Appendices

APPENDIX A

Federal Focus "Discussion Draft" Executive order
(May 1989)

DISCUSSION DRAFT

Executive Order _____ of _____, 1990

Risk Assessment and Risk Management

By the authority vested in me as President by the Constitution and laws of the United States of America, and in order to improve consistency among federal agencies in assessing the risks of potential environmental hazards, communicating risks to the public, and utilizing those assessments as the basis for regulatory decisions, it is hereby ordered as follows:

Sec. 1. Definitions. For purposes of this Order:

(a) "Risk assessment" refers to the process of identifying hazards and quantifying the degree of risk they pose for exposed individuals and populations. It also refers to the document containing the explanation of how the process has been applied to an individual substance or activity.

(b) "Risk assessment guideline" refers to a document specifying the assumptions and methodologies that will be employed in performing risk assessments.

(c) "Risk management" refers to the regulation of risks.

(d) "Negative data" refers to data indicating that under certain conditions a given substance or activity did not induce a toxic response.

(e) "De minimis risk" refers to risk which is clearly insignificant in relation to the ordinary day-to-day risks faced by the average individual.

Sec. 2. Fundamental Principles. In conducting risk assessments, formulating risk assessment guidelines, and making risk management decisions, Executive departments and agencies shall be guided by the following principles:

(a) Consistency should be achieved to the maximum extent possible in risk assessment procedures and risk management decisions among federal agencies and programs and environmental media.

(b) Risks should be communicated to the public in a manner that emphasizes the most scientifically realistic appraisal of expected risk for the average individual, with due regard for more sensitive individuals.

(c) Scientific justification for assumptions employed in assessing risks should be reviewed periodically.

(d) Risk management decisions should be based to the maximum extent possible on scientific data and judgment concerning actual risks rather than assumptions concerning hypothetical risks.

(e) Risk management decisions should be made with due regard for allocation of scarce societal resources available for mitigation of other significant environmental risks and with due regard for substitution risks.

Sec. 3. Specific Criteria for Risk Assessments. In addition to the above fundamental principles, Executive departments and agencies, shall adhere, to the extent permitted by law, to the following criteria when preparing risk assessments and risk assessment guidelines:

(a) Risk assessments shall employ the most scientifically realistic assumptions.

(b) Where practicable, empirical data shall be employed in risk assessments rather than assumptions or modeling.

(c) Agency risk assessment guidelines shall provide for the use of alternatives to linear dose-response models in cases where valid scientific data indicates the likelihood of a non-linear mechanism of action.

(d) Due weight shall be given to relevant negative laboratory and epidemiological data.

Sec. 4. Characterization and Communication of Risk. In communicating risk to the public through official documents such as rulemaking notices, health advisories, and health criteria documents, Executive departments and agencies shall exercise their scientific judgment to:

(a) Provide (1) an estimate of the most likely risk for the average individual as well as sensitive populations and highly exposed individuals, (2) an estimate of the most likely risk to the entire exposed population in terms of expected annual cancer incidences, as well as the most likely risk to sensitive or highly exposed populations in terms of annual cancer incidences, and (3) an estimate of the "worst case" risk for (1) and (2).

(b) If available scientific data are inadequate to permit statement of most likely risk levels, state risk in terms of a range with upper and lower bounds.

(c) Where possible, describe risk (1) by comparison with risks from other activities or exposures familiar to and routinely encountered by individuals, (2) with

explanation of major sensitivity and uncertainty factors, including an assessment of the weight of the evidence.

(d) Include an explanation and assessment of any risks that could be increased through regulation of the substance or activity under review - for example, by reason of substitution products or activities, or diminishment of resources available for other programs.

Sec. 5. Specific Criteria for Risk Management Decisions. In addition to the fundamental principles in Sec. 2, Executive departments and agencies shall adhere, to the extent permitted by law, to the following criteria in making risk management decisions:

(a) Strive for consistency among departments and agencies and across programs and environmental media. Where a department or agency regulates a substance or activity which has also been regulated by another department or agency, or has been regulated by the same department or agency under a different program or authority, the regulatory decision shall explain the extent to which it is consistent or inconsistent with such preexisting regulatory actions and provide a clear and reasonable justification for any inconsistency.

(b) Avoid the regulation of de minimis risks.

(c) In cases where the health benefits of regulatory action are estimated, base such estimates on assessment of most likely risk. If the data do not permit an assessment of most likely risk, benefits shall be based on a mid-point in the range, adjusted for the weight of the evidence.

(d) In determining margins of safety, give due regard to (1) the extent to which the selection of assumptions for the risk assessment incorporates a margin of safety into the ultimate risk assessment, and (2) the margins of safety utilized in other programs.

(e) Take into account alternative means to obtain benefits, and approaches that would allow regulated parties maximum flexibility in achieving the regulatory objectives.

APPENDIX B

Regulatory Policy Guidelines 4 and 5, Accompanying Explanatory Text, and Supplemental Discussion of Risk Assessment (Presidential Task Force on Regulatory Relief, Aug. 1983)

4. **Regulations that seek to reduce health or safety risks should be based upon scientific risk-assessment procedures, and should address risks that are real and significant rather than hypothetical or remote.**

Individual decisions involving risk are a routine of everyday life: in driving, choosing an occupation, doing home repairs, engaging in sports, and so on. Regulatory agencies are frequently called upon to make decisions involving risk on behalf of large segments of the public: in setting product, workplace, or environmental standards, and in determining whether certain products should be marketed at all. Agencies have sometimes made these decisions much too conservatively, forcing expenditures on risk reduction that greatly exceed the ordinary prudence of private decisions involving risk. In many circumstances, it is indeed appropriate that government should err on the side of caution in protecting public health and safety. But a sense of balance is required, and a sense of the limits of regulatory policy in affecting the overall level of private and public risk in society. All decisions involving risk--public and private--have costs as well as benefits, and excessive costs in any one area can be counterproductive on the whole, reducing resources or incentives for increased health and safety in other areas.

In recent years, scientists and policy analysts have developed increasingly sophisticated procedures for assessing health and safety risks and for incorporating these assessments into benefit-cost analyses. When evaluating product, workplace,

and environmental standards, the calculation of potential benefits should derive from the best estimate of the expected reduction in risk attributable to the standard, using the best estimate of the value of that risk reduction to the affected group. When deciding whether to issue a license or permit for a new project or product, the calculation of the decision's potential costs should include the potential costs of any increase in risk--again using the best estimate of expected risk and the best estimate of the value of avoiding it. In the first case, the benefits of risk reduction are compared to the costs of reduction; in the second, the costs of risk increase are compared to the benefits of the new product or production process (these benefits will often include reduced risks elsewhere).

To be useful in determining overall benefits and costs, risk assessments must be scientifically objective and include all relevant information. In particular, risk assessments must be unbiased best estimates, not hypothetical "worst cases" or "best cases." Extreme "best" or "worst" safety or health results should be weighted (along with intermediate results) by the probability of their occurrence to estimate the expected result implied by the available evidence. In addition, the distribution of probabilities for various possible results should be presented separately, so as to allow for an explicit "margin of safety" in final decisions.

Risk assessment is not simply a matter of estimating the health effects of a particular substance at assumed levels of exposure, or the safety effects of a particular product at assumed levels of use. The risk of a substance in the real world is the product of two factors: hazard (defined as health effects at given levels of exposure), and actual exposure. At the extremes, there is no risk from even an extremely hazardous substance if no one is exposed to it, just as there is no risk from widespread exposure to a substance that is not hazardous in itself. Regulatory decisions cannot, therefore, be based solely on abstract inferences from laboratory or epidemiological data. Real-world exposure among those who would be affected by the regulatory decision must be considered as well.

5. **Health, safety, and environmental regulations should address ends rather than means.**

Regulations to limit environmental pollution, or to protect the health or safety of workers or consumers, frequently dictate the exact engineering methods for achieving their intended results. Under the Clean Air Act, for example, some "new source performance standards" for controlling air pollution adopt extremely narrow definitions of what constitutes a "new source" of pollution, and effectively prescribe how each "new source" must be designed and operated. As a result, individual manufacturing plants often contain many separate "sources" of a single air pollutant. Modernizing a single production facility within a plant may create a new pollution "source"--even when the net effect is to reduce pollution by replacing an older and less efficient facility. And every such "source" in the nation may be required to meet the same federal design and operating specifications, regardless of whether other approaches would reduce pollution at less expense.

Regulations that impose precise engineering requirements are generally cost-ineffective, especially when applied on a uniform nationwide basis. The best means of accomplishing a given environmental end varies from firm to firm, region to region, and over time; imposing engineering uniformity on these natural variations almost always results in too much economic cost for a given environmental benefit, or too little environmental benefit for a given economic cost. Moreover, government-prescribed

uniform technology retards productivity growth--dampening market competition and reducing incentives for innovation in both production and pollution-control markets.

For these reasons, regulatory standards should be adopted in terms of results or performance rather than specifying the means employed to achieve the results. With performance-oriented standards, regulated firms are responsible for meeting some regulatory target, but are free to choose--or invent--the easiest or cheapest methods to reach the target. The Food and Drug Administration's "tamper-resistant packaging" regulation is a good example of this approach.

In practice, the distinction between performance standards and design standards is a continuum rather than a simple dichotomy. Regulatory policymaking usually involves selecting a point on a spectrum running from pure design standards to pure performance standards. The variety of approaches to protecting workers from airborne health hazards is a typical example. There is a spectrum of choices theoretically available in defining a standard, each one closer to the ultimate objective of protecting workers' health:

- An engineering standard would prescribe the design and operation of each source of the airborne chemical or other substance in question.

- An emission standard would prescribe the maximum rate of emission from each source, but without prescribing the engineering techniques for achieving these rates.

- A concentration standard would prescribe maximum ambient concentrations of the chemical or substance, while allowing any combination of emission rates from different sources that would achieve the concentration level.

- An exposure standard would prescribe maximum levels of exposure for employees themselves, while allowing any combination of ambient concentration controls, personal protective devices (such as respirators), and work rotation practices that would achieve the exposure level.

- A health standard would prescribe maximum levels of intermediate health effects (blood levels of a given indicator) or ultimate health effects (disease rates), while allowing these levels to be achieved through combinations of exposure limits and direct medical surveillance.

While moving towards the performance end of the spectrum produces more cost-effective regulations, this tendency is offset by difficulties of enforcing performance standards in certain

circumstances. A pure health standard would be impractical, for example, where the disease being protected against manifests itself only after decades of exposure. Uncertain causation between a given design and a given performance objective is not, however, an argument against performance standards, since this uncertainty also applies to any design requirement prescribed by the government. To the degree that performance can be measured or reasonably imputed, a standard based on this level of performance is always superior to more means-oriented regulation.

Performance standards should also be applied as broadly as possible without creating too much variation in regulatory benefits. An example is the Environmental Protection Agency's "emissions bubble" policy, which effectively regulates existing air pollution sources on a plant-wide (sometimes even firm-wide) basis rather than source-by-source. Under this policy, plant managers can exceed emissions standards at any one source where control costs are relatively high, so long as they achieve equivalent reductions at other sources where control costs are lower. EPA is actively considering ways to apply this approach to new sources as well.

Automobile regulations offer similar opportunities for regulating broadly rather than narrowly. The federal government's automobile fuel economy regulations, while an unnecessary form of regulation in themselves, at least have the virtue of applying to a manufacturer's entire fleet rather than to individual vehicle

odels, permitting manufacturers to vary fuel economy among odels within a single constraint on average economy. Permitting eetwide avaraging of automobile emissions, not currently ermitted, would yield substantial cost savings without ompromising air quality. EPA has recently proposed a rule to ermit averaging under its motor vehicle diesel particulates tandard, and is preparing rules to permit averaging in other reas as well.

Performance standards are also an important counterweight to he "soak the rich" regulatory tendencies that appear to be articularly damaging to productivity growth. For example, under he Clean Air Act and Clean Water Act, EPA is to set echnology-based pollution-control standards; in various sections f the Acts, the technology is to be the "best technology" that s "economically feasible," (and other similar formulations). hile the importance of control costs under these standards is ften less than clear, one interpretation with some support in he Congress and strong traditional support at EPA is that ontrols are to require the best technologies affordable by the ndustries to which they apply. Until recently, EPA analyses of he economics of its rules placed heavy emphasis on industries' inancial circumstances and other "affordability" factors, and ittle emphasis on cost-effectiveness. This is a sure formula or loading disproportunate costs on the most productive sectors f the economy, thus squandering both environmental protection

and potential economic growth; and the result has indeed been wide variations in the cost-effectiveness of pollution controls from industry to industry.

Under E.O. 12291, the Administration has placed major emphasis on making technology-based pollution controls more effective. EPA now calculates incremental removal costs as part of all air and water pollution control rulemaking proceedings. To facilitate cross-industry comparisons in its water program, the Agency has established a formal weighting scheme for the toxicity of various pollutants, using the toxicity of copper as a "numeraire"; the "copper-equivalent cost-effectiveness" of every new water effluent guideline is now gauged against those of existing guidelines. Similar cost-effectiveness benchmarks are being introduced into the air pollution standards. While the Clean Air and Water Acts remain strongly biased towards the adoption of uniform engineering standards, cost-effectiveness approaches such as these are helping to select engineering controls with as much attention to ultimate performance as the statutes permit. This will reduce the needless productivity costs of pollution controls and also produce greater benefits for the nation's environmental investments.

The Justice Department has also returned antitrust enforcement to its proper and vital function of policing against serious restraints of trade. In recent decades--primarily through government consent agreements and private treble-damage suits--antitrust had come to prohibit numerous trade practices that promote rather than harm consumer welfare and efficient resource use. In the past two years, the Department has begun a comprehensive reexamination of over 1,200 existing judicial decrees; several have already been the subject of successful court actions revising or eliminating the decrees, and about 200 more are under active consideration. It also has begun an active program of intervention in private antitrust actions, pressing for new antitrust interpretations to allow economically beneficial marketing and distribution arrangements, joint ventures among competing firms, and protection of property rights in new innovations. These modifications of existing antitrust doctrine will be especially important to the nation's high-technology industries.

These reforms have permitted the Department to focus renewed attention on serious trade restraints, especially price-fixing and bid-rigging: in fiscal year 1982 the Antitrust Division filed more cases, and more criminal cases, than in any other year since passage of the Sherman Act of 1890.

12. Risk Assessment. The advance of science has brought with it the discovery of new substances that may be hazardous to

human health. In addition to discovering potentially-risky new substances, new techniques and knowledge have led to discoveries that substances previously considered safe may pose low level risks, and have enabled scientists to detect the presence of substances in products and the environment at far lower levels than was previously possible. These advances have made some traditional regulatory policies obsolete--especially those that aim simply at identifying potentially hazardous substances, and then requiring that they be controlled to the extent technologically feasible or banned outright. As discussed in Part I of this Report, the Administration needs to ensure that scientifically consistent policies for assessing the significance of low-level risks are established and that these assessments are incorporated in regulatory decisions.

Over the past two years, the Administration has undertaken several major efforts to refine and update policies to reduce the risks posed by hazardous substances in the environment, in the workplace, and in consumer products. The President's Office of Science and Technology Policy (OSTP) is chairing a regulatory work group composed of scientists from eight federal regulatory and research agencies; this group is developing a consensus regarding the scientific principles needed for assessment of carcinogenic risks, so that all federal agencies will have a common scientific foundation upon which to base regulatory policies concerning potential carcinogens and other hazardous substances.

In addition to the OSTP effort, the Occupational Safety and Health Administration (OSHA), the Environmental Protection Agency (EPA), and the Food and Drug Administration (FDA) are separately reviewing their existing policies on carcinogenic risk assessment. OSHA is reviewing its "carcinogen policy" in light of important judicial and scientific developments since the policy was issued in 1980. EPA withdrew its 1979 proposed policy for airborne carcinogens under Section 112 of the Clean Air Act, and has formed a special task group that is now soliciting comments on a new draft policy. And FDA has solicited public comments on its "constituents" approach to protecting the public health against carcinogens in the food supply; this approach uses risk assessment to determine whether an additive as a whole is safe, even though one of its "constituents" may technically be a carcinogen.

The Administration will complete these initiatives during the next year and issue a consistent set of policies that, to the extent allowed by the individual regulatory statutes, will permit all federal agencies to employ risk assessment procedures based on modern scientific thinking and the general policies of Executive Order 12291.

13. Agricultural Marketing Orders. Fruit and vegetable marketing orders were identified early in 1981 by the Task Force for reassessment and possible modification. After intensive study, the Department published a review of marketing orders in

APPENDIX C

Office of Science and Technology Policy Cancer Principles (U.S. Interagency Staff Group on Carcinogens, Feb. 1985)

Principles

Preface

The principles contained herein were derived from the information detailed by the members of the Interagency Staff Group. This section attempts to provide, in a nontechnical form, some important general statements relevant to the evaluation of the role of chemicals in carcinogenesis. These statements are intended to serve as a bridge connecting the basic science with multifaceted process of risk assessment.

Since there are gaps in the information available, differences in evaluations and in scientific opinion may exist about certain of the points highlighted as principles. However, these principles derive from a Weltanschauung utilizing a balanced approach with an appreciation of all elements of the problem, from hazard identification and estimation through exposure and risk assessment. It is clearly understood that new information and newly emerging concepts may modify some of these statements. Indeed, an unstated "zeroth" principle is that regulatory judgments should embody an openness to advances in science and emerging scientific understanding. As a consequence, it is necessary that the process which led to this document be a continuing one, with periodic updates as new advances in science dictate. For the time being however, as a result of an arduous cooperative effort, these statements, we believe, represent an up-to-date summary on a number of important topics.

Principles Derived from the Mechanisms of Carcinogenesis

Principle 1. Carcinogenesis is a multistage phenomenon that may involve the genome both directly and indirectly. These stages of carcinogenesis may be, to varying degrees, influenced by a number of variables such as age at exposure, diet, hormonal status, and intra- and interspecies variability, which should be considered when trying to predict human response to potentially carcinogenic agents (see pp. 209–212, 220–224).

Principle 2. Appropriate *in vitro* and *in vivo* tests can indicate that an agent has a certain action such as genetic toxicity or promotion. Such information is valuable and may be useful in evaluating mechanism(s) of cancer induction. However, in the evaluation of human risk, the attribution of observed findings of carcinogenicity to a particular biological effect must rest upon sound evidence that the effect is responsible for the cancer induction. It must be kept in mind that a chemical may contribute to carcinogenesis in multiple ways (see pp. 220–224).

Principle 3. At the present stage of knowledge, mechanistic considerations such as DNA repair and other biological responses, in general, do not prove the existence of, the absence of, or the location of a threshold for carcinogenesis. The presence or absence of a threshold for one step of the carcinogenic process does not necessarily determine the presence or absence of a threshold for the whole process (see pp. 217–219).

Principle 4. The carcinogenic effects of agents may be influenced by nonphysiological responses (such as extensive organ damage, radical disruption of hormonal function, saturation of metabolic pathways, formation of stones in the urinary tract, saturation of DNA repair with a functional loss of the system) induced in the model systems. Testing regimes inducing these responses should be evaluated for their relevance to the human response to an agent, and evidence from such a study, whether positive or negative, must be carefully reviewed (see pp. 217–220).

Principles from Tests of Cancer Induction

Principle 5. Short-term tests, such as assays for point mutations, chromosomal aberrations, DNA damage, and *in vitro* transformation are useful in screening for potential carcinogens, reaching a judgment on the carcinogenicity of a chemical, and providing information on carcinogenic mechanisms (see pp. 227–230).

Principle 6. Short-term tests are presently limited in their ability to predict the presence or absence of carcinogenicity and cannot supplant data from long-term animal studies or epidemiological investigations, since the tests do not necessarily screen for all potential means of cancer induction and do not necessarily mimic all reactions that would occur *in vivo*. Additional research is required to improve existing tests and develop ones that identify chemicals which act by genetic mechanisms not yet determined or which act by other, nongenetic mechanisms (see pp. 227, 232–233).

Principle 7. Short-term tests should be carefully selected to ensure they have been adequately validated. Several tests with different endpoints may be required to characterize a chemical's response (see pp. 230–232).

Principle 8. In the evaluation of long-term test results, the term "carcinogen" should be used in a broad sense, i.e., a substance which is capable under appropriate test conditions (Principles 10–13) of increasing the incidence of neoplasms (combining benign and malignant when scientifically defensible) or decreasing the time it takes for them to develop. Careful consideration to the relevant issues cited in Principles 4, 9, and 14 should be given prior to a determination that a chemical is an animal carcinogen (see pp. 234–235).

Deference should be given to the IARC principle: "that in the absence of adequate data in humans, it is reasonable, for practical purposes, to regard chemicals for which there is sufficient evidence of carcinogenicity in animals as if they presented a carcinogenic risk to humans." However, this presumption is evaluated along with other relevant information (Principle 25) in making a final judgment concerning human carcinogenicity and should not foreclose further inquiry into the human relevance of animal carcinogens (see pp. 253–254).

Principle 9. Some experimental animal models ordinarily have high incidences of certain tumors. The evaluation of tumor data from such animals can pose special problems. For example, the interpretation of cancer incidence in some strains of rats with testicular or mammary tumors or in some strains of mice with lung or liver tumors must be approached carefully in the light of other biological evidence bearing on potential carcinogenicity (see pp. 240–241).

Principle 10. Protocols for long-term tests should be designed to achieve an appropriate balance between the two essential characteristics of a biological assay: adequate biological and statistical sensitivity (a low false negative rate) and adequate biological and statistical specificity (a low false positive rate). The absence of biases in selection and allocation of animals between control and treatment groups as regards diet, husbandry, necropsy, pathology, and from insufficient quality control, is crucial (see pp. 234–238, 241–243).

Principle 11. It is appropriate to use test doses that generally exceed human exposure levels in order to overcome the inherent insensitivity of the traditional design of the long-term animal test. The highest dose should be selected after an adequate prechronic study and after evaluating other relevant information, as necessary, to determine the highest dose consistent with predicted minimal target organ toxicity and normal lifespan, except as a consequence of the possible induction of cancer (see pp. 236–239).

Principle 12. The diagnosis of pathologic lesions is complicated and requires judgment and appropriate experience. Diagnoses can differ, depending on the tissues and species involved and can change with time as techniques improve and data on bioassays accumulate. Accurate interpretation of tumor data is contingent upon careful attention to gross observation, tissue sampling, slide preparation and histologic examination. Diagnosis of tumors should be guided by evidence of their histogenic origin and stage of progression (see pp. 240–241).

Principle 13. Appropriate statistical analysis should be performed on data from long-term studies to help determine whether the effects are treatment related or possibly due to chance. These should include a statistical test for trend and a test based on pairwise comparisons, including appropriate correction for differences in survival. The weight to be given to the level of statistical significance (the p-value) and to other available pieces of information is a matter of overall scientific judgment (see pp. 241–243).

Principle 14. Decisions on carcinogenicity of chemicals in animals should be based on consideration of relevant biological and biochemical data. Use of background or recent historical control incidence of tissue specific tumors can be an aid, in addition to concurrent controls, in the evaluation of tumor data. Care should be exercised when combining different control groups.

Evidence of probable reproducibility is important. This evidence can consist of independent confirmation of the original findings or may be derived from intergroup comparisons of tumor incidence data, between dose groups, sexes, strains or species. Evidence of dose response increases confidence that the effect is treatment-related; similarly, lack of an observed dose response may reduce the likelihood that the effect is associated with the treatment.

Confidence is increased when: (1) the incidence of tumors is markedly elevated in the treated groups compared to controls; (2) tumor incidence is significantly increased at multiple anatomical sites; and (3) tumor latency is significantly reduced. In addition to tumor incidence at specific sites, the stage in the development of neoplasia should be evaluated. For example, the finding that the majority of neoplastic lesions at a specific site is more advanced in a treated group compared to its control may provide additional evidence of a treatment related effect. Conversely, the finding that the control group lesions are more advanced might argue that a marginal elevation of tumor incidence is not treatment-related. The incidence of preneoplastic lesions in treatment or control groups may, in certain instances, provide evidence for the biological plausibility of a neoplastic response and contribute to the interpretation of a bioassay. Identification of effects from prechronic studies on the target organ(s) can aid in the evaluation of long-term studies. Information on the activity of chemicals at the physiological, cellular and molecular level may be important to the evaluation of carcinogenicity data on a case-by-case basis (see pp. 238–243).

Principles for Epidemiology

Principle 15. The major strength of the epidemiological method is that it is the only means of assessing directly the carcinogenic risk of environmental agents in humans. However, the observational (nonexperimental) nature of most epidemiological studies, as well as the frequent paucity of relevant data, can impose serious limitations on the method (see pp. 243–244).

Principle 16. Descriptive epidemiological studies (based on the measurement of disease rates for various populations), including correlational studies (in which the rate of disease in a population is compared with the spatial or temporal distribution of suspected risk factors), are useful to generate and refine hypotheses, or provide supporting evidence in evaluating relationships detected by other means, but rarely, if ever, provide information allowing a causal inference (see pp. 244–245).

Principle 17. Well designed, conducted, and evaluated analytic epidemiological investigations of either the case-control or cohort variety can provide the basis for causal inferences especially useful for public health decisions (see pp. 245–246).

Principle 18. Elements in interpreting the likely causality of epidemiological observations include the magnitude of the risk estimates (strength of associations); the possibility of their being due to chance (statistical significance); the rigor of the study design to

avoid various kinds of bias, including those related to selection, confounding, classification and measurement; dose-response relationships; the temporal relationships between exposure and disease; the specificity of the associations; their biological plausibility; and the reproducibility of the findings (see p. 244).

Principle 19. A high-quality negative epidemiological study, while useful, cannot prove the absence of an association between chemical exposure and human cancer. Within the scope of the study, specifically for the populations studied (including concomitant exposures), for the levels and durations of exposure to the agents evaluated and for the time assessed following exposure, a likely range can be determined for the estimates of risk and the statistical likelihood of the study to detect an effect can be assessed (see p. 247).

Principles for Exposure Assessment

Principle 20. It is desirable that exposure routes employed in animal health effects studies are comparable to human exposure routes both for the simplification of risk assessment and because there may be important route-dependent differences in molecular, biochemical, and physical parameters in organs (see pp. 209–212).

Principle 21. At present, a single generally applicable procedure for a complete exposure assessment does not exist. Therefore, in the near term, it is expected that integrated exposure assessments (utilizing monitoring data, results from physical and chemical models, and considerations of all routes of exposure through all media) will be conducted on a case-by-case basis (see pp. 248–251).

Principle 22. The depth and accuracy of an exposure assessment should be tailored to provide the degree of knowledge required to support analytical needs. A preliminary assessment using available crude data can often shed light on the upper or lower bounds of potential risks (see pp. 248–251).

Principle 23. An exposure assessment should describe the strengths, limitations and uncertainties of the available data and models and should indicate the assumptions made to derive the exposure estimates (see pp. 251–252).

Principle 24. In general, an array or range of exposure values is preferable to a single numerical estimate (see pp. 251–252).

Principles for Risk Assessment

Principle 25. Decisions on the carcinogenicity of chemicals in humans should be based on considerations of relevant data, whether they are indicative of a positive or negative response, and should use sound biological and statistical principles. This weight of evidence approach should include consideration of all relevant factors and should give appropriate weight to each on a case-by-case basis. Examples of the types of information that should be taken into account include: findings from long-term animal studies (see Principle 14); results from epidemiological studies (see Principles 16–19); results from *in vivo* and *in vitro* short-term tests (see Principles 5 and 6); and data from studies of mechanism, including factors such as structure-activity relationships, and known similarities and differences in metabolic and kinetic profiles for different species (see Principles 1–4 and pp. 253–254).

Principle 26. No single mathematical procedure is recognized as the most appropriate for low-dose extrapolation in carcinogenesis. When relevant biological evidence on mechanism of action (e.g., pharmacokinetics, target organ dose) exists, the models or procedures employed should be consistent with the evidence. However, when data and information are limited, and when much uncertainty exists regarding the mechanisms of carcinogenic action, models or procedures which incorporate low-dose linearity are preferred when compatible with the limited information (see pp. 255–258).

Principle 27. The quantification of the various sources of uncertainty involved in cancer risk assessment can be as important as the projection of the risk estimate itself. The sources that might be addressed include the statistical uncertainty associated with the given risk estimate (often expressed as upper and lower confidence bounds); the variability introduced by the selection of a particular low-dose extrapolation procedure (often expressed as an envelope of risk estimates from a variety of plausible models); when risk estimation is based on laboratory-generated data, the biological variability associated with the use of a particular test organism and its scaling or extrapolation to man (see pp. 257–258).

Principle 28. An estimate of cancer risk for humans exposed to an agent can be no more accurate than an exposure assessment that it utilizes. Lack of adequate exposure data is frequently a major limiting factor in evaluation of carcinogenic risks for humans (see Principles 20–24 and p. 254).

Principle 29. While several considerations often enter the risk assessment process, it is important to try to maintain a clear distinction among facts (statements supported by data), consensus (statements generally held in the scientific community), assumptions (statements made to fill data gaps), and science policy decisions (statements made to resolve points of current controversy) (see pp. 257–258).

Principle 30. Differences in human susceptibility, and variable and extreme exposures to chemicals suggest the likelihood that there are subpopulations that are at greater than average risk. Consideration should be given to the identification of high risk populations (see pp. 209–212, 251).

Principle 31. Because of the uncertainties associated with risk assessment, a full evaluation of risk to humans should include a qualitative consideration of the basic strengths and weaknesses of the available hazard and exposure data, in addition to any numerical estimations that are made (see pp. 257–258).

APPENDIX D

Discussion in the 1990–91 *Regulatory Program*

Two Executive orders establish principles of sound regulatory management for agencies to follow in developing regulations. The orders combine with the Paperwork Reduction Act of 1980 to ensure that the paperwork and regulatory burdens the Government imposes are necessary, tolerable, and cost-effective.

Both Executive orders are the outgrowth of successive Presidents' efforts since the 1960s to establish procedures for Executive regulatory oversight. Executive Order 12291 (see appendix I for the complete text) sets out fundamental regulatory principles: it directs agencies to justify the need for regulations, weigh their costs and benefits, and choose the most cost-effective regulatory options. It also directs OIRA to review the agencies' rationales and assumptions and to ensure that agency regulations are consistent with Presidential policies and statutory intent. Executive Order 12498 established the *Regulatory Program of the United States Government* as a vehicle for agency regulatory planning and coordination (see appendix II for the complete text).

Taken together, the two Executive Orders form a coherent framework for creating a regulatory structure that is humane, effective, and economically sensible. Congress and the President, through the enactment and approval of agency enabling legislation, have charged agencies with the ultimate responsibility for acting within their specific policy areas. Presidential regulatory oversight promotes a careful weighing of such actions, and also seeks to harmonize conflicts between competing agency mandates. It thus fulfills the President's Constitutional obligation to manage the executive branch. The establishment of a formal oversight process during the past decade provides a method for rationalizing the Federal Government's maze of regulatory requirements, increasing benefits and lowering costs. In the 1990s, this process will guide the continuing evolution of regulations that are both effective and prudent.

The Regulatory Program of the United States and Executive Order 12498

Executive Order 12498 requires the annual publication of the *Regulatory Program of the United States Government*. This document outlines the regulatory priorities and important upcoming actions of all the regulatory agencies. Most important is what the document represents—a process for planning and coordinating agency actions in advance of rulemaking.

The 1990–1991 Regulatory Program sets forth each agency's objectives for the 1990 program year, spanning April 1, 1990, to March 31, 1991. It outlines the issues agencies see as requiring immediate attention, as well as the steps each agency is taking to ensure the cost-effectiveness of the regulatory approach it proposes. The Regulatory Program thus allows Congress and the American people to understand the policy directions of the regulatory agencies. The actions listed in the Regulatory Program represent the major initiatives of the regulatory agencies, and may be substantially revised over time, through Administration decisions to guide and coordinate these agency actions.

Future Policies and Proposed Conferences

This overview discusses areas for further action on regulatory issues, including possible next steps in reforming economic regulation, efforts to develop a system of regulatory budgeting, key issues in developing scientific risk assessments as a part of the development of sound regulatory policy, and the use of information as a regulatory tool. The review also provides a discussion of OMB's final regulatory impact analysis (RIA) guidelines, and OIRA's response to the comments received on the draft guidelines published for comment in the 1988 Regulatory Program. Earlier Regulatory Programs identified some major inconsistencies and shortcomings in Federal agency analysis of major regulatory actions.[1] Because of these inconsistencies and shortcomings, and the potentially large net benefits that would result from improved analysis, OMB recognized the need for specific guidelines for preparing RIAs. The revised final RIA guidelines are published in this Regulatory Program in Appendix V.

OIRA plans to seek advice on these regulatory issues from the affected public and from academic and other experts on regulations and regulatory policies

[1] Regulatory Program (1986–1987), pp. xix–xxvi; Regulatory Program 1987–1988, pp. xv–xxii; and Regulatory Program 1988–1989, pp. 31–37.

and practices. OMB is considering a series of workshops and conferences, open to the public, on the following:

- The impact of the regulatory reform initiatives of the 1980s;
- Guidance and coordination for ranking risks (risk assessment, management, and communication);
- The potential for development of the regulatory budget concept; and
- The use of information strategies and public disclosure requirements to complement and possibly replace more direct regulatory intervention.

Introductory discussions of these topics are presented below in the sections entitled "Reforming Economic Regulation," "Regulatory Review and the Case for a Regulatory Budget," "Current Regulatory Issues in Risk Assessment," and "Information as a Regulatory Strategy." OIRA seeks comments on all these topics. Please send comments to the Office of Information and Regulatory Affairs, Washington, DC 20503.

The Council on Competitiveness

Maintaining American economic competitiveness into the next century depends on the initiative and innovativeness of the private sector. The Federal Government can foster competitiveness by encouraging a vigorous and competitive market environment both in this country and in the world economy. One of the more important steps the Federal Government can take in promoting a competitive market environment is to avoid unnecessary regulation.

Government regulation has an important role in advancing societal goals—such as public health and safety—where the market fails to protect such goals. But regulation can also impose substantial costs on American business, State and local governments, and consumers that burden our competitiveness abroad and our welfare at home. It is thus important to assess, on a continuing basis, the need for both new and existing regulations, balance the immediate objectives of such regulation with the broader objectives of promoting the Nation's welfare, and promote a reliance on markets wherever such opportunities exist.

To assist regulatory oversight, President Bush announced in *Building a Better America*, on February 9, 1989, that Vice President Quayle would chair the Council on Competitiveness:

> The Council will review regulatory issues, and such other issues as may be referred by the President, bearing on competitiveness. In reviewing regulatory matters, the Council will be continuing the work of the former President's Task Force on Regulatory Relief—chaired in the Reagan Administration by then Vice President Bush.

The Council will work closely with OIRA to augment the regulatory review process, ensure that the benefits of regulation outweigh their costs, and coordinate development of legislative and administrative initiatives to reduce unnecessary regulatory burdens.

Reforming Economic Regulation

The United States has traditionally relied on a "cost-of-service" approach to setting prices or rates for public utilities or other industries that are considered to be "natural monopolies." Cost-of-service regulation involves cumbersome and highly judgmental determinations of (1) the value of a firm's rate base, that is, its assets dedicated to providing a good or service; (2) the fair rate of profit (return) on those assets; and (3) all other costs associated with providing the good or service. In addition, the process of determining the "price" involves substantial costs to both the regulatory commission and the regulated industry because of complex and time-consuming ratemaking procedures.

Most students of public utility regulation believe that regulated industries are not as efficient or as innovative as they could be.[2] When regulated rates are tied directly to an individual firm's costs, those firms have a reduced incentive to achieve long-term cost reductions. Under this regulatory system, all cost savings are eventually "passed through" to consumers; the firm benefits only during the period of "regulatory lag," (that is, until rates are adjusted to reflect the firm's reduced costs). Firms subject to cost-of-service

[2] See Alfred E. Kahn. *The Economics of Regulation: Principles and Institutions*, New York: John Wiley & Sons, 1970, Chapter 2, pp. 47-94.

Supporters of this system argue that it would provide an incentive for better estimates of the costs of legislative proposals and a basis for an explicit discussion of the costs and tradeoffs of such proposals. High cost ceilings would focus attention on the expected benefits of the program, and alternative approaches; cost ceilings that were too low would prevent agencies from issuing implementing regulations. Such an approach would, needless to say, give agencies an incentive to choose regulatory approaches that would produce the greatest benefits at the lowest costs.

ISSUES AND AREAS FOR FURTHER STUDY

While the fiscal budget process provides a continuous record of actual expenditures, there is no comparable record of the cost of meeting regulatory requirements.[20] Members of Congress and the past two Administrations have considered developing an accounting framework to record direct regulatory expenditures, but more work needs to be done to solve the practical accounting problems inherent in measuring the private expenditures that Federal regulations mandate. These include:

- Developing a record of actual expenditures while minimizing the recordkeeping burden on the private sector;
- Identifying an appropriate "baseline," recognizing that some costs would be incurred even in the absence of Federal regulation; and
- Estimating the costs of forgoing certain products where Federal regulation prohibits production or distribution.

Each of these raises difficult issues in designing an effective regulatory budget process. For example, the costs of banning a product are not directly measurable and can only be estimated using complex statistical models. However, measuring only the direct compliance costs for oversight purposes creates a bias toward banning substances and products instead of controlling them.

As a first step in determining the feasibility of the regulatory budget concept, OMB has begun systematically to collect the costs of all significant published regulatory actions. Analysis of these data should aid in the development of ways to overcome the problems of regulatory budgeting, uncover unforeseen problems in developing cost estimates, and more fully refine a workable regulatory budgeting process.

Current Regulatory Issues in Risk Assessment and Risk Management

Many Federal agency regulatory decisions are intended to reduce risks to human life and health. Government regulations control which agricultural chemicals may be used to reduce insect damage, increase farm yields, and improve the quality of food products. Other rules govern hazards in the Nation's workplaces and emissions from its factories. There are regulations directing the way in which automobiles must be manufactured, commercial aircraft maintained, and trains operated. Hardly any widespread human activity that entails risk is free of some degree of social control, often achieved through government regulation.

Regulatory decisions involving risk require agencies to address questions such as, "How safe is 'safe'?" and "How clean is 'clean'?" When government agencies promulgate regulations intended to reduce a risk or mitigate a hazard, they are engaging in what has become known as *risk management*. These policy choices inevitably involve consideration of both the risks entailed by the underlying activity and the social consequences of regulatory intervention. Thus, the first challenge of risk management is to set priorities to determine which risks are worth reducing and which are not.

For government to carry out its risk-management responsibilities, there must be an extensive investment in the careful assessment and quantification of risks. The term *risk assessment* means the application of credible scientific principles and statistical methods to develop estimates of the likely effects of natural phenomena and human activities.

The need to keep risk assessment and risk management separate has long been the objective of responsible public officials. In 1983, the National Academy of Sciences (NAS) studied the process of managing risk

[20] Researchers, using different methods, assumptions, and time periods, have formed incomplete estimates by adding up the cost of individual regulations. These estimates accordingly show considerable variation for current annual costs ranging from $60 billion to $175 billion a year—5 to 15 percent of current Federal outlays.

in the Federal Government and offered the following recommendations, among others:

> *Recommendation 1:* Regulatory agencies should take steps to establish and maintain a clear conceptual distinction between assessment of risks and the consideration of risk management alternatives; that is, the scientific findings and policy judgments embodied in risk assessments should be explicitly distinguished from the political, economic, and technical considerations that influence the design and choice of regulatory strategies.[21]

> *Recommendation 2:* Before an agency decides whether a substance should or should not be regulated as a health hazard, a detailed and comprehensive written risk assessment should be prepared and made publicly available. This written assessment should clearly distinguish between the scientific basis and the policy basis for the agency's conclusions.[22]

The belief that risk assessment and risk management should be kept separate enjoys widespread support among professional risk-assessment practitioners and risk-management officials. [23]Others have emphasized the importance of ensuring that policy biases do not distort the analysis of alternative risk-management choices.[24] The NAS principles have also have been endorsed by a number of Federal agencies, including the Office of Science and Technology Policy (OSTP), the Environmental Protection Agency (EPA), and the Department of Health and Human Services (HHS).[25]

Unfortunately, risk-assessment practices continue to rely on conservative models and assumptions that effectively intermingle important policy judgments within the scientific assessment of risk. Policymakers must make decisions based on risk assessments in which scientific findings cannot be readily differentiated from embedded policy judgments. This policy environment makes it difficult to discern serious hazards from trivial ones, and distorts the ordering of the Government's regulatory priorities. In some cases, the distortion of priorities may actually increase health and safety risks.

This section explores some of the continuing difficulties that plague the practice of risk assessment, and describes briefly their policy implications. It can be summarized in three observations:

The continued reliance on conservative (worst-case) assumptions distorts risk assessment, yielding estimates that may overstate likely risks by several orders of magnitude. Many risk assessments are based on animal bioassays utilizing sensitive rodent species dosed at extremely high levels. Conservative statistical models are used to predict low-dose human health risks, based on the assumption that human biological response mimics that observed in laboratory animals. Worst-case assumptions concerning actual human exposure are commonly used instead of empirical data, further exaggerating predicted risk levels.

Conservative biases embedded in risk assessment impart a substantial "margin of safety". The choice of an appropriate margin of safety should remain the province of responsible risk-management officials, and should not be preempted through biased risk assessments. Estimates of risk often fail to acknowledge the presence of considerable uncertainty, nor do they present the extent to which conservative assumptions overstate likely risks. Analyses of risk-management alternatives routinely ignore these uncertainties and treat the resulting upper-bound estimates as reliable guides to the likely consequences of regulatory action. Decisionmakers and the general public often incorrectly infer a level of scientific precision and accuracy in the risk-assessment process that does not exist.

Conservatism in risk assessment distorts the regulatory priorities of the Federal Government, directing societal resources to reduce what are often trivial carcinogenic risks while failing to address more substantial threats to life and health. Distortions are probably most severe in the area of cancer-risk assessment, because many conservative models and assumptions were developed specifically for estimat-

[21] National Academy of Sciences, *Risk Assessment in the Federal Government: Managing the Process*, Washington, DC: National Academy Press, 1983 (hereinafter, *NAS Risk Management Study*), p. 151.

[22] Ibid., p. 153.

[23] For representative views of risk-assessment practitioners see, e.g., Lester B. Lave, *The Strategy of Social Regulation: Decision Frameworks for Policy*, Washington, DC: Brookings, 1981; Lester B. Lave, "Methods of Risk Assessment," Chapter 2 in *Quantitative Risk Assessment in Regulation*, Lester B. Lave, ed., Washington, DC: Brookings, 1982, esp. pp. 52–54. For representative views of risk-management officials see, e.g., William D. Ruckelshaus, "Science, Risk, and Public Policy," *Vital Speeches of the Day*, Volume 49, No. 20, August 1, 1983, pp. 612-615.

[24] See, e.g., Howard Kunreuther and Lisa Bendixen, "Benefits Assessment for Regulatory Problems," and Baruch Fischhoff and Louis Anthony Cox, Jr., "Conceptual Framework for Regulatory Benefits Assessment," Chapters 3 and 4, respectively, in *Benefits Assessment: The State of the Art*, Judith D. Bentkover, Vincent T. Covello, and Jeryl Mumpower, eds., Dordrecht, Netherlands: D. Reidel, 1986, pp. 44–45, 59–61.

[25] See U.S. Office of Science and Technology Policy, "Chemical Carcinogens: A Review of the Science and Its Associated Principles," Principle 29 (50 FR 10378, March 14, 1985, hereinafter, *OSTP Risk Assessment Guidelines*); U.S. Environmental Protection Agency, "Guidelines for Carcinogen Risk Assessment," 51 FR 34001 (September 24, 1986, hereinafter, *EPA Carcinogen Risk Assessment Guidelines*); U.S. Department of Health and Human Services, *Risk Assessment and Risk Management of Toxic Substances*, April 1985, p. 20.

ing upper bounds for these risks. Risk-assessment methods with similar conservative biases are less common elsewhere, particularly in those areas where real-world data are available, or where the mechanism by which injury or illness occurs is better understood.

A renewed commitment to the NAS recommendations is clearly warranted. As quantitative risk assessment plays an increasingly significant role in risk management, the need to separate science from policy becomes ever more important, if either process is to maintain public confidence. As former EPA Administrator William D. Ruckelshaus has noted:

> Risk assessment...must be based on scientific evidence and scientific consensus *only*. Nothing will erode public confidence faster than the suspicion that policy considerations have been allowed to influence the assessment of risk. [26]

ALTERNATIVE RISK-ASSESSMENT METHODOLOGIES

Risk assessments of chemical substances in general (and of possible carcinogens in particular) involve a mixture of facts, models, and assumptions. There is considerable debate concerning the scientific merits of the models and assumptions commonly used in risk assessments. In some cases, a scientific consensus has developed to support a particular model or assumption. In other instances, however, certain models and assumptions are relied upon because they reflect past practices rather than the leading edge of science. Furthermore, a scientific basis for several of the most critical models and assumptions simply does not exist.

Most scientists agree that these models and assumptions impart a conservative bias: that is, they lead to risk projections that the actual (but unknown) risk is very unlikely to exceed. These "upper-bound" estimates are often useful as a screening device, to exclude from regulatory concern potential hazards that are insignificant even under worst-case conditions. Unfortunately, upper-bound risk estimates are routinely employed for altogether different purposes, such as estimating the likely benefits of regulatory actions. Policymakers are required to act on the basis of biased representations of both the magnitude of the underlying hazard and the extent to which Government action will ameliorate it.

Contemporary risk assessment relies heavily upon animal bioassay and epidemiology. Each approach has theoretical advantages and disadvantages. In practice, both can be misused to bolster preestablished conclusions. The following discussion emphasizes problems in carcinogenic risk assessment, because the prevention and cure of cancer plays such a major role in policy issues involving risks to life and health.

Animal Bioassay

Animal testing enables scientists to estimate risks ex ante, before human health effects materialize, whereas epidemiological studies can only detect such effects ex post. In addition, animal tests can be conducted under tightly controlled laboratory conditions, which provide more reliable estimates of exposure and avoid many of the confounding factors that often plague epidemiological investigations. The relatively short lifetimes of experimental mammals (such as rats and mice) allow scientists to ascertain the possible effects of long-term exposure in just a few years.

Animal testing suffers serious limitations, however, arising from certain critical assumptions. Despite its routine application, there is no accepted scientific basis for the assumption that results can be meaningfully extrapolated from test animals to humans.[27] Some scientists believe that animal data should not be used in assessing human health risks.[28]

Another critical limitation is the reliance on very high doses to generate adverse effects in test animals.[29] A mathematical model must be used to bridge the gap between these high-dose exposures and the low-dose exposures more typically faced by people. Many different mathematical models can be constructed to fit the data at high doses. These models often vary enormously, however, in their predictions of risk at low doses.

Beyond these unavoidable methodological constraints, the results of animal bioassays may be subject to conflicting scientific interpretation or strongly influenced by the choice of research method.

[26] William D. Ruckelshaus, (*op. cit.*), p. 614.
[27] *OSTP Guidelines*, Guideline 8, p. 10376.
[28] See, e.g., Bruce Ames, Renae Magaw, and Lois Swirsky Gold, "Ranking Possible Carcinogenic Hazards," *Science*, Vol. 236, April 17, 1987; Gio Batta Gori, "The Regulation of Carcinogenic Hazards," *Science*, Vol. 208, April 18, 1980.
[29] *OSTP Guidelines*, Guideline 11, p. 10377.

Tissue preparation and histology present obvious opportunities for error, as experts may disagree as to how slides should be interpreted.[30] This problem generally is not significant at high doses, where malignancies are often obvious. At low doses, however, pathologists often differ in how they distinguish tumors from hyperplasia. Subjectivity cannot be avoided where such interpretations of the data must be made.[31]

Epidemiology

Epidemiology is attractive because it largely avoids these two problems. It focuses on observable human health effects instead of on hypothesized outcomes based on animal experimentation, and it relies upon real-world exposures to generate empirical data. Many of the serious problems associated with animal studies can be avoided, allowing researchers to develop risk estimates that are directly related to human health.

Unfortunately, epidemiological research suffers from its own set of limitations. For example, retrospective studies often have difficulty correlating morbidity and mortality with exposure to specific substances. Exposure data are commonly lacking, incomplete, imprecise, or affected by systematic recall or selection biases. Furthermore, the risks these studies seek to detect are often very small relative to background, thus making statistically significant effects difficult to observe. When health effects are latent, correlating exposures to illness is even harder.

Besides these unavoidable methodological limitations, epidemiological studies often suffer from outright bias. Many studies employ scientifically questionable procedures aimed at demonstrating positive relationships between specific substances and human illness.[32] Some researchers use inappropriate statistical procedures to "mine" existing databases in search of associations. One result of these practices is that epidemiological studies often display contradictory results.[33]

Despite these constraints, properly conducted animal bioassays and epidemiological studies both have useful roles to play in quantitative risk assessment. Indeed, they are complementary. The usual weaknesses of epidemiological investigations—unreliable exposure data, confounding effects—are readily avoided in laboratory experiments on animals. The weaknesses of animal bioassays—high- to low-dose extrapolation, animal-to-man conversion—do not arise in epidemiological studies. Careful risk assessment incorporates both types of analysis to ensure that the emerging picture of human health risk is as complete as possible, and that inferences derived from this picture are themselves internally consistent.

ISSUES IN RISK ASSESSMENTS DERIVED LARGELY FROM ANIMAL BIOASSAYS

Animal bioassays tend to dominate current risk assessments. An important reason for this is that the derivation of dose-response relationships is a critical regulatory motive for performing quantitative risk assessment. Animal studies are ideally suited to serve this purpose by virtue of the controlled conditions under which dose and response can be calibrated. Epidemiological studies often are relegated to providing merely a "reality check" to ensure that the implications of animal bioassays are plausibly consistent with real-world experience. Because of this heavy emphasis on animal testing, the focus here is on several major problems that arise with respect to risk assessments primarily based on the results of animal bioassays.

The Use of Sensitive Test Animals

To enhance the power of animal tests, scientists typically rely on genetically sensitive test animals. It

[30] In the original analysis of the rat bioassay used to derive the dose-response function for dioxin, 9 of 85 controls were said to develop liver tumors. An independent review of this data resulted in 16 of the 85 controls being classified as having such tumors. See U.S. Environmental Protection Agency, *A Cancer Risk-Specific Dose Estimate for 2, 3, 7, 8-TCDD, Appendix A*, EPA/600/6-88/007Ab, June 1988 (hereinafter, *Dioxin Risk Assessment Appendix A*), pp. 2–3.

[31] Colin N. Park and Ronald D. Snee, "Quantitative Risk Assessment: State-of-the-Art for Carcinogenesis," Chapter 4 in *Risk Management of Existing Chemicals*, Rockville, MD: Government Institutes, 1983, p. 56.

[32] Alvan R. Feinstein, "Scientific Standards in Epidemiological Studies of the Menace of Daily Life," *Science*, Vol. 242, December 2, 1988, pp. 1257–1263.

[33] Linda C. Mayes, Ralph I. Horowitz, and Alvan R. Feinstein, "A Collection of 56 Topics with Contradictory Results in Case-Control Research," *International Journal of Epidemiology*, Vol. 17, No. 3 (1988), pp. 680–685.

is unclear whether these species accurately mimic biological responses in humans.

Some test species are extremely sensitive. For example, approximately one-third of all male B6C3F1 mice, a common test species, spontaneously develop liver tumors.[34] The same phenomenon occurred in an important bioassay concerning dioxin using female Sprague-Dawley (Spartan) rats. Tumors observed in dosed animals were predominantly located in the liver. However, approximately one-fifth of the animals in the control group also developed liver tumors.[35] The relevance of elevated liver tumors in hypersensitive species has been questioned by scientists and is not universally considered probative evidence of carcinogenicity. Nevertheless, cancer risk assessments often proceed on the assumption that these data are sufficient to conclude that a substance is indeed a carcinogen.[36]

The reliance on sensitive test animals also biases risk assessments in a more subtle way. It establishes powerful incentives to search for and develop increasingly sensitive test species. As test animals become more sensitive, repeated testing using identical protocols will tend to result in higher and higher estimates of risk even if all other factors are held constant.

Selective Use of Alternative Studies

In their respective risk-assessment guidelines, both OSTP and EPA recommend that relevant animal studies should be considered irrespective of whether they indicate a positive relationship.[37] In practice, however, studies that demonstrate a statistically significant positive relationship routinely receive more weight than studies that indicate no relationship at all.[38] For example, the plant growth regulator daminozide (Alar) and its metabolite unsymmetrical 1,1-dimethylhydrazine (UDMH) recently received B2 classifications ("probable human carcinogen"). Each of these classifications was based on a single positive animal bioassay.[39] Overcoming such a classification requires, at a minimum, two "essentially identical" studies showing no such relationship.[40] In the case of Alar and UDMH, however, a more stringent test was apparently applied: Three high-quality negative studies showed no significant effects; these studies appear to have received little or no weight in the classification decision.[41]

Selective Interpretation of Results

Risk-assessment guidelines generally give the greatest weight to the most sensitive test animals. Thus, if a substance has been found to cause cancer in one

[34] Ames *et al., (op. cit.)*, p. 276.

[35] *Dioxin Risk Assessment Appendix A*, pp. 2–3.

[36] See Ames *et al., (op. cit.)*, p. 276 (arguing that such data are irrevelant); *OSTP Guidelines* Guideline 9, p. 10377 (concluding that such data "must be approached carefully"); and *EPA Carcinogen Risk Assessment Guidelines*, p. 33995 (making the policy judgment that such data are sufficient evidence of carcinogenesis). Liver tumors dominated in EPA's dioxin risk assessment. See *Dioxin Risk Assessment*, appendix A, pp.2–3.

[37] See *OSTP Guidelines*, Guideline 25, p. 10378; *EPA Carcinogen Risk Assessment Guidelines*, p. 33995.

[38] See *EPA Carcinogen Risk Assessment Guidelines*, p. 33999–34000. A single animal test that shows a positive result "to an unusual degree" (p. 33999) is sufficient to warrant at least a B2 classification ("probable human carcinogen"), even if this result occurs in a species known to have a high rate of spontaneous tumors. A strong animal bioassay or epidemiological study showing no evidence of carcinogenic effect cannot overcome this presumption (p. 34000).

[39] See "Second Peer Review of Daminozide (Alar) and UDMH (Unsymmetrical 1,1-dimethylhydrazine)," Memorandum from John A. Quest to Mark Boodee, U.S. Environmental Protection Agency, OPTS, May 15, 1989 (hereinafter, *Alar/UDMH Internal Peer Review No. 2*). This internal OPTS panel reviewed several recent studies on Alar and UDMH.

One study of Alar yielded a statistically significant increase in common lung tumors in mice, but only for one of three dosage levels. Results were not statistically significant at one higher and two lower dosages, and controls also displayed unusually high tumor incidence. 90% of the lung tumors in dosed mice were benign, versus 89% in the controls.

One study of UDMH yielded statistically significant increases in common lung and uncommon liver tumors in mice, but only for the higher of two dosages. 97% of the lung tumors in dosed mice were benign, versus 100% in the controls. 29% of the liver tumors in dosed mice were benign; no tumors were observed in the controls.

Prior studies that purported to show a carcinogenic response had been judged inadequate by EPA's Scientific Advisory Panel, an external peer review group. The Office of Pesticides and Toxic Substances (OPTS) panel noted that a different internal EPA risk-assessment panel (the Carcinogen Assessment Group) considered these studies sufficient to justify B2 classifications when it evaluated them for EPA's Office of Solid Waste and Emergency Response. Despite the scientific controversy, the OPTS panel interpreted these prior studies as "supporting evidence" under EPA's risk-assessment guidelines.

[40] See *EPA Carcinogen Risk Assessment Guidelines*, p. 33995 (establishing the need for replicate identical studies showing no effect), and p. 33999 (establishing the minimum requirement of two well-designed studies showing no increased tumor incidence to warrant a "no evidence" determination).

[41] *Alar/UDMH Internal Peer Review No. 2*, pp. 6, 8, 9. EPA's scheme for carcinogen classification is itself an issue among scientists. See, e.g., U.S. Environmental Protection Agency, Risk Assessment Forum, *Workshop Report on EPA Guidelines for Carcinogen Risk Assessment*, EPA/625/3-89/015, Washington, DC: March 1989, pp. 21–26.

species or gender but shown to exhibit no effects elsewhere, the results pertaining to the sensitive species or gender typically will be used to develop estimates of human-health risks. For example, if male mice develop cancer from a substance but female mice and rats of both genders do not, then the results from the male mouse often will be used to derive estimates of cancer risks to humans.[42]

Once a positive result has been obtained in an animal bioassay, a substance often will be provisionally classified as a probable human carcinogen. The statistical burden of proof then shifts to the no-effect hypothesis. Because it is logically impossible to prove a negative, however, this practice establishes a virtually irrebuttable presumption in favor of the carcinogenesis hypothesis.

Severe Testing Conditions

Current risk-assessment protocols require the use of very high doses. Unfortunately, high doses are often toxic for reasons unrelated to their capacity to cause cancer. A common procedure is to use what is called the maximum tolerated dose (MTD), which is the most that can be administered to a test animal without causing acute toxicity. At such exposure levels, substances often cause severe inflammation and chronic cell killing. For example, formaldehyde causes nasal tumors in rats when administered in high doses. However, MTD administration severely inflames nasal passage tissues. It is therefore unclear whether the cancers induced are caused by formaldehyde per se or by the toxic effects of high doses.

Results such as these have caused some scientists to question the validity of rodent tests performed at the MTD for estimating human health risks that arise from exposure at low doses.[43] By combining very high doses with highly sensitive test subjects, some animal bioassays are predisposed to discover apparent carcinogenic effects.

Relevance of Animal Bioassay Results

An important reason why animals vary in their sensitivity is that they have different physiologies, metabolic processes, reproductive cycles, and a host of other species-specific characteristics that largely result from unique evolutionary paths. Each of these factors needs to be carefully considered in evaluating the significance of animal data with respect to human health. This is recognized in both the OSTP and EPA guidelines, but it is often neglected when the guidelines are applied to specific substances.[44]

The most important assumption in this regard is that animal test results can be meaningfully extrapolated to humans. A recent study of chemicals tested under the auspices of the U.S. National Toxicology Program shows that this assumption can lead to the erroneous classification of many chemicals as probable human carcinogens.[45] Positive associations have been obtained in either rats or mice for half of 214 chemicals tested. However, results were consistent across these two genetically similar species only 70 percent of the time. If it is assumed that rodent bioassays have the same sensitivity and selectivity with respect to human carcinogens as they do between rodent species, and it is further assumed that 10 percent of all chemicals are in fact human carcinogens, then 27 of every 100 randomly selected chemicals would be misclassified as probable human carcinogens. Only three chemicals would be misclassified as noncarcinogens. Thus, "false positives" would be 9 times more common than "false negatives."[46]

Of course, this ratio of false positives to false negatives reflects highly conservative "upper-bound" assumptions concerning sensitivity and selectivity. Given the high degree of similarity between rats and mice and the limited resemblance between rodents and humans, the sensitivity of rodent bioassays with respect to human carcinogenicity is probably much lower than 70 percent. Furthermore, other research indicates that selectivity may be as low as 5 percent.

---

[42] See *EPA Carcinogen Risk Assessment Guidelines*, p. 33997 (data from long-term animal studies showing the greatest sensitivity should generally be given the greatest emphasis).

[43] See, e.g., Ames *et al.*, *(op. cit.)*, pp. 276–277.

[44] *OSTP Guidelines*, Guideline 25, p. 10378; *EPA Carcinogen Risk Assessment Guidelines*, p. 34003 (responding to comments on the draft guidelines and affirming agreement with OSTP Guideline 25).

[45] Lester B. Lave, Fanny K. Ennever, Herbert S. Rosenkranz, and Gilbert S. Omenn, "Information Value of the Rodent Bioassay," *Nature*, Vol. 336 (December 15, 1988), pp. 631–633.

[46] *False negatives* occur when a test fails to detect effects when they are in fact present. *Sensitivity* refers to the capacity of a test to minimize false negatives. *False positives* occur when a test appears to detect effects that in fact are absent. *Selectivity* refers to a test's ability to minimize false positives. The 9 to 1 ratio of false positives to false negatives calculated by Lave *et al.* assumes that both selectivity and sensitivity equal about 70%.

Adjusting only for this lower selectivity suggests that false positives are almost 30 times more common than false negatives. This raises serious questions concerning the practical utility of the current approach to animal bioassays for the purpose of quantitative risk assessment.[47]

Other factors should also be considered when relying upon animal bioassay results as the primary basis for quantitative risk assessments. For example, certain substances are toxic or even carcinogenic by one pathway but not by others. Nevertheless, animal bioassay protocols often emphasize the most sensitive pathway. As long as human exposure is likely to arise the same way, then this choice may be reasonable. However, the pathway to which the test species is sensitive sometimes reflects an exposure route that is implausible or irrelevant for humans. For example, formaldehyde causes nasal tumors in rats at 12 times the rate observed in the next most sensitive animal species. This extreme sensitivity may be related to the fact that rats breathe only through the nose.

There may be important differences between animals and humans that make specific tumors irrelevant. For example, some chemicals cause cancer in the zymbal gland of the rat; because humans lack such a gland it is unclear whether these results matter in estimating human health risk. Other substances induce cancer through biochemical mechanisms not found in humans.

A greater controversy surrounds the question whether the same weight should be given to benign and malignant tumors. The scientific consensus is that benign and malignant tumors should be aggregated only when it is scientifically defensible to do so.[48] In practice, however, benign and malignant tumors are routinely aggregated unless a strong case can be made *against* the practice.[49] The difference between these default assumptions is significant: One approach counts only carcinomas that *are* present, whereas the other counts tumors that *might become* carcinomas. In an extreme case, a substance that promotes benign tumors but never causes cancer could be classified as a probable human carcinogen simply because benign and malignant tumors are treated equally.

In addition, tumor incidence is commonly pooled across sites to obtain a total estimate of carcinogenic effects.[50] This implicitly assumes that cancer induction is independent across sites and not the result of either metastasis or the same biological mechanism. Given the extreme sensitivity of test species and the regular use of MTD administration, other explanations for tumors occurring at multiple sites appear just as plausible.

The Choice of Dose-Response Model

No single mathematical model is accepted as generally superior for extrapolating from high to low doses.[51] Consequently, Federal agencies often use a variety of different models. Rather than being a scientific footnote to the risk-assessment process, however, the choice of model is actually an important policy issue. The multistage model appears to be the most commonly used method for estimating low-dose risks from chemicals, and there are two major sources of bias embedded in this choice: its inherent conservatism at low doses, and the routine use of the "linearized" form in which the 95 percent upper bound is used instead of the unbiased estimate.

The *multistage model* essentially involves fitting a polynomial to a data set, with the number of "stages" identified by the number of terms in the polynomial. Since animal bioassays rarely have more than three dose levels, it is unusual to see applications of the multistage model with more than two stages. Although the multistage model enjoys some scientific support because it is compatible with multistage theories of carcinogenesis, in practice the model fails to include enough stages, due to the absence of sufficient alternative exposure cohorts.

The multistage model typically yields low-dose risk estimates that are higher than most other models. For example, when five different dose-response models were analyzed in a recent risk assessment of cadmium, estimates of cancer risks at moderate doses varied by a factor of 100. This difference among

[47] Lave *et al., (op. cit.)*, p. 631. Adjusting also for less sensitivity reduces the ratio of false positives to false negatives. For example, if sensitivity is only 10 percent and all other parameters remain unchanged, then this ratio declines to 9.5 to 1. However, this implies that both types of statistical errors are rampant, which raises questions concerning the practical utility of animal bioassays. This is, in fact, precisely the concern raised by Lave *et al., (op. cit.)*, who conclude that such tests are cost-effective investments in information only under extraordinary conditions.

[48] *OSTP Guidelines*, p. 10376.

[49] *EPA Carcinogen Risk Assessment Guidelines*, p. 33997.

[50] *Id.*

[51] *OSTP Guidelines*, Guideline 26, p. 10378; Ames *et al., (op. cit.)*, p. 276.

estimates widened as doses declined toward the very low levels within the range of regulatory concern. At very low doses, two of the five models predicted excess lifetime cancer risks greater than one in one thousand (10^{-3}), a risk oftentimes regarded by policymakers as unacceptable. However, two other equally plausible models predicted essentially no excess cancer risk at all. Since none of the five models offers a scientifically superior basis for deriving low-dose risks, the choice of model is therefore a pivotal policy decision. The accepted practice under these circumstances is to develop a subjectively-derived "best" estimate while fully informing decisionmakers as to the extent of uncertainty surrounding it.[52] In the cadmium case, as in most others, this practice was not followed: Estimates of the number of statistical cancers that would be prevented by regulation were presented based only on the multistage model.[53]

The *linearized multistage model (LMS)* is a special version of the multistage model in which the 95 percent upper confidence limit of the linear term is used instead of the unbiased estimate. That is, the model identifies the largest value for the linear term that cannot be rejected at the 95 percent confidence level and uses it in place of the unbiased estimate. Assuming that the model has been correctly specified, there is only a 5 percent chance that the true risk exceeds this level.

The LMS has become the preferred statistical approach because estimates derived from it appear to be more "stable" than estimates obtained from the ordinary multistage model. The "stability" issue originally arose because unbiased estimates of low-dose risks are very sensitive to the maximum-likelihood estimate (MLE) of the value of the linear term. When the MLE of the linear term is positive, it dominates estimated risks at low doses. In some instances, however, the MLE of the linear term is zero, and low-dose risk estimates decline precipitously. Using the 95 percent upper confidence limit ensures that the linear term is always positive, thus eliminating the inherent "instability" of low-dose risk estimates derived from the multistage model. [54]

Another often-cited advantage of the LMS procedure is that it provides a "yardstick" for comparing potencies across chemicals.[55] A uniform risk-assessment procedure such as the LMS, it is argued, enables policymakers to better understand the relative significance of a broad array of chemical hazards and set regulatory priorities accordingly.

Finally, the LMS is often defended on the ground that it is prudent to err on the side of caution when dealing with potentially carcinogenic chemicals. Because the LMS generates upper-bound risk estimates, policymakers can be confident that actual risks are likely to be lower.

None of these purported advantages of the LMS approach has a sound statistical basis. It is a fundamental axiom of statistics that unbiased estimates are generally preferred to biased ones. Using the upper confidence limit instead of the unbiased estimate exaggerates underlying specification errors instead of eliminating them. "Instability" is overcome, but at the cost of greater errors in specification.

The inherent instability of the multistage model reflects a generalized misspecification of dose-response—that is, the real human dose-response relationship is often very different from what the multistage model constrains it to be. The model is extremely sensitive to small differences in observed tumor incidence, which can cause dramatic changes in estimated low-dose risks. The LMS procedure eliminates this sensitivity without remedying the underlying specification error. Proper statistical procedure requires correcting model misspecification, not masking its symptoms behind biased parameter estimates.

The LMS procedure inflates low-dose risk estimates by a factor of two or three when the MLE of the linear term is positive. However, it increases low-dose risk estimates by orders of magnitude when the MLE of the linear term is zero.[56] This means that the degree of hidden conservative bias is substantially greater for what are demonstrably lower risks.

By its very nature, the LMS cannot serve as a useful yardstick for comparing the relative risk of a variety of potential carcinogens. If a given statistical procedure generated identical biases across substances tested,

[52] See, e.g., *OSTP Guidelines*, Guidelines 27, 29, and 31, p. 10378; *EPA Carcinogen Risk Assessment Guidelines*, pp. 33999, 34003.

[53] Occupational Safety and Health Administration, "Occupational Exposure to Cadmium; Proposed Rule," 55 FR 4076 (February 6, 1990).

[54] Albert L. Nichols and Richard J. Zeckhauser, "The Dangers of Caution: Conservatism in Assessment and the Mismanagement of Risk," Chapter 3 in *Advances in Applied Micro-Economics, Volume 4: Risk, Uncertainty, and the Valuation of Benefits and Costs*, V. Kerry Smith, ed., Greenwich, CT: JAI Press, 1986, pp. 55–82, esp. pp. 62–63. A nontechnical version of this paper is available by the same authors as "The Perils of Prudence: How Conservative Risk Assessments Distort Regulation," *Regulation*, November/December 1986, pp. 13–24.

[55] U.S. Environmental Protection Agency, *A Cancer Risk-Specific Dose Estimate for 2,3,7,8-TCDD*, EPA/600/6-88/007Aa, June 1988 (hereinafter, *Dioxin Risk Assessment*), pp. 45–46.

[56] Nichols and Zeckhauser, *op. cit.*, pp. 62–63.

then it would still yield an accurate rank-ordering of theoretical hazards. Similarly, if the procedure added a stochastic bias from a uniformly distributed random variable, the resulting rank-ordering would still be accurate on an expected-value basis. The problem with the LMS is that it generates biases that intensify with the degree to which the multistage model misspecifies the true dose-response relationship. Even if the multistage model provided an accurate rank-ordering of hazards, the LMS could not do so, because it injects biases that are systematic with statistical misspecification.

The LMS procedure (and the multistage model itself) is also fatally flawed as a yardstick for regulatory priority setting because it fails to take account of human exposure in the calculation of unit risks. Regardless of the procedure's capacity to accurately rank-order *hazards*, failing to adjust unit risks by relative human *exposure* virtually guarantees that regulatory priorities will be misordered. Resources tend to be focused on reducing the greatest theoretical *hazards* rather than the most significant human health *risks*.[57]

Finally, the "margin of safety" argument in favor of the LMS unequivocally contradicts the widely recognized need to distinguish science from policy.[58] The LMS introduces into each risk assessment a conservative bias of varying but unknown magnitude. This practice fundamentally alters regulatory decisionmaking. Instead of leaving policy decisions to policymakers, the LMS disguises fundamental policy decisions concerning the appropriate margin of safety behind the veil of science.

In summary, the LMS cannot be justified as a method of scientific risk assessment. The "yardstick" defense implicitly asserts that scientific advancements in risk-assessment methodology should take a back seat to the preservation of an outdated and misguided statistical procedure. The "margin of safety" argument tacitly usurps from policymakers the authority and responsibility for risk-management decisions. Finally, the statistical "instability" overcome by the LMS is an artifact of specification error, not any scientific theory of human carcinogenesis that warrants the intentional use of biased parameter estimates. The habitual reliance upon either the multistage model or its LMS descendant cannot be supported by sound scientific principles.

Alternative models are available, of course, and they have been applied in many quantitative risk assessments. Because proper model specification is the foundation of applied statistical methodology, alternatives to the multistage model should be expected and encouraged. Indeed, innovation is the hallmark of scientific inquiry; policies that institutionalize any particular model specification effectively stifle scientific advancement.

Unfortunately, models other than the multistage model are often discouraged in practice.[59] Agencies may require substantial scientific evidence in support of an alternative model before allowing it to be used. Alternative models thus face a burden of demonstrating scientific plausibility that the multistage model cannot satisfy. Even in the extraordinary case in which this burden can be satisfied, estimates may be required from the linearized multistage model anyway.[60]

The potential human health threat posed by dioxins provides an excellent example of the problem of model selection. Using the same linearized multistage model, EPA, the Centers for Disease Control (CDC), and the Food and Drug Administration (FDA) have arrived at upper-bound risk estimates that span an order of magnitude.[61] Depending on the data and assumptions used, the linearized multistage model predicts unit risk factors that vary by as much as 1,200, with the

[57] Some scientists have attempted to devise alternative indexes of relative human health risk that explicitly account for variations in human exposure. Ames *et al.*, (*op. cit.*), pp. 272–273, describe one such alternative (the Human Exposure/Rodent Potency index, or HERP) and report index values for 36 substances. Because the HERP index is based on a relative rather than absolute scale, the distorting effect of conservative biases embedded in the underlying risk assessments has been significantly reduced. Many substances suspected of being environmental carcinogens rank very low on the HERP index, suggesting that regulatory priorities have been seriously misdirected.

[58] See, *e.g.*, *NAS Risk Management Study*, p. 161; *OSTP Risk Assessment Guidelines*, Principle 29, p. 10378; and *EPA Carcinogen Risk Assessment Guidelines*, p. 34001.

[59] See, *e.g.*, Ames *et al.*, (*op. cit.*), p. 276 (continued reliance on linear models despite the accumulation of evidence against linearity); and Lester B. Lave, "Health and Safety Risk Analysis: Information for Better Decisions," *Science*, Vol. 236, April 17, 1987, pp. 291–295, esp. p. 292 (agencies often resist modeling improvements and data that yield lower risk estimates).

[60] *EPA Carcinogen Risk Assessment Guidelines*, pp. 33997–33998. "In the absence of adequate information to the contrary, the linearized multistage procedure will be employed. . . . Considerable uncertainty will remain concerning responses at low doses; therefore, in most cases, an upper-limit risk estimate using the linearized multistage procedure should also be presented."

[61] *Dioxin Risk Assessment Appendix A*, p. 13. Unbiased risk estimates vary by a similar factor.

three risk estimates mentioned earlier clustered at the high end of the range.[62] Risk assessments based on different models have led other governments to establish unit risk factors that are a thousand times less stringent than the most commonly used of these three; one study suggests that this particular estimate overstates the most likely risk estimate by a factor of almost 5,000.[63]

Conversion from Animals to Humans

Once risk has been extrapolated to low doses in rodents, scientists must convert them to human dose-equivalents. The two most common approaches involve the use of body-weight or surface-area conversions, and there are scientific reasons for choosing either approach in individual cases. The surface-area approach leads to estimates of risk that are between 7 and 12 times greater than those based on the body-weight method, depending upon the test species. Despite the ambiguity of the underlying science, the more conservative surface-area method is often applied reflexively.[64]

ISSUES ARISING FROM HUMAN EXPOSURE ESTIMATES

In addition to developing estimates of the dose-response function, agencies must estimate the likely level of human exposure. This section examines some of the issues and problems that arise in conducting an exposure assessment.

It is a generally accepted principle of exposure assessment that estimates should be based on the most likely scenario, with appropriate consideration of uncertainty.[65] Nevertheless, agencies often use conservative assumptions for exposure when real-world data are unavailable. When each of these assumptions tends to overstate likely human risks, the multiplicative effect of even a small overstatement at each stage in an exposure assessment will yield a substantial overestimate of actual exposure. For example, the multiplicative effect of overstating risk by a factor of two at five different points in an exposure assessment will overstate actual risk by a factor of thirty-two.

Worst-Case Environmental Conditions

When data are available they often relate to unusually sensitive environments or highly contaminated conditions. When estimating regional or nationwide exposures, agencies often use data from these local "hot spots" in developing more general national estimates of health risks. However, such data are never representative and estimates extrapolated from them are generally unreliable and misleading.

In addition, chemicals often degrade naturally after they have been released to the environment. In some cases, degradation occurs very quickly, whereas in others the process may take many years or even decades. A common practice in exposure assessment modeling is to assume that exposures remain constant over time—that is, chemicals are assumed never to degrade, or degradation by-products are assumed to pose identical risks.

The Maximum-Exposed Individual

In addition to estimating the amount of a substance that may actually be present in the environment, a risk analysis must also consider the conditions under which humans may be exposed. Actual risks vary considerably depending on location, mobility, and a host of other factors. Nevertheless, estimates often are based on the upper-bound lifetime cancer risk to the maximum-exposed individual (MEI), the hypothetical person whose exposure is greater than all others. Sometimes, risks to the entire population are estimated by assuming that everyone is exposed at the MEI level. Because environmental regulations are often justified using MEI-based risk assessments, actual risks may be substantially lower than what decisionmakers and the general public perceive them to be.

---

[62] *Dioxin Risk Assessment*, pp. 46–49. 10^{-6} risk-specific doses (RsDs) derived from the linearized multistage model span the range from 0.001 to 1.2 picogram/kg/day. The RsDs of EPA, CDC, and FDA are 0.006, 0.03, and 0.06 pg/kg/day, respectively.

[63] *Dioxin Risk Assessment*, p. 4.

[64] *EPA Carcinogen Assessment Guidelines*, p. 33998. "EPA will continue to use this [surface area] scaling factor unless data on a specific agent suggest that a different scaling factor is justified."

[65] EPA guidance documents have historically called for unbiased estimates of exposure. See, e.g., U.S. Environmental Protection Agency, "Guidelines for Exposure Assessment," 50 FR 34042–34054 (September 24, 1986, hereinafter, *EPA Exposure Assessment Guidelines*); U.S. Environmental Protection Agency, *Superfund Public Health Evaluation Manual*, OSWER Directive 9285.4–1, October 1986; and U.S. Environmental Protection Agency, *Superfund Exposure Assessment Manual* (Revised Draft), OSWER Directive 9285.5–1, December 1986. EPA recently abandoned the calculation of unbiased exposure estimates for Superfund sites on the ground that it was insufficiently conservative. EPA's new protocol requires the estimation of "reasonable maximum exposure" instead of the average and upper-bound estimates. Reasonable maximum exposure constitutes a new term of art that EPA intends to be "well above the average case" but not as extreme as the upper-bound. It provides a new opportunity for embedding conservative assumptions into exposure assessment and exaggerating estimates of actual human-health risk at Superfund sites. See *Risk Assessment Guidance for Superfund, Volume I: Human Health Evaluation Manual (Part A), Interim Final*, EPA/540/1-89/002, December 1989, Chapter 6, pp. 5, 47–50.

In developing the MEI risk level, analyses invariably assume that the level of exposure is continuous over a 70-year lifetime. This assumption overstates actual risks, because people are mobile, encounter a constantly changing portfolio of daily risks to life and health, and can take actions that reduce risk.

Assumptions vs. Real-World Exposure Data

The thread that connects these exposure assessment issues is that simple constructs which overstate exposure are typically used in lieu of real-world data, often because such data are unavailable. The risk estimates generated by these models depend on the validity of their assumptions; even small biases in exposure assessment assumptions can result in a substantial overstatement of risk.

For example, regulatory agencies may not have statistically reliable real-world data on pesticide residues in agricultural products. They also may not know the proportion of a given crop that has been treated with a particular pesticide. A common resolution of these uncertainties is to assume that residues are equal to the regulatory "tolerance"—the maximum level allowed to be present in food sold in interstate commerce—and that 100 percent of the relevant crop has been treated. Both assumptions overstate actual exposure, but are encouraged by agency guidance as a way to instill conservatism in risk assessment.[66] When data are available, however, the extent of this conservative bias becomes evident. In a recent special review for the pesticide Captan, for example, EPA reduced its earlier upper-bound lifetime cancer risk estimate by two orders of magnitude when it replaced the original conservative assumptions with real-world data. Even with these improvements, EPA still reported that upper-bound risks were probably overstated. For example, field tests were performed based on applications at the maximum legal rate and as close to harvest as the label permits. Similarly, feeding studies assumed that animal diets were dominated by feedstuffs that happened to contain high residues relative to other feedstuffs, such as almond hulls and raisin waste. As EPA noted, even if these assumptions accurately represented typical animal diets, they would do so only for portions of California where these crops are grown; nationwide extrapolations based on these "hot-spots" would very likely overstate exposure.[67] Since two of the highest product-specific risks were attributed to milk and meat, these remaining conservative biases can be expected to be significant.

IMPLICATIONS OF CONSERVATIVE RISK ASSESSMENT FOR RISK MANAGEMENT AND REGULATORY DECISIONMAKING

The primary purpose of risk assessment is to provide data as a basis for risk management decisions. Providing useful data requires the synthesis of information concerning risks and exposure levels into a coherent package that can be used to develop regulatory options. Decisionmakers then can use these risk estimates in evaluating regulatory alternatives. Unfortunately, the way in which risk information is characterized tends to overstate risks, making them appear much greater than they are likely to be. As a result, decisionmakers may make regulatory choices that are very different from the ones they would make if they were fully informed.

Quantification of Uncertainty

In accordance with the recommendations of the National Academy of Sciences, the *OSTP Guidelines* explicitly call for the quantification of uncertainty, particularly as it arises in the selection of dose-response models and exposure assumptions.[68] Unfortunately, Federal regulatory proposals that utilize risk assessment rarely provide this information, nor do they analyze the implications of uncertainty for decisionmaking. Instead, many risk assessments only identify a lifetime upper-bound level of risk.[69]

The differences between upper-bound and expected-value estimates may be considerable. As we indicated earlier, the upper-bound risk estimate for dioxin may be 5,000 times greater than the most likely estimate. Plausible risk estimates for perchloroethylene (the primary solvent used in dry cleaning) vary by a factor of about 35,000.[70]

In some instances, decisionmakers may not be informed that risk estimates differ because of policy choices hidden in the risk-assessment methodology. In EPA's proposed rule limiting emissions from coke

[66] *EPA Exposure Assessment Guidelines*, p. 34053. "When there is uncertainty in the scientific facts, it is Agency policy to err on the side of public safety."

[67] See, e.g., U.S. Environmental Protection Agency, "Captan: Intent to Cancel Registrations; Conclusion of Special Review," 54 FR 8127–8128 (February 24, 1989).

[68] *OSTP Guidelines*, (Guideline 27), p. 10378.

[69] See, e.g., *EPA Carcinogen Risk Assessment Guidelines*, p. 33998.

[70] Nichols and Zeckhauser, *(op. cit.)*, pp. 64–65.

ovens, for example, cancer risks were estimated based on the LMS model—a model that is designed to yield upper-bound estimates of risk. In previous rules involving similar types of risks, however, EPA used the unbiased maximum likelihood estimate. To the extent that decisionmakers were not informed that the higher estimate of risk was largely due to a different low-dose extrapolation procedure, regulatory decisions based on this risk assessment were likely to reflect misunderstanding rather than science.[71]

Plausible estimates of likely cancer risk can often be found buried in regulatory background documents. However, *Federal Register* rulemaking notices seldom present such estimates alongside upper-bound estimates. This practice overstates baseline human health threats, as well as the amount of risk reduction that may be accomplished by regulation. Policymakers and the public are misled because they typically see only the upper-bound estimates of the threat.

The prevalent Federal agency practice is to calculate the benefits of Federal regulatory initiatives based solely on upper-bound estimates of risk and exposure. In a recent proposal to reduce occupational exposure to cadmium, for example, the Occupational Safety and Health Administration (OSHA) developed risk estimates based on five alternative models for animal data, and two alternative models for human data. Across these seven data/model combinations, estimated excess lifetime cancer risk at the least stringent of the two proposed exposure standards varied from 0 to 153 cases per 10,000 workers occupationally exposed for 45 years. OSHA based its proposed exposure standards on *one* of these data/model combinations—the multistage model applied to animal data. This data/model combination predicted an excess lifetime cancer risk of 106 per 10,000 exposed workers, and was used to estimate aggregate cancer incidence and the risk-reduction benefits attributable to the new standard. Uncertainties in the underlying risk assessment, which span several orders of magnitude, were not carried forward through the exposure assessment and benefit calculation stages. This analytic error effectively obscured the uncertainty surrounding the true incidence of cadmium-induced lung cancer, and resulted in benefit estimates that may exceed actual reductions in occupational illness by several orders of magnitude.[72]

Misordered Priorities, Perverse Outcomes

Logically, one would expect that the routine overstatement of likely risks would lead to inefficient regulatory choices. Decisionmakers, convinced that a certain substance or activity poses a significant threat to public health, might well take actions that they would otherwise resist. Alternatively, they might take actions that address the wrong real-life risks.

To the extent that risk assessments differ in the degree to which they adopt conservative assumptions, it is difficult to determine which activities pose the greatest risks and hard to establish reasonable priorities for regulatory action. Because conservatism in risk assessment is especially severe with respect to carcinogens, it is reasonable to expect that other health and safety risks tend to receive relatively less attention and weight. As a result, society may actually incur greater total risk, because of misordered priorities caused by conservative biases in cancer risk assessment.[73]

A perverse and unfortunate outcome of using upper-bound estimates based on compounded conservative assumptions is that the practice may actually increase risk, even in situations where cancer is the only concern. Regulatory actions taken to address what are in fact insignificant threats may implicitly tolerate or ignore better known, documented risks that are far more serious. For example, before it was banned, ethylene dibromide (EDB) was used as a grain and soil fumigant to combat vermin and molds. Vermin transmit disease, and molds harbor the natural and potent carcinogen aflatoxin B. The estimated human cancer risk from the aflatoxin contained in one peanut butter sandwich is about 75 times greater than a full day's dietary risk from EDB exposure. On this basis alone, it might have been appropriate to accept a small increase in cancer risk from EDB to reduce the much larger cancer risk from aflatoxin. By eliminating the relatively small hazard from EDB, Federal risk managers may have intensi-

[71] Letter from Wendy Gramm (Administrator of the Office of Information and Regulatory Affairs) to Lee Thomas (Administrator of the Environmental Protection Agency), August 12, 1986, p. 3.

[72] Occupational Safety and Health Administration, "Occupational Exposure to Cadmium; Proposed Rule," 55 *Federal Register* 4076, 4080, 4093.

[73] This is precisely the policy issue raised by Nichols and Zeckhauser, (*op. cit.*), pp. 69–71, who note that EPA's 1985 decision to limit lead in gasoline was threatened by concerns about potential increases in benzene exposure. Any tradeoff between lead and benzene risks would have been biased against lead; as estimates of benzene risks are more conservative simply because it is a carcinogen, whereas lead is not.

fied the relatively potent threat of aflatoxin associated with an increase in the prevalence of mold contamination.[74]

The emphasis on risks faced by the maximum-exposed individual may also cause a perverse result by increasing overall population risks. For example, EPA's proposed regulation of the disposal of sewage sludge would probably create more public health risk than it eliminates. The proposal outlines a regulatory scheme that would shift disposal from generally safe practices to relatively risky alternatives. Thus, setting sludge quality standards to achieve an MEI upper-bound lifetime cancer risk of one in 100,000 (10^{-5}) would prevent 0.2 statistical cancer cases resulting from monofilling and land application. However, it would cause 2.0 additional statistical cancers by forcing a shift away from these disposal approaches toward incineration.[75]

These problems can be addressed by providing decisionmakers with the full range of information on the risks of a substance or an activity. Thus, decisionmakers should be given the likely risks as well as estimates of uncertainty and the outer ranges of the potential risk. Then, if regulatory decisionmakers want to choose a very cautious risk management strategy, they can do so and a margin of safety can be applied explicitly in the final decision. This approach is superior to one in which the expected risk and an unknown margin of safety are hidden behind the veil of a succession of upper-bound estimates adopted at key points in the risk-assessment process.

The public and affected parties also benefit from knowing both the expected risk and the margin of safety rather than being given upper-bound estimates that are probably very different from actual risks. People are likely to have a better intuitive understanding of the significance of averages than they have of unlikely extremes. To the extent that a margin of safety is appropriate—perhaps to protect unusually sensitive subpopulations—the magnitude of this margin can be more readily communicated if made explicit. In addition, providing information in this way should help improve public confidence in quantitative risk assessment as the basis for decisionmaking.

AVOIDING CONSERVATIVE BIASES IN RISK ASSESSMENT

Risk assessment remains a powerful and useful scientific tool for estimating many of the risks that arise in a technologically advanced society. Unfortunately, it is also susceptible to hidden biases that may undermine its scientific integrity and the basis for policymakers' reliance on such information in risk management decisions. For policymakers and the public to continue to rely on risk assessment in the development of regulatory initiatives, a renewed effort must be made to separate science from policy and provide risk information that is both meaningful and reliable.

Expected Value Estimates

Perhaps the most important current need in regulatory decisionmaking is for carefully prepared and scientifically credible estimates of the likely risks involved. Relying on worst-case analysis based on extremely conservative risk assessment and exposure models leads to widespread misunderstanding on the part of both Government officials and individual citizens. Decisionmakers at all levels need unbiased and impartial risk information so they can focus their attention on significant problems and avoid being distracted by minutiae.[76]

Weight-of-Evidence Determinations

Similar procedures are needed for assigning weights to each relevant study in the risk-assessment literature. Current practice gives undue weight to studies that show positive relationships. Resulting risk classifications are thus conservatively biased estimates derived from samples of similarly biased observations.

Full Disclosure

Efficient and responsible decisionmaking requires that policymakers and the public be fully informed about the implications of the regulatory alternatives among which they must choose. Meeting this requirement demands a careful discrimination between science and policy. When risk estimates depend on assumptions and judgments instead of data, the meaning and implications of these nonscientific parameters must be clearly articulated.

Avoiding Perverse Outcomes

Careful attention needs to be paid to the likely results of regulatory alternatives, with an eye toward avoiding choices that have the perverse effect of increasing net risk. All human activity involves risk.

[74] Ames *et al.*, *op. cit.*, p. 273.

[75] U.S. Environmental Protection Agency, "Standards for the Disposal of Sewage Sludge; Proposed Rule," 54 FR 5746–5902 (February 6, 1989).

[76] Nichols and Zeckhauser, *op. cit.*, pp. 72–76.

responses to scientific uncertainty. In combination, however, they result in a distortion equal to the product of the individual conservative biases. To illustrate, suppose that there are ten independent steps in a risk assessment and prudence dictates assumptions that in each instance result in risk estimates two times the expected value. Such a process would yield a summary risk estimate that is [illegible] in quantitative risk assessments, policy choices are distorted from the course that would have been selected if decisionmakers had been better informed of the actual risks. Ironically, these policy decisions may actually increase total societal risk. Too much attention is focused on relatively small hazards that have been exaggerated by conservative risk assessments, leaving alone larger risks that have been estimated using unbiased procedures.

Information as an Alternative Regulatory Strategy

Federal regulation was initiated to deal with economic problems caused by monopoly and so-called "excess competition." Subsequent events have shown that, in general, economic regulation—fixing prices, establishing restrictive terms of trade, and erecting barriers to entry—is usually inefficient and detrimental to innovation. In response to these lessons, Federal regulation of this type has been under increasing criticism. As indicated above, however, much more needs to be done to reform economic regulation and restore competition.

Federal regulation has more recently been initiated to deal with what economists call externalities, situations in which participants in voluntary market transactions do not bear the full costs or capture all of the benefits of these exchanges. Common examples of externalities include environmental pollution and traffic congestion, common property resources such as fisheries and public forests, and "public goods" such as basic scientific research. In each of these instances, regulation may be an appropriate mechanism to modify or restore distorted market processes, or to establish markets where heretofore they have not existed, to maximize net social benefits (including environmental, health, and safety benefits). The key ingredient is the determination that existing markets are, in some significant manner, failing to perform efficiently.

The traditional regulatory approach to externalities has been the promulgation of standards. Because this approach often remedies existing externalities by creating new ones, economic incentive instruments are becoming an increasingly popular alternative to standards. The principal attraction of economic incentives is that they rely on market forces rather than attempt to suppress them.

This section explores another alternative regulatory strategy—the production, provision, or mandated disclosure of information. The first subsection briefly summarizes the economics of information as it relates to regulatory decisionmaking. Three points stand out in this discussion. First, because information is costly to acquire and the capacity to process it is limited, there is an optimal level of information for every market transaction. Second, differences in the amount and quality of information between buyers and sellers are normal and do not necessarily indicate market failure. Rather, these differences generally reflect variations in the costs and benefits that are attributable to information. Third, competitive markets provide powerful incentives for buyers and sellers to reveal relevant information. Market processes, not government regulations, provide the dominant motivation for generating, acquiring, and disclosing information. The role of government regulation thus should be to supplement these processes when they prove to be inadequate, not to supplant them when they work well.

The second subsection identifies three rationales for government intervention in the production or mandated disclosure of information. Two of these are economic—the public-good character of some types of