed prior to 1984, when the agency instituted a "more standard" risk assessment process. But from 1984, when FDA's more standardized process was implemented, to the end of 1985, when our period of observation ended, FDA promulgated only two final direct food additive regulations, both of which were the result of industry petitions. With methylene chloride, this made three actions to choose from. We eliminated one of these because it was the subject of another congressionally requested GAO review. We then selected methylene chloride randomly from the remaining two actions, in accordance with the case selection procedures we describe in the report.

20. This passage, now on pages 21-22, has been modified.

21. We did not rate monitoring and evaluation in the methylene chloride case because the regulation has not yet been promulgated. It was not our intention to imply that epidemiological studies to evaluate risk reduction are possible for food additive regulations. Our remarks were limited to observing that the monitoring activities that are performed do not permit an assessment of regulatory effectiveness. We have clarified this passage, now on page 36.

22. We discuss in the report the chronological order in which FDA responds to premarketing approval petitions, and we recognize that the order in which risk analyses are concluded is determined by the difficulty of the cases as well as the order in which they are initiated. The problem we identify in the report is that priority-setting is not used in any of these procedures. FDA's procedures do not ensure that to the extent feasible, the potential hazards of substances currently in use are analyzed in order of potential severity. The enabling legislation authorizes this priority-setting; it does not require FDA to wait until it "becomes aware of data" indicating a potential hazard. FDA has developed a sound plan for prioritization, but it has failed to implement it

HHS's comment also confuses known hazards with potential hazards. W do not imply that there is a list of known hazards that should be addressed with specific priority. The priority-setting we refer to deals with potential hazards and, according to FDA's own plan, is to "assess relative concerns for [food] additives so that it [FDA] may devote more o its resources to those additives that are of the highest potential public health concern." The agency is not now doing this. FDA's implementing its plan, however, would mean that the agency would not have to wait until it happened to become aware of data indicating that a substance used in foods or cosmetics might be hazardous. A priority-setting syste would permit the agency to take the initiative.

23. This comment is based on our use of the term "toxicological study" to distinguish "tissue bioassays" from the analysis of the potential biological effects of the various chemical properties of substances such as those examined in "quantitative structure activity relationship analyses." We have modified the relevant passages to include more specific terminology. We assigned no rating for the absence of this possible com ponent of the risk analysis process, because we did not view its omissic as a deficiency. Our decision reflects the modest results of this type of analysis and our view that it is not applicable when better data are available. The reference is to pages 23 and 99-100.

24. The general issue of agency accountability is addressed in chapter 5 While we do rate aspects of the general risk analysis process that are beyond agency control, we do not hold the agency accountable for ther See also comment 13.

25. The criteria on external expert review reflect the view expressed i the literature that reviews by external experts should be conducted early in the risk analysis process. Our intention was to acknowledge th the research FDA used early in the process was reviewed by experts outside FDA. We recognize that it is not appropriate to appear to credit FDA for routine activities of the National Toxicology Program that FDA did not request. However, since an external FDA review, in addition to the program's review, would have served no useful purpose, we now consider the external expert review for these phases of the process ina plicable in this case. The passage, now on pages 23-26, has been modified.

26. We do not suggest that a standing group of external experts does exist. Our rating reflects the view expressed in the risk analysis literature that such bodies should exist, that they are necessary, and that an

delay they might entail would be worth the confidence in the results. A review of FDA's work by any appropriate outside panel of experts would have fulfilled the criterion. We also see no reason why the establishment of demonstrated procedures for improving the quality of scientific work should be restricted to those specifically required by law.

27. This comment confuses setting priorities for in-depth analysis of potential hazards with in-depth risk assessment to determine if or what action is necessary, and the evaluation of previous agency actions. We do not suggest that one monitoring program could accomplish all three of these tasks. We now specifically state in the report that FDA monitors compliance with its actions. The reference is to page 35.

28. We do not agree that there is a conflict between responsiveness and full documentation of decisionmaking for rulemaking actions.

# Comments From the Department of Labor

Note: GAO comments supplementing those in the report text appear at the end of this appendix.

**U.S. Department of Labor**

Assistant Secretary for
Occupational Safety and Health
Washington, D.C. 20210

APR 9 1987

The Honorable Richard L. Fogel
Assistant Comptroller General
Human Resources Division
U.S. General Accounting Office
Washington, D.C. 20548

Dear Mr. Fogel:

This is in response to your letter of March 12 to Secretary of Labor William E. Brock requesting comments on the proposed report of the General Accounting Office (GAO) on health risk analysis.

See comment 1.

GAO has attempted to characterize the risk analysis process as it is practiced in three agencies: the Occupational Safety and Health Administration (OSHA), the Environmental Protection Agency (EPA) and the Food and Drug Administration (FDA). Its method for this characterization was to set up a "model" risk analysis process, and then compare one regulatory action from each agency with the model.

We support GAO's efforts in this study to improve regulatory decisionmaking. We have a number of comments, and technical corrections to offer, however. These comments follow.

Now pages 54-55.

o In the discussion on "regulatory decisionmaking" (pages 3-35 and 3-36 of the draft report), suggestions are made for improvements in the presentation of OSHA's "risk management" decisions. While the OSHA process is described as resulting in clear risk management options for the decisionmaker, the report notes that internal documents do not discuss the uncertainties inherent in risk estimates; that assumptions used in making the risk assessments are not presented; and that internal documents do not contain "citations of the studies used in the documents," are "excessively brief" and do "not fully discuss key issues." The report cites as a specific example that discussion of linear versus quadratic dose/response models was not presented.

See comment 2.

The risk analysis performed for inorganic arsenic, which was selected by GAO for its case study of OSHA's risk analysis process, was done early in the decade. OSHA's risk analysis procedures have changed since then. Detailed documentation or

risk assessment now accompanies all rulemaking actions. Recent examples to which GAO may wish to refer include benzene, formaldehyde, and ethylene oxide.

See comment 3.

- On page 3-7 of the draft report, GAO's description of OSHA's risk analysis process is inaccurate. OSHA's risk analysis consists of the following three steps: (1) critical review of the scientific literature, (2) quantitative risk assessment with appropriate studies, including dose-response assessment, and (3) determination of the significance of the risk by evaluating the information in steps (1) and (2) from both scientific and policy perspectives.

See comment 4.
Now page 12.

- GAO includes as a final step in its model risk analysis process an evaluation of whether a regulatory action had the anticipated effects (page 1-4 of the draft report). In measuring agency performance against this yardstick, GAO looked at whether the agencies had checked on the efficacy of their regulations once they were in place. GAO noted that OSHA had not done such a review of the effect of its standard for inorganic arsenic (page 3-36).

Now page 55.

To have done such an evaluation, would have required OSHA to start a prospective epidemiologic study of smelter workers to see if cancer incidence dropped after the regulation went into effect and to have monitored the economic impact of the standard. OSHA believes that while prospective studies of this kind might prove valuable, present resource constraints make such studies unlikely except on a very limited scale. In addition, the research nature of such studies makes NIOSH a more likely candidate for carrying them out.

Now page 53.
See comment 5.

- On page 3-31 of the draft report, GAO criticizes OSHA for the lack of formal economic analysis guidelines. While it is true that there was no written policy for economic and regulatory analyses at the time the arsenic decision was completed, Executive Order 12291 was issued in 1981. Since the GAO refers to this problem in the past tense, we assume that it realizes that the problem has been corrected.

Now pages 48-49.
See comment 6.

- In the suggestions for improvements in exposure assessment (pages 3-20 through 3-22), GAO is critical of the epidemiologic studies on which OSHA's exposure assessments for inorganic arsenic are based. It appears to us that GAO misread OSHA's risk assessment with respect to exposure assessment. OSHA believes that its reconstruction of past exposures to workers for the period of the studies was the most accurate that could be done. We believe that the epidemiologic studies for arsenic were of the highest quality and yielded excellent data for regulatory purposes.

See comment 7.

- In the first sentence of the second paragraph on page 3-8 of the draft report, the word "alternatives" should be changed to "regulatory." Before the last sentence in the same paragraph, the following sentence should be added: "Standards are set to eliminate significant risks, within the confines of feasibility."

See comment 8.

- On page 3-10, first paragraph, GAO states that no hazard identification work was performed for arsenic. That is inaccurate. Hazard identification work was done in 1978 (see 43 CFR 19584).

See comment 9.

- GAO's discussion of OSHA's use of epidemiologic studies (page 3-16 through page 3-20) makes it appear that weaknesses in those studies are attributable to OSHA when they are, in fact, weaknesses in the studies themselves. OSHA cannot improve the technical literature.

See comment 10.

- On page 3-17, second paragraph, GAO inaccurately characterizes sulfur dioxide and lead fumes as carcinogens. We suggest changing the word "carcinogenic" to "hazardous" and deleting the word "potentially."

See comment 11.

- The last sentence, second paragraph, page 3-34, should read: "This study, according to one industry group, suggested that there may be a threshold level for arsenic."

See comment 12.

- On page 3-4, second paragraph, the last sentence in GAO's discussion of the Supreme Court's ruling should be changed to read: "Instead, the Court ruled that once OSHA determines a substantial reduction in significant risk will occur, it must reduce exposures to the lowest feasible level."

See comment 13.

- The following should be added at the end of the third sentence in the second paragraph on page 3-5; "...risk management since feasibility must be considered as well as substantial reduction in risk."

Thank you for the opportunity to comment on the proposed report. If you have any questions, please call my Executive Assistant, Ms. Jan Williams, on 523-7480.

Sincerely,

John A. Pendergrass
Assistant Secretary

## GAO Comments

1. This statement by the Department of Labor (DOL) represents a misunderstanding of our approach. We do not attempt to characterize risk analysis as it is practiced in three agencies, and we did not attempt to define a "model" or an ideal risk analysis process. We developed a descriptive model of the generic risk analysis process based primarily on the NAS attempt to do the same. We then evaluated three risk analysis cases, using criteria derived from attempts in the risk analysis literature to propose guidelines for conducting adequate risk analyses. Our approach was more tentative and exploratory than is reflected in DOL's statement.

2. The general issue concerning the appropriateness of the cases we selected is addressed in chapter 5. We recognize that risk analysis policies and procedures are constantly evolving, and we make no claim that the 1983 inorganic arsenic action represents OSHA's overall efforts. However, we must point out that the 1983 inorganic arsenic action was on the list of actions OSHA validated as appropriate for this study at the outset of our project.

3. This passage, now on page 43, has been changed to reflect this point.

4. DOL does not state that an empirical evaluation of the effects of the inorganic arsenic regulation is not technically feasible. The absence of such evaluations, especially when they are technically feasible, is a serious deficiency in the risk analysis process, because it means that no feedback is possible from expensive regulatory actions and, thus, there is no means of learning if the regulations are reducing risks as intended. In addition, this deprives researchers of an important means of improving predictive models. DOL's comments about resource constraints and the proper organization for conducting evaluations are important considerations but they were beyond the scope of our pilot study.

5. We do not see any connection between executive order 12291 and the presence or absence of formal OSHA guidelines for economic analysis for determining feasibility. The executive order requires benefit-cost analysis for major rules, but OSHA does not use benefit-cost analysis to support its regulatory actions. Since OSHA's standards are set to eliminate significant risks within the confines of economic and technological feasibility, formal guidelines for economic analysis would be anticipated to specify what is required to demonstrate economic feasibility.

6. The general issues of the distinction between science and process and agency accountability are discussed in chapter 5. We recognize that

OSHA's effort to reconstruct exposures was as good as circumstances allowed; OSHA should recognize that the quality of the exposure assessment was nevertheless low. We make no suggestions or recommendations for improving the exposure assessment. Our evaluation is limited to identifying areas in the specific cases where improvement would have been useful. This does not necessarily imply that the agency is at fault or would be able to effect improvements.

7. The passage, now on page 43, has been changed.

8. We do not state that hazard identification was not performed for inorganic arsenic, and we do not assign a low rating for this phase of the process. We state that no hazard identification was performed to support the action we selected, which stemmed from a limited judicial remand. Consequently, hazard identification was not applicable in this case. In order to develop a sampling frame, it was necessary to strictly define what we meant by "risk analysis case." We elected to define it in terms of the work required to support a specific agency action published in the Federal Register. The hazard identification work was performed for the 1978 action. That portion of the work was not remanded by the court and thus was not relevant to the 1983 action we selected. The passage referred to is now on page 44.

9. The general issue of agency accountability is discussed in chapter 5. We state that OSHA relied on 13 epidemiological studies, and table 3.7 shows that OSHA neither conducted nor contracted for any of the 13. We have also further clarified this point, now on page 44.

10. We did not characterize sulfur dioxide or lead fumes as carcinogens. We stated that they were treated as potential carcinogens in the research we referred to. These substances were controlled in the research to help prevent them from being confounded with arsenic in the attribution of greater mortality from cancer than expected. The research concluded that people with the highest exposure to arsenic and moderate to heavy exposure to sulfur dioxide were at the greatest risk, but the research could not fully separate their effects. We are aware that research on smoking has not identified sulfur dioxide as a carcinogenic constituent of cigarette smoke. The research we referred to in our report found no effect of exposure to lead fumes but did not indicate the specific lead compounds that were involved. The International Agency for Research on Cancer has determined that evidence for some lead compounds is sufficient to define carcinogenicity in animals. In order to avoid creating the impression that these substances have been firmly

established as human carcinogens, we have deleted reference to specific substances in the passage, now on page 48.

11. The statement, now on page 54, has been changed.

12. The statement, now on page 41, has been changed.

13. The statement has been added. The reference is to page 41.

# Comments From the Environmental Protection Agency

Note: GAO comments supplementing those in the report text appear at the end of this appendix.

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF
POLICY, PLANNING AND EVALUATION

APR 13 1987

Mr. J. Dexter Peach
Assistant Comptroller General
Resources, Community and Economic
Development Division
General Accounting Office
441 G Street, N.W.
Washington, D.C. 20548

Dear Mr. Peach:

This letter is in response to your letters of March 11 and March 30, 1987, concerning the General Accounting Office (GAO) draft report, "Health Risk Analysis: Technical Adequacy In Three Selected Cases." Previously EPA and GAO staffs met to discuss the report, and I am now providing an official Agency response to the report. Under separate cover, I will send to you two detailed documents relating to the report.

After a thorough review of the report, I believe that evaluation of the decision mechanism for a technology-based standard, such as the New Source Performance Standards-Onshore (NSPS-Onshore) reviewed in the report, is inconsistent with a report on the technical adequacy of health risk analysis. My reasoning for this belief follows.

The purpose of the GAO report is, as stated in its introduction, "...to address the adequacy of the risk analysis work supporting federal health regulations," and more specifically "...to examine how the risk analysis process is conducted." To accomplish this, GAO "developed criteria for evaluating the adequacy of the risk analysis work...." (Emphasis added). However, the selection of the NSPS-Onshore was totally inappropriate because NSPS are not based upon risk analyses.

See comment 1.

In a NSPS standard, risk analysis does not lead the decision and/or is not integrated into the decision; therefore, any evaluation of the technical adequacy of the health risk analysis is irrelevant to the decision-making process or to the Agency's use of risk analysis in general. Preparers of the report imply an understanding of this when they express discomfort about the lack of integration of health risk analysis into the decisionmaking progress. The report attempts to apply GAO's criteria for evaluating risk analysis work to NSPS-Onshore which involves no risk analysis. Not surprisingly, the results are confusing, misleading, and erroneous.

See comment 2.

GAO evaluates both process and scientific aspects. They dilute and confuse what they could say constructively about each one, independently. Furthermore, in their criticism of the scientific analyses, they do not identify which may be generic problems in each Agency's scientific procedures or processes and which are due to the presence or absence of specific information needed for that analysis. For instance, did Occupational Health and Safety Administration's (OSHA's) careful and complete analysis of arsenic risk occur because of OSHA's own procedures or because there is an extensive data base about arsenic?

See comment 3.

OSHA also received the most praise for conduct of its risk balancing and cost/benefit analysis. How much of that praise resulted because of the quality of OSHA's work, and how much because this case was the only example of a risk balancing regulation in the study? It would be better to have evaluated the analysis behind risk-balancing regulations for several agencies, although it still might be difficult to make any valid generalizations.

See comment 4.

The Food and Drug Administration (FDA) is criticized about its lack of guidelines; however, EPA staff is aware of FDA's scientific analysis on methylene chloride, and believes it is done carefully and competently. GAO also did not distinguish between the FDA's handling of the cosmetic issue (where methylene chloride is a major ingredient) and decaffeinated coffee (where it is a process contaminant). Furthermore, the regulatory process is not complete on methylene chloride, and the GAO analysis could become outdated very quickly.

See comment 5.

Another inconsistency is the choice of health effects. Methylene chloride and arsenic are regulated as carcinogens; ozone is not. Some of the differences in the scientific analyses could be caused by that basic difference.

The draft report presents statements that demonstrate a misunderstanding of the Clean Air Act and related EPA programs. Examples of these are:

See comment 6.

--A lack of understanding of the intent of section 111 and the case law which has served to clarify the appropriate basis for these standards. (GAO describes section 111 as "the technological control approach to risk management." This is an erroneous perspective.)

See comment 7.

--An incorrect understanding of the EPA internal review process. (GAO identified it as solely administrative, a characterization that is not at all accurate.)

See comment 8.

--A lack of understanding of the NSPS decisionmaking process and the role of costs in NSPS decisions. (The GAO report erroneously considers cost effectiveness or cost/benefit to be the sole basis for NSPS decisions and then analyzes them from a risk management perspective.)

Thank you for the opportunity to comment officially on this report. I hope that this letter and the two documents that I will send to you aid GAO when revising this report.

Sincerely yours,

John M Campbell

John M. Campbell
Acting Assistant Administrator

## GAO Comments

1. The general issue concerning the appropriateness of the cases we selected is addressed in chapter 5. This criticism is based in part on the difference between the way we and EPA use the term "risk analysis," but it is based primarily on EPA's interpretation of the requirements for new source performance standards. Although the risk analysis literature does not always use terms consistently, we use the term "risk analysis" to refer to the general process of regulatory decisionmaking that includes both risk assessment and risk management activities. EPA uses the term risk analysis to refer only to what we call "risk assessment."

EPA's criticism is based primarily on the unstated argument that since risk assessment work is not conducted to fulfill specific section 111 requirements for developing a new source performance standard, the standard is not based on risk analysis (risk assessment in our model). However, EPA does not regulate sources of harmless substances; it regulates sources of substances that the administrator has determined are harmful to human health or welfare. This determination requires risk assessment work and, thus, new source performance standards are indeed based on risk assessment (risk analysis according to EPA's usage). It is not a relevant concern that the necessary risk assessment work is usually conducted to fulfill the specific requirements of other sections of the act, because that work remains the basis for regulating sources of air pollution under section 111. EPA explicitly recognized the connection between new source performance standards and the supporting risk assessment work performed under other sections of the act in the case we selected, a connection EPA seems now to deny. EPA stated in the background information document for promulgation that "The administrator clearly determined the need to regulate VOC [volatile organic compounds] to protect public health and welfare in the EPA publication 'Air Quality Criteria for Ozone and Other Photochemical Oxidants,'" which was published in accordance with section 108 of the act. Moreover, we fully explained our design at the outset of this project and EPA did not object to it.

We do not express "discomfort" about the lack of integration of risk assessment work in this case. Since the technological control approach to risk management does not require a quantitative assessment of the magnitude of risk, we did not expect to see such analyses. The hazard identification portion of risk assessment is all that is explicitly required. If, as EPA seems to suggest, the presence of quantitative risk characterizations completely defined the risk analysis process, then the technological control approach would be excluded by definition. However, this conceptualization contradicts the literature. We do express concern that

the risk assessment work that was used in developing the benefit-cost analyses that influenced EPA's decisionmaking for this case was not included in the public record.

2. The general issues of the distinction between science and process and agency accountability are addressed in chapter 5. We believe it is essential to examine both the scientific and process dimensions of risk cases in order to determine the degree of adequacy. We do not believe we dilute or confuse either the scientific or the process dimensions. And, while we certainly do evaluate aspects of the general risk analysis process that are beyond agency control, we neither blame nor praise the agencies for factors that are beyond their control.

3. The general issue concerning the appropriateness of the cases we selected is addressed in chapter 5. We evaluated the OSHA case in the same fashion that we evaluated the EPA case; we applied our evaluation criteria where they were appropriate. We considered the type of risk analysis we were evaluating, as well as other factors, in deciding the appropriate criteria. We explicitly recognize that different types of scientific information are required, according to the assumptions of the different approaches to risk management. Because balancing is acknowledged only in the OSHA case, it is our only example of a risk-balancing regulation. However, the OSHA case received its favorable review not because it was an example of the risk-balancing approach but, rather, because of the extent to which it met our evaluation criteria. Moreover, the balancing OSHA conducts does not include benefit-cost analysis. It would have been better to evaluate all risk-balancing regulations only if our purpose had been to compare how well risks are balanced with other factors within each agency. This was not our purpose. Additionally, we would not have been able to generalize, nor did we generalize, from only three cases, regardless of their type.

4. By and large, we concluded that FDA's assessment of the risks of methylene chloride was careful and competent, despite the lack of formal guidelines. The criticism that we do not distinguish between the uses of methylene chloride that FDA considered is not well founded. Throughout our discussion, we distinguish between the use of methylene chloride in decaffeinating coffee and in cosmetics. We recognize that the risk analysis process is not complete for methylene chloride and qualify our remarks in the report accordingly.

5. This comment reflects the general issue concerning the validity and applicability of our criteria, which is addressed in chapter 5. EPA seems

to be suggesting that our criteria apply to risk assessments for carcinogens but not other harmful substances such as ozone. While some of our criteria do assume a carcinogenic risk source, the general model of the risk analysis process and many of the specific criteria apply to noncarcinogens as well. Of course, specific criteria have to reflect specific types of research. We did not apply criteria that were not applicable. We point out instances in which the risk analysis literature does not contain criteria for specific types of research we encountered in the cases, and we note instances in which scientific analyses differ because the threats to health differ in the three cases.

6. The general issue concerning the appropriateness of the cases we selected is addressed in chapter 5. Before we selected this case for evaluation, we provided EPA with a description of our design and a list of cases that we thought represented the technological control approach to risk management. Several EPA officials, including a representative of the office of the general counsel, approved the cases on our list as representing the technological control approach to risk management.

EPA's assertion that new source performance standards do not represent the technological control approach to risk management is without merit. The language in section 111 has often been categorized in the literature as representing the general technological control approach, also referred to as the "technology-based" approach, to risk management. The literature also indicates that this approach necessarily requires a consideration of costs as well as technical feasibility, although the consideration of cost is sometimes unstated.

We do not believe that the intent of new source performance standards affects the characterization of a risk management approach. EPA's comment apparently refers to the congressional intent to establish minimum national standards regardless of local air quality, in order to prevent pollution and the relocation of sources of air pollution to relatively clean areas. We recognize that new source performance standards are not the only approach to managing the risks associated with air pollution as required by the Clean Air Act and that much more extensive risk assessment work is conducted for air pollutants under other circumstances. EPA should recognize that minimum national standards based on technology are part of the overall strategy for managing the risks associated with air pollution referred to in the Clean Air Act.

EPA's comment refers to case law as demonstrating that new source performance standards are not technological control actions. As we state in

the report, case law has clarified what it means to consider cost by establishing that benefit-cost analysis is not required by section 111. If benefit-cost analysis were the sole basis of new source performance standards, then the standards would represent the balancing approach to risk management rather than the technological control approach. However, in the past EPA has not indicated any intention of basing new source performance standards on a balancing approach; using this approach would require that the agency include the research supporting it in the public record. The cost-related research EPA places on record is limited to the economic effect of regulation, which is consistent with technological control but not with balancing. That the comment addresses the categorization of new source performance standards and not our findings seems to indicate agreement that statements made to the public in this case were misleading because they did not acknowledge that benefit-cost analysis did influence decisionmaking.

7. We do not state that EPA's internal review is solely administrative. We state specifically that the scientific and technical aspects of a new source performance standard are reviewed during what is primarily an administrative review. Our rating reflects only the fact that there is no distinct internal expert review. While we state that our criteria represent what has been proposed in the literature, we do not argue that all the criteria are necessarily required for adequacy. One of the several considerations we are reviewing for any follow-on to this study is whether it is necessary that internal expert and administrative reviews be distinct. Since there are arguments on both sides, we decided not to alter our criteria at this time.

8. The general issue of the factors considered in decisionmaking is addressed in chapter 5. Our understanding of the decisionmaking process for the specific new source performance standard we examined was obtained from official EPA documents and the statements of the several EPA representatives we interviewed. We state specifically that benefit-cost analysis influenced but did not completely determine the decision; we have not stated that it was the sole basis for decisionmaking. EPA's comment here seems to acknowledge the accuracy of our statement. The problem, as we see it, is that EPA has not previously acknowledged that such analyses play any role in decisionmaking.

# Bibliography


Allera, E. J. "An Overview of How the FDA Regulates Carcinogens under the Federal Food, Drug, and Cosmetic Act." Food Drug Cosmetic Law Journal, 33 (1978), 59-77.

Arthur D. Little, Inc. The Economic Impact of Proposed OSHA Airborne Arsenic Standards. Report to American Smelting and Refining Company. N.p.: June 1975.

Arthur Young and Company. Technological Feasibility Analysis and Inflationary Impact Statement for the Proposed Standard for Inorganic Arsenic (40 FR 3392). Prepared for the Occupational Safety and Health Administration. Washington, D.C.: April 28, 1976.

Atkisson, Arthur A., Michael E. Kraft, and Lloyd L. Philipson. Risk Analysis Methods and Their Employment in Governmental Risk Management. Technical report 85-1398-1 prepared for the National Science Foundation, Policy Research and Analysis Division, grant PRA-8007228. Redondo Beach, Calif.: J. H. Wiggins Company, February 1985.

Blejer, Hector P., and William Wagner. "Case Study 4: Inorganic Arsenic-Ambient Level Approach to Control of Occupational Cancerigenic Exposures." Annals of the New York Academy of Sciences, 271 (1976), 179-86.

Bogen, Kenneth T. Coordination of Regulatory Risk Analysis: Current Framework and Legislative Proposals. Washington, D.C.: Congressional Research Service, June 19, 1981.

Calabrese, Edward J. Methodological Approaches to Deriving Environmental and Occupational Health Standards. New York: John Wiley and Sons, 1978.

Cheney, David W. Risk Assessment in Health and Environmental Regulation. Washington, D.C.: Congressional Research Service, September 14, 1984.

Chu, Kenneth C. "Addendum B: Risk Assessment for Inorganic Arsenic." Prepared for the Occupational Safety and Health Administration. Washington, D.C., January 14, 1981.

Clement Associates, Inc. Assessment of Human Respiratory Cancer Risk Due to Arsenic Exposure in the Workplace. Prepared for the Occupational Safety and Health Administration. Washington, D.C.: September 5, 1980.

Congressional Research Service. Risk: Assessment, Acceptability, and Management. Prepared for the House Committee on Science and Technology, Subcommittee on Science, Research, and Technology. Washington, D.C.: November 1981.

———. A Review of Risk Assessment Methodologies. Prepared for the House Committee on Science and Technology, Subcommittee on Science, Research, and Technology. Washington, D.C.: March 1983.

Conservation Foundation. Risk Assessment and Risk Control. Washington, D.C.: 1985.

Consultants in Epidemiology and Occupational Health, Inc. Arsenic Risk Assessment: Critique and Alternative. Prepared for the Arsenic Project Panel, Chemical Manufacturers Association. Washington, D.C.: June 2, 1982.

Dodge, Christopher H., and Robert L. Civiak. Risk Assessment and Regulatory Policy. Washington, D.C.: Congressional Research Service, August 24, 1981.

Dubose, D. A., J. I. Steinmetz, and G. E. Harris. Frequency of Leak Occurrence and Emission Factors for Natural Gas Liquid Plants. Research Triangle Park, N.C.: U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, July 1982.

Eaton, W. S., et al. Fugitive Hydrocarbon Emissions from Petroleum Production Operations. Washington, D.C.: American Petroleum Institute, March 1980.

Enterline, Philip E., and Gary M. Marsh. "Mortality Studies of Smelter Workers." American Journal of Industrial Medicine, 1 (1980), 251-59.

———. "Mortality Among Workers Exposed to Arsenic and Other Substances in a Copper Smelter." Presented at the Arsenic Symposium, National Bureau of Standards, Gaithersburg, Maryland, November 4-6, 1981.

Erickson, D. G., and V. Kalcevic. Organic Chemical Manufacturing. Vol. 3. Storage, Fugitive, and Secondary Sources. Research Triangle Park, N.C.: U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, September 1980.

Fischhoff, Baruch, et al. Acceptable Risk. New York: Cambridge University Press, 1981.

Higgins, Ian, et al. "Influence of Arsenic Exposure and Smoking on Lung Cancer Among Smelter Workers: A Pilot Study." American Journal of Industrial Medicine, 2 (1981), 33-43.

Hill, A. B., and E. L. Faning. "Studies in the Incidence of Cancer in Factory Handling Inorganic Compounds of Arsenic. I. Mortality Experience in the Factory." British Journal of Industrial Medicine, 5 (1948), 1-6.

Holsti, Ole R. Content Analysis for the Social Sciences and Humanities. Reading, Mass.: Addison-Wesley Publishing Company, 1969.

Hustvedt, K. C., R. A. Quaney, and W. E. Kelly. Control of Volatile Organic Compound Leaks from Petroleum Refinery Equipment. Research Triangle Park, N.C.: U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, June 1978.

Kasper, Raphael. "Cost-Benefit Analysis in Environmental Decisionmaking." The George Washington University Law Review, 45:5 (1977), 1013-24.

Lave, Lester B. "Health and Safety Risk Analyses: Information for Better Decisions," Science, April 17, 1987, pp. 291-95.

———-, ed. Quantitative Risk Assessment in Regulation. Washington, D.C.: Brookings Institution, 1982.

Lee, Anna M., and Joseph F. Fraumeni, Jr. "Arsenic and Respiratory Cancer in Man: An Occupational Study." Journal of the National Cancer Institute, 42:6 (1969), 1045-52.

Lee-Fieldstein, Anna. "Arsenic and Respiratory Cancer in Man: Follow-up of an Occupational Study." Presented at the Arsenic Symposium, National Bureau of Standards, Gaithersburg, Maryland, November 4-6, 1981.

Lippmann, W., and C. Merz. "A Test of the News." The New Republic, special supplement, 23 (1920), 1-42.

McCaffery, David P. OSHA and the Politics of Health Regulation. New York: Plenum Press, 1982.

Martin, James G. "The Delaney Clause and Zero Risk Tolerance." Food Drug Cosmetic Law Journal, 34:1 (1979), 43-49.

Milham, S., Jr., and T. Strong. "Human Arsenic Exposure in Relation to a Copper Smelter." Environmental Research, 7 (1974), 176-82.

Moreau, David H. "Quantitative Assessments of Health Risks by Selected Federal Regulatory Agencies: A Review of Present Practices with Special Attention to Non-Carcinogenic Substances." Manuscript prepared for the Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, N.C., October 1980.

Muller, Frank G. "Benefit-Cost Analysis: A Questionable Part of Environmental Decisionmaking." Journal of Environmental Systems, 4:4 (Winter 1974), 299-307.

———. "Benefit-Cost Analysis—A Necessary Part of Environmental Decisioning—A Reply." Journal of Environmental Systems, 7:3 (1977-78), 293-95.

National Academy of Sciences, National Research Council. Risk Assessment in the Federal Government: Managing the Process. Washington, D.C.: National Academy Press, 1983.

Office of Technology Assessment. Assessment of Technologies for Determining Cancer Risks From the Environment. Washington, D.C.: June 1981.

Ott, Marvin G., et al. "Respiratory Cancer and Occupational Exposure to Arsenicals." Archives of Environmental Health, 29 (1974), 250-56.

Pinto, Sherman S., et al. "Arsenic Trioxide Absorption and Excretion in Industry." Journal of Occupational Medicine, 18 (1976), 677-80.

Pinto, Sherman S., et al. "Mortality Experience in Relation to a Measured Arsenic Trioxide Exposure." Environmental Health Perspective, 19 (1977), 127-30.

Porter, Donna V. The Delaney Clause: Current Application and Proposed Changes. Washington, D.C.: Congressional Research Service, June 18, 1984.

Regens, James L., Thomas M. Dietz, and Robert W. Rycroft. "Risk Assessment in the Policymaking Process: Environmental Health and Safety Protection." Public Administration Review, 43 (March-April 1983), 137-45.

Rencher, Alvin C., and Melvin W. Carter. "A Retrospective Epidemiological Study at Kennecott's Utah Smelter." Brigham Young University, Center for Business and Economic Research, Provo, Utah, April 1971.

Russell, Milton, and Michael Gruber. "Risk Assessment in Environmental Policy-Making." Science, April 17, 1987, pp. 286-90.

Solomon, Kenneth A., William E. Kastenberg, and Pamela F. Nelson. Dealing with Uncertainty Arising Out of Probabilistic Risk Assessment. Santa Monica, Calif.: Rand Corp., September 1983.

Steigerwald, B. J. Emissions of Hydrocarbons to the Atmosphere from Seals on Pumps and Compressors. Report 6, Joint District, State, and Federal Project for the Evaluation of Refinery Emissions. Los Angeles: Air Pollution Control District of the County of Los Angeles, April 1958.

U.S. Department of Health and Human Services, Task Force on Health Risk Assessment. Determining Risks to Health: Federal Policy and Practice. Dover, Mass.: Auburn House Publishing Company, 1986.

U.S. Department of the Interior, U.S. Geological Survey, Conservation Division. Outer Continental Shelf Statistics: 1953 Through Calendar Year 1979. Washington, D.C.: June 1980.

U.S. Environmental Protection Agency. "The Carcinogen Assessment Group's Final Risk Assessment on Arsenic." Report prepared by Carcinogen Assessment Group, Washington, D.C.: May 2, 1980.

———, Office of Air Quality Planning and Standards. Fugitive Emission Sources of Organic Compounds—Additional Information on Emissions, Emission Reduction, and Costs. Research Triangle Park, N.C.: April 1982.

U.S. Food and Drug Administration, Bureau of Foods. Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food. Washington, D.C.: 1982.

U.S. General Accounting Office. Adequacy of Safety and Efficiency Data Provided to EPA by Nongovernmental Laboratories, RED-76-63. Washington, D.C.: January 26, 1976.

———. Improving the Scientific and Technical Information Available to the Environmental Protection Agency in Its Decisionmaking Process, CED-79-115. Washington, D.C.: September 21, 1979.

———. Does Nitrite Cause Cancer? Concerns About Validity of FDA-Sponsored Study Delay Answer, HRD-80-46. Washington, D.C.: January 31, 1980.

———. Regulation of Cancer-Causing Food Additives—Time for A Change? HRD-82-3. Washington, D.C.: December 11, 1981.

———. Cost-Benefit Analyses Can Be Useful in Assessing Environmental Regulations, Despite Limitations, RCED-84-62. Washington, D.C.: April 6, 1984.

———. EPA's Efforts to Identify and Control Harmful Chemicals in Use, RCED-84-100. Washington, D.C.: June 13, 1984.

Wetherold, R. G., and L. Provost. Emission Factors and Frequency of Leak Occurrence for Fittings in Refinery Process Units. Research Triangle Park, N.C.: U.S. Environmental Protection Agency, Office of Research and Development, February 1979.

Wetherold, R. G., et al. Assessment of Atmospheric Emissions from Petroleum Refining. Vol. 3. Appendix B. Research Triangle Park, N.C.: U.S. Environmental Protection Agency, Office of Research and Development, April 1980.

Wetherold, R. G., et al. Evaluation of Maintenance for Fugitive VOC Emissions Control. Prepared for the U.S. Environmental Protection Agency, Industrial Environmental Research Laboratory, Cincinnati, Ohio: Radian Corp., February 1981.

World Health Organization. Environmental Health Criteria 18: Arsenic. Geneva, Switzerland: 1981.

Yin, Robert K. "The Case Study Crisis: Some Answers." Administrative Science Quarterly, 26 (1981), 58-65.

———. Case Study Research: Design and Methods. Beverly Hills, Calif.: Sage Publications, 1984.

———, and Karen A. Heald. "Using the Case Survey Method to Analyze Policy Studies." Administrative Science Quarterly, 20 (September 1975), 371-81.

# Glossary

| | |
|---|---|
| Benefits Analysis | The analysis of the benefits of a particular regulatory action. Such analyses are often quantified in terms of monetary damage caused by uncontrolled hazard sources or the amount that people would be willing to pay to avoid certain outcomes. |
| Best Demonstrated Technology | A term EPA uses to refer to the level of emission control required by section 111 of the Clean Air Act. The act states that the level shall reflect<br><br>"the degree of emission limitation and the percentage reduction achievable through the application of the best technological system of continuous emission reduction which (taking into consideration the cost of achieving such emission reduction, and any nonair quality health and environmental impact and energy requirements) the Administrator determines has been adequately demonstrated." (42 U.S.C. 7411(1)) |
| Criteria Air Pollutants | An air pollutant that is emitted from numerous and diverse stationary or mobile sources and that adversely affects public health or welfare. For each pollutant, EPA is required, under the Clean Air Act, to publish a scientific compendium or "criteria" document showing the adverse effects of these substances at various concentrations. |
| Delaney Clause | Section 409(c)(3)(A) (sponsored in the House of Representatives by James Delaney) of the Federal Food, Drug, and Cosmetic Act. The provision states that FDA can approve no direct food additive if it is carcinogenic in animals or humans. This section of the act is often referred to as the "Delaney anticancer clause" or simply the "Delaney clause." |
| De Minimis Risk | A finite level of risk that is considered negligible. When an exposure threshold of zero risk cannot be established, it may be determined for regulatory purposes that the minimum level is effectively zero. Chances of hazard occurrence of 1 in 100,000 and 1 in 1,000,000 have been suggested as appropriately representing a de minimis risk. |
| Development and Evaluation of Risk Management Options | The process of formulating possible regulatory options and assessing their public health, economic, social, and political consequences. Risk management options are often partially determined by legislative guidance. |

| | |
|---|---|
| Dose-Response Assessment | The process of describing the relationship between the dose of an agent administered or received and the incidence of adverse effects in a population exposed to the agent. The incidence of adverse effects is characterized as a function of the degree of exposure to the agent while the influence of other possible factors such as age, sex, and smoking history are held constant. |
| Epidemiology | The study of the incidence, distribution, causes, and control of disease in a human population. Epidemiological studies are generally statistical analyses of the associations between variables rather than experimental, because of the inability to assign humans randomly to exposure groups. |
| Executive Order 12291 | A 1981 executive order that requires the review of all regulatory actions by the Office of Management and Budget. For major actions, defined in part as those expected to cost more than $100 million, a "regulatory impact analysis" containing a statement of benefits and costs of the proposed rule must be submitted. |
| Exposure Assessment | The process of characterizing human exposure to a particular substance. It includes identifying the sources, routes, and concentration of exposure and the populations at risk. Exposure assessment is often used in the development and evaluation of control options and is salient to risk characterization. Technically, exposure (the amount individuals come in contact with) is distinct from dose (the amount that enters or interacts with individuals), but they are often used synonymously. |
| Exposure Routes and Concentration | In exposure assessment, the "route" of exposure refers to the way an agent moves through the environment, enters the human body—for example, through the skin or through inhalation—and is processed in the body. It includes the types of human activity in which exposure occurs, such as through specific occupations or food consumption. Concentration refers to the degree of exposure or dosage—that is, the intensity, duration, and frequency of exposure. |
| Hazard Identification | The qualitative determination of whether exposure to a substance can cause an increase in the incidence of ill health (such as cancer, birth |

| | |
|---|---|
| | defects, and the like) and whether anyone is currently exposed or likely to be exposed as a result of an activity under consideration. |
| Hazardous Air Pollutant | A particularly dangerous air pollutant, as defined by section 112 of the Clean Air Act, that is not emitted by a sufficiently wide range of sources to justify national ambient air quality standards but that is controlled instead by national emission standards for hazardous air pollutants. |
| Interspecies Extrapolation | The adjustment of short-term or long-term bioassay dose-response results to allow for differences between the test species and humans. Such extrapolations usually assume that effects are equivalent when dosage is standardized in terms of body weight, body surface area, lifetime, and so forth. No empirical basis exists for selecting one standardizing assumption over another. Animal species may be more or less sensitive to a particular substance than humans. |
| Linear Model | A mathematical function that when used in risk characterization assumes that risk is directly proportional to dose. See also Quadratic model. |
| Long-Term Bioassay | An experimental animal study in which specimens are randomly assigned to exposure groups. The animals are routinely observed for the majority of their lifetimes. At the termination of the study, the surviving animals are destroyed, and their internal organs are examined in detail for abnormal tissues. |
| Low-Dose Extrapolation | The adjustment of observed empirical dose-response information, typically derived from circumstances in which animals or humans experienced high doses, to predict the frequency of adverse effects at low doses outside the range of observation but expected, both in most actual existing situations of human exposure and as the result of regulation. Several linear and nonlinear relationships may "fit" the empirical data, and two or more may be equally plausible in terms of current biological knowledge. |
| Monitoring and Evaluation | The process of determining the extent to which an implemented policy or regulation results in the anticipated outcome. |

| | |
|---|---|
| Mutagenicity | The capacity to produce mutations. Short-term bioassays (in which living animal tissues are exposed to suspected agents) test for mutagenicity and other genetic damage. Because many mutagens are also carcinogens, a positive response to a short-term bioassay is considered supportive but not conclusive evidence of carcinogenicity. |
| New Source Performance Standard | A term EPA uses to refer to the type of standards required by section 111 of the Clean Air Act. The act states that "standards of performance for new stationary sources" are applicable to a category of new stationary sources of emissions if, in the judgment of the administrator, "it causes, or contributes significantly to, air pollution which may reasonably be anticipated to endanger public health or welfare" (42 U.S.C. 7411(b)(1)(A)). See also Best demonstrated technology. |
| Probability Sample | A number of individuals taken from a larger population with a known probability of selection. The legitimate use of statistical inference is impossible without probability sampling. Nonprobability samples may actually be quite representative of the population from which they are drawn and probability samples quite unrepresentative; it is a matter of chance. But it is not possible to quantify uncertainty arising from sampling error without a probability sample. |
| Quadratic Model | A mathematical function that when used in risk characterization assumes that risk is proportional to the square of the dose. See also Linear model. |
| Regression Analysis | Statistical techniques for describing the relationship between two or more variables. The analysis may be used to estimate the unknown value of one variable from the known value of another. |
| Regulatory Decisionmaking | The process of selecting an appropriate regulatory response ranging from no action to banning a substance. The selection necessarily involves making value judgments about such issues as the acceptability of risk and the reasonableness of costs to control it. |
| Risk Analysis | The process of examining information concerning the level of risk posed by a hazard source, the acceptability of that risk level, and possible |

| | |
|---|---|
| Mutagenicity | The capacity to produce mutations. Short-term bioassays (in which living animal tissues are exposed to suspected agents) test for mutagenicity and other genetic damage. Because many mutagens are also carcinogens, a positive response to a short-term bioassay is considered supportive but not conclusive evidence of carcinogenicity. |
| New Source Performance Standard | A term EPA uses to refer to the type of standards required by section 111 of the Clean Air Act. The act states that "standards of performance for new stationary sources" are applicable to a category of new stationary sources of emissions if, in the judgment of the administrator, "it causes, or contributes significantly to, air pollution which may reasonably be anticipated to endanger public health or welfare" (42 U.S.C. 7411(b)(1)(A)). See also Best demonstrated technology. |
| Probability Sample | A number of individuals taken from a larger population with a known probability of selection. The legitimate use of statistical inference is impossible without probability sampling. Nonprobability samples may actually be quite representative of the population from which they are drawn and probability samples quite unrepresentative; it is a matter of chance. But it is not possible to quantify uncertainty arising from sampling error without a probability sample. |
| Quadratic Model | A mathematical function that when used in risk characterization assumes that risk is proportional to the square of the dose. See also Linear model. |
| Regression Analysis | Statistical techniques for describing the relationship between two or more variables. The analysis may be used to estimate the unknown value of one variable from the known value of another. |
| Regulatory Decisionmaking | The process of selecting an appropriate regulatory response ranging from no action to banning a substance. The selection necessarily involves making value judgments about such issues as the acceptability of risk and the reasonableness of costs to control it. |
| Risk Analysis | The process of examining information concerning the level of risk posed by a hazard source, the acceptability of that risk level, and possible |