Statistical Science
1988, Vol. 3, No. 3, 328–338

Sources and Policy Implications of Uncertainty in Risk Assessment

Jack Needleman

Abstract. While risk assessment is an element of both regulatory and nonregulatory decision making, the role played by these studies in agency risk management decisions has proven to be limited. Part of the reason for this is the role played by considerations other than risk in the decision making of public agencies. In the risk management process, either formal or informal consideration is given to at least four factors: the feasibility of controlling exposure, the costs of control and economic impacts, the balance of costs and benefits and the importance of the product or agent suspected of causing harm. Part of the reason is uncertainty regarding the findings of risk assessment because of methodologic limits or disagreements among analysts, and the impact of these uncertainties on policymakers. Policymakers come into office with different attitudes about how risk averse government policy should be, and uncertainty in the risk assessment process allows policymakers considerable latitude to interpret the evidence and make decisions consistent with attitudes. Outside groups are actively involved in reviewing and conducting risk assessments, however, and this has both brought additional resources to these tasks and served as a source of external peer review.

Key words and phrases: Risk assessment, carcinogenesis, environmental health policy, health effects of toxic substances, risk methodology, bioassay.

Many agencies of the government are involved in discussions affecting the public and private management of risks. Most of the Public Health Service (PHS) agencies are involved—including the Food and Drug Administration (FDA), National Institutes of Health (NIH), Centers for Disease Control (CDC) and agencies in the Office of the Assistant Secretary for Health (OASH). The Environmental Protection Agency (EPA), the Occupational Safety and Health Administration (OSHA), the Consumer Product Safety Commission (CPSC), the Department of Transportation (DOT), the Mine Safety and Health Administration (MSHA), the United States Department of Agriculture (USDA) and other agencies are also closely involved in these activities.

Risk assessment has become a major activity of these agencies as they pursue their risk management efforts. Risk assessments may be conducted for a variety of reasons. In many cases, they are part of the process of regulating exposure to physical or chemical hazards, either pre-emptively through premarketing product approval processes such as FDA's procedure for new drugs and EPA's approval of new pesticides, or through corrective regulation designed to eliminate or reduce hazards already present in the environment. Nonregulatory uses of risk assessment include education and guidance to individuals and health professionals on actions they can take to reduce risks to health, and advice on appropriate medical screening protocols, diet and life style. Risk assessments may also be undertaken to help agencies make program decisions—such as whether to undertake mass screening or how to target direct research funds—or to provide information needed for litigation or compensation. Because the stakes in risk management decisions (whether measured in terms of health and personal well-being, social or economic costs) can be high, the methods used to assess risk need to provide estimates of risk that have the full and justified confidence of both policymakers and the public.

It is clear that without risk assessment studies providing substantial indications that a substance or agent is a potential hazard to human health, no action is likely. Even with such studies, action may not be taken. This paper explores the role played by risk

Jack Needleman is Vice President, Lewin/ICF, 1090 Vermont Avenue, Washington, D. C. 20005.

assessment in risk management, and the impact of uncertainty in the risk assessment process.

The paper is based upon an examination of the risk assessments of the ten hazards by a range of federal agencies. These cases, listed in Table 1, were selected to include a variety of different hazards and risk assessment methods and to illustrate approaches in both regulatory and nonregulatory settings. The cases were analyzed by reviewing formal risk assessment documents, interviewing agency personnel involved in the development of the risk assessments, reviewing other background materials and analyses, and in some cases, interviewing senior agency officials and policymakers and knowledgeable outside parties about the particular risk assessment and the general problems of using risk assessment methods.

This review suggested several reasons why the role played by risk assessments in risk management is limited:

- Health effects and the assessment of health risks are only one consideration in the risk management process. This is not an issue that can be addressed within the risk assessment process per se.
- Methodologic limits and weaknesses create uncertainties about the conclusions that can be drawn from risk assessment. Many of the assumptions underlying risk assessment methods, while plausible, have not been demonstrated empirically. Ongoing disagreements over methods allow policymakers considerable latitude to interpret evidence and make decisions consistent with their personal experience and attitudes regarding how averse public policy should be.
- Risk management decisions are influenced by political considerations and publicity regarding individual issues. Outside groups are involved in significant risk management decisions. This has served as a source of outside review and brought additional resources to risk assessment activities.

Each of these factors is reviewed in this paper.

THE ROLE OF HEALTH EFFECTS IN RISK MANAGEMENT DECISIONS

Health effects may be only one consideration in the risk management process. In the cases studied, health effects dominated those from PHS that involved public and professional education. In the regulatory cases, other factors, many specified by statute, also proved important.

The risk management process, formally or informally, considers the following factors in addition to health effects:

- Technical feasibility of control or of reduction exposure.
- Costs of controlling or reducing exposure, and the economic impact of imposing those costs on specific industries and the economy as a whole.
- The balance of costs and benefits.
- The importance of the product or agent, as indicated by the amount used and the availability of substitutes.

Political pressures and public concern also played a role in the risk management process.

TABLE 1
Case studies

Agent	Hazard(s) examined	Agencies	Anticipated use of risk assessment
Ethylene dibromide	Carcinogenesis Reproductive	EPA OSHA	Regulatory-corrective[a]
Formaldehyde	Carcinogenesis Irritation	EPA OSHA CPSC	Regulatory-corrective
Tris	Carcinogenesis	CPSC	Regulatory-corrective
Dioxin in Missouri	Carcinogenesis	CDC	Regulatory-corrective
Lead	Reproductive	EPA OSHA	Regulatory-corrective
Cotton dust	Microbial or chemical	OSHA	Regulatory-corrective
Noise	Physical	EPA OSHA	Regulatory-corrective
Passive smoking	Carcinogenesis	OSH	Public professional education[b]
Dietary fat	Carcinogenesis	NCI	Public professional education
Mammography	Carcinogenesis	NCI	Public professional education

[a] Risk assessment used as part of a regulation setting process to control existing exposures.
[b] Risk assessment used as a basis for public and professional education programs.

Legislation specifies some of the factors to be considered in risk management. Regulatory agencies, for example, can develop either standards, which are enforceable regulations, or criteria, which are recommended levels of exposure without the force of law or regulation. For the most part, criteria are to be based on health effects only, with no consideration of technical or economic factors. This was the case with the EPA noise analysis, for example.

Standards or enforceable regulations usually involve consideration of both technical and economic factors. However, these considerations may be specifically excluded. The Delaney clause of the Food, Drug and Cosmetic Act, which bans food additives shown to be carcinogenic, provides for no offsetting criteria. Air quality standards issued under Section 109 of the Clean Air Act are by statute to be based solely on scientific and medical criteria. In contrast, under Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), the EPA analysis of ethylene dibromide (EDB) had to consider not only health effects, but the impact of restrictions in pesticide use on agricultural production, food prices and the agricultural economy in general (EPA, 1980).

In many of these cases, feasibility and other issues dominate the risk management decision. For example, OSHA found that technical and cost considerations made it infeasible to set lead and noise standards that their risk assessment indicated would be fully protective of reproductive capacity and hearing, respectively (OSHA, 1978, 1981). The risk management decisions, however, used the assessments to establish a basis for medical surveillance activities that would complement other efforts to limit exposures (OSHA, 1978, 1981).

Although statutes may specify that certain factors be considered in regulatory decisions, they rarely specify the weight to be given the different factors. Economic and political concerns appear to play a substantial role in the weight given. For example, in the case of EDB, notwithstanding high estimates of risk backed by good data, action was delayed by other factors such as availability of safe substitutes, economic impact and industry objectives. By contrast, in the case of Tris, a chemical flame retardant used briefly on children's sleeping apparel, action occurred more quickly because the risk identified was not disputed, other flame retardant products and processes were available, the chemical's major use was not as a flame retardant (and thus the economic impact of the banning was limited) and the publicity associated with exposing children to carcinogens solidified a political consensus for action.

Concern about litigation influences how agencies approach risk assessment. Those interviewed at EPA and OSHA expect their decisions to be reviewed by the courts. This expectation has affected the presentation of risk estimates, weighing of evidence and documentation provided. OSHA, for example, presents risks in the form of rates per thousand because that is the form cited in the Supreme Court's benzene decision.

All the regulatory cases examined involved products already in use or exposure currently taking place. It may be that the economic stakes involved in such decisions are higher than in pre-emptive regulation, such as initial approval of a drug or pesticide. As a result, nonhealth effects may play a larger role in the cases examined in this study than in pre-emptive regulation.

Risk assessments done for reasons other than regulation deal less with nonhealth effects. Advisories based on these assessments have no coercive power. Because these advisories can have economic consequences, factors other than health can still play a role.

HOW REGULATORY AND OTHER GOVERNMENTAL DECISION MAKERS VIEW RISK ASSESSMENT

During this study, there were opportunities to meet with officials involved in some of the risk management decisions supported by the risk assessments examined. There were also opportunities to talk with technical staff and others about how the information developed in the risk assessments was viewed and how it was used in risk management decisions. The intent in examining these matters was not to second guess decisions but to understand the role of risk assessment information and the factors that contribute to or hinder its use. The conclusions presented here, because of the limited sample, can only suggest the range of experience with the use of risk assessment analysis.

These conclusions are discussed in three categories:

- The background of decision making officials reviewing risk assessment results.
- Methodologic issues they confront reviewing risk assessments.
- How decision making officials interpret risk assessment.

1. The Background of Risk Managers

The roles and scientific backgrounds of risk managers varied widely. Some are the administrators and senior staff of independent regulatory agencies. Others are the senior staff of agencies with principally scientific missions. Most of those involved in nonregulatory risk management decisions were scientists. They were more likely to have formal training in the disciplines used in the risk assessment and to be

familiar with the methods and sources of data used and their strengths and limitations.

Some decision makers in the regulatory cases had scientific backgrounds and some did not. Those without significant scientific backgrounds included lawyers, business people and economists. In these cases, there were extensive efforts to explain the risk assessment results to the officials. Most were described as attentive to these efforts, and a hearing record from the CPSC on Tris shows the commissioners actively questioning scientists on the methods and assumptions involved in the risk assessment (CPSC, 1977).

Several of the scientific and technical staff who were interviewed noted that the nonscientist decision maker was not always able to understand the nuances or audit or evaluate the quality of the data or analyses. They also noted that in some cases, a manager's personal experience with specific hazards shaped their attitude toward the credibility and importance of a risk assessment. Some reportedly conferred with their personal physicians regarding risk assessments and the level of risk associated with different hazards. It is unlikely that these clinicians had substantial experience in either risk assessment methods or epidemiologic research. One agency scientist, citing his experience, stated that these physicians didn't trust animal data.

2. Uncertainties in Risk Assessments

Risk assessments incorporate a large number of professional judgments, plausible assumptions and compromises resulting from limited data. These leave these studies vulnerable to challenge and, indeed, they are routinely challenged by parties with substantial interests in the risks being analyzed. Some of the limitations that can create substantial uncertainty as to the findings of a risk assessment are noted below.

Limitations in the Methodologies for Assessing Carcinogenic Risk. From a strictly scientific perspective, human data are the preferred source of information to assess risks to human health. In practice, however, sound studies of human exposures and/or outcomes are rarely available. There are two principal problems in epidemiologic studies of carcinogens. One is small study size, which results in low statistical power to detect effects. In their risk management decisions, agencies are interested in cancer risks that range between one in a thousand and one in a million at actual exposures. Finding effects at these levels requires substantial study groups or human populations with very high exposure to the agent. Thus, although positive epidemiologic findings are considered extremely useful in identifying carcinogenic hazards (and were used as the basis for risk assessments of mammography (Jablon and Kato, 1971; Myrden and Hiltz, 1969; NRC, 1972, 1980, 1984)), negative epidemiologic studies (and to a lesser extent negative animal studies) have limited usefulness. There were positive epidemiologic studies for passive smoking (DHHS, 1986), fat (Greenwald and Ershow, 1985) and radiation in mammography.

A second major problem in the epidemiologic studies has been the difficulty of determining the actual exposure of those being studied. The epidemiologic studies of fat, formaldehyde, EDB, passive smoking and dioxin have all been challenged on this basis. The effect of these problems has been to reduce the role of epidemiologic data in cancer risk assessment of chemical hazards.

The difficulties in using human data were reflected by the Interagency Regulatory Liaison Group (IRLG) when it noted, "A carcinogenic effect can be easily missed by epidemiologic methods, especially when common types of cancer are studied ··· Absence of a positive statistical correlation does not by itself demonstrate absence of a hazard ··· (but) may indicate the upper limits for the rate at which a specific type of exposure could affect the incidence and/or mortality of specific human cancers under the conditions of observation" (IRLG, 1979).

Some persons interviewed for this study went even further than IRLG in questioning epidemiologic studies. They argued that it is almost always possible to "poke holes" in such studies and that interested parties aggressively seek to do so in contentious cases. Because the regulatory agencies have limited resources to rebut these claims, they are even less inclined to rely on the epidemiologic findings.

Long-term animal bioassay have become the major source of information in the hazard identification stage of risk assessment of carcinogens. There is general acceptance by the technical staffs involved in risk assessment of the proposition that for public health and regulatory policy purposes, proven animal carcinogens should be treated as at least potential human carcinogens. This policy is based on an understanding that the mechanisms of carcinogenesis are much the same across mammalian species and that most known human carcinogens have tested positive in animal bioassays (IRLG, 1979; EPA, 1984).

Agencies involved in risk assessment have had to develop a policy for using and accepting animal data. For example, Tris was the first case to come before CPSC where the Commission had to rely exclusively on animal data (Esch, 1978). At that time, staff recommended that careful thought be given to what animal data were to be accepted and how they were to be used. Staff further recommended that the Commission explicitly decide what would constitute adequate

evidence of carcinogenicity for this and future cases. The Commission used the bioassay and in vitro tests as adequate evidence of carcinogenicity.

A plausible assumption is not a proven proposition, however; agencies doing risk assessments have been careful to label as proven human carcinogens only those substances for which there are human data. They have characterized the results of animal studies as suggesting substances are probable carcinogens or possible carcinogens, and each agency has drawn the line differently. The OSHA cancer policy (no longer in effect) required two positive animal studies in order to justify regulating a substance as a human carcinogen. The National Cancer Institute (NCI) analysis of dietary fat and the Office of Smoking and Health (OSH) analysis of passive smoking did not characterize the power of the animal studies alone, but assessed the weight of evidence from both animal and human studies (DHHS, 1986; NCI, 1984). Regardless of where agencies have drawn the line, animal data has routinely been challenged by outside groups as not conclusively demonstrating a cancer risk in human beings.

Relying on animal studies raises many methodologic questions. In bioassays, a small number of animals are exposed to large doses of the suspected carcinogens. If different species show different sensitivity to the agent, a choice must be made as to which species to use. Standard practice of some agencies has been to make the most conservative choice (i.e., the choice that increases the estimated cancer risk) by using data from the most sensitive species but some critics have challenged this assumption. In some cases, where there is a known high rate of spontaneous cancers in some species, e.g., liver cancers in mice, agencies have preferred to use data on tumors at other sites or other species for their analysis. Even when this issue is resolved, other issues remain.

For example, agencies differ in how they handle animals that die or are killed before the end of the study. High levels of mortality in the study population decrease the reliability of the study and increase the uncertainty of the predicted risk.

The possibility that a suspected carcinogen may act at multiple sites, and the problem of interpreting information on carcinogenic potential when available human data does not show site concordance of tumors with those in animal studies, likewise remain unresolved issues in risk assessment. The agencies involved in the cases studied also differed in how they counted tumors in multiple sites.

Once a decision has been made to use animal bioassays and the studies for analysis have been chosen, the data must be extrapolated from animals to humans. Important questions persist in how to equate animal and human dose or exposure. Risk assessments use several methods in extrapolating dose—body weight, surface area, concentration in the air. EPA guidelines call for modifying body weight for surface area, a very conservative choice (EPA, 1984). Critics have said that this is not always appropriate and agencies involved in other cases in this study used a variety of other scaling methods.

A second issue in determining dose extrapolations is whether to use the actual dose rate to which the animal is exposed or to use the cumulative dose received over a lifetime, expressed as an average daily exposure. The latter is the EPA recommended procedure and was used for dioxin and by the Environmental Defense Fund (EDF) in assessing the risk of Tris (Harris, 1977). The use of cumulative dose assumes that a high dose over a short period is equivalent to a low dose over a lifetime. This is assumed to be a conservative approach. This assumption becomes more doubtful as exposures become more intense, but less frequent, especially when there is evidence that an agent has shown dose-rate effects. Under these circumstances, the use of cumulative dose may not be conservative because the effect of high intermittent dose or high doses early in life followed by no further exposure may be greater than dose spread evenly over a lifetime. These kinds of problems were encountered with EDB, where exposures were intermittent but sometimes very high.

The significance of the timing and intensity of dose is quite controversial and not well understood. For example, one EPA scientist explained in an internal memo:

> "NCI bioassays ··· do not yield results which can be used to answer such questions as—what are the estimated risks attributable to only a limited exposure period? What in fact is being gathered in all such routine studies of carcinogenicity is information on the cancer-induction potential following a given series of rates of exposure ··· rather than the total dose administered over a given period. It follows, therefore, that statements on the risk of neoplasia which are made as a consequence of any extrapolation procedure can have reference at most to such rates of exposure and not to quantities of exposure ··· If it is deemed compelling ··· to know what would be the risks attributable to various exposure times, then ··· carcinogenesis studies of a very different kind would be required. These would involve experimental exposure at various rates for a day, a week, a month, etc., and for a number of multiples of such units yet with the observation time remaining ··· close to a lifetime for the experimental animals ··· Experiments such as this would soon reach prohibitively large sizes" (EPA, 1981).

On the other hand, the EDF risk assessment of Tris quotes the National Academy of Sciences (NAS) and other scientists as saying that dose rate does not seem to affect risk (Harris, 1977). This was the conclusion drawn in the evaluation of the radiation studies used in mammography as well, although age at exposure did influence risk.

Thus the standard bioassays used for assessing the risk of carcinogenesis have no way of dealing with differences in the timing of exposure (dose schedule). OSHA and, to a lesser extent, EPA expressed concern about how to deal with this problem in terms of occupational exposure to EDB which is generally intermittent. This is a critical issue in risk assessment because human exposure is rarely constant over a lifetime, and decision makers as well as risk assessors try to make adjustments for these differences whether they can be justified with the data or not.

Different agencies also have different policies regarding the choice of a mathematical model for describing the dose-response curve. The model preferred by EPA has changed over time as knowledge about cancer (and about models) has accumulated. Other agencies choose models on the basis of their biologic plausibility and how well they fit the data. Some scientists and statisticians interviewed, however, question the practice of picking a model that best fits the data, noting that biologic plausibility should be the determining factor. This view ignores the uncertainty about the biologic plausibility of any of the competing models. Often a report of risk assessment will include the fitted points for several models in order to see whether the estimates are close to one another, or to give a sense of the uncertainty of the estimate.

Certain agreements have been reached with regard to cancer dose-response models. There is little argument in the scientific community that models should be linear at low doses and should not have a threshold. There is some argument, however, about how to model repair mechanisms that may operate at low doses but may be overwhelmed at high doses so that dose-response is quite different at high and low doses. This effect is sometimes described as an irritation effect. There is some debate on the relationship of irritation to carcinogenicity. In discussion concerns raised by industry in the EDB risk assessment, EPA dismissed the argument as having been put to rest (EPA, 1980). Brown (1981), however, believed that irritation was at least in part responsible for the difference in rates of contact site and remote site tumors caused by EDB.

In most risk assessments reviewed, pharmacokinetic and metabolic data were either not mentioned or used only in a peripheral way—to support the conclusion that an agent was carcinogenic or mutagenic, or to provide a biologic explanation of an effect observed in epidemiologic or whole animal studies. Although synergies among carcinogens are known to exist, no efforts were made to identify these or take them into account.

There is also disagreement among risk assessors regarding the appropriate statistic to present as the estimate of risk. Estimates may be presented in various ways, such as a maximum likelihood estimate (MLE), an upper or lower confidence bound or a range. Different agencies prefer different statistics. OSHA prefers the MLE because, say agency staff, they feel constrained to come up with a "realistic number rather than worst case estimates of risk." Others, commenting on the OSHA position, have argued that the "worst case" is not really the worst, and that is not even unrealistic, just less likely than the MLE. Part of OSHA's confidence in the MLE rests on the fact that the level of occupational exposures it is examining in its risk estimates do not extend far beyond the range of the experimental data. Thus extrapolation from the maximum likelihood curve seems highly appropriate (OSHA, 1985).

Agencies that regulate for the general public, which includes infants, children, pregnant women and the elderly and other potentially vulnerable populations, seem to prefer the upper confidence limit. For example, EPA consistently uses the 95% upper confidence limit although, according to interviews, this practice is controversial both within and without the agency. One EPA scientist commented, "· · · although actual risk is usually lower than the estimate of the upper limit, it is not necessarily so, particularly when the estimation process is flawed."

These issues—weakness of the human data base, selection of species, high-to-low dose extrapolation, animal-to-human extrapolation, intermittent exposures and synergies and others—keep re-emerging in the process of reviewing carcinogens. They have not been resolved because current knowledge of carcinogenesis does not permit them to be resolved at this time.

Limitation in the Methodologies for Assessing Noncarcinogenic Risk. The range of effects examined in risk assessments for hazards other than cancer are quite wide: injury or trauma to specific body systems, such as the central nervous or auditory systems; infection; and reproductive problems, including failure to conceive, miscarriage and birth defects. Some effects are temporary or reversible; others are permanent. They may vary in severity from life-threatening or disabling to mildly inconvenient.

The breadth of effects and variety of agents that can be examined make it difficult to generalize about risk assessment in this area. Some risks are easily observed and the relationship to dose level can be quickly established. For others, these relationships are

harder to evaluate and the cases studied suggest some of the obstacles to and issues in conducting effective risk assessments of noncarcinogenic risks.

More of these risks were identified using human data than was the case for carcinogens. It appears that the existence of adequate human data for hazard identification depends on one or more of the following: exposure of large numbers of people (cotton dust, noise, lead, formaldehyde), long history of exposure (noise, cotton dust, lead) and easily observable effects that appear soon after exposure (formaldehyde). Hazard identification is also enhanced by the inclusion of certain end points in standard animal bioassay protocols and in legislation (e.g., reproductive end points in FIFRA). It is much more difficult to identify hazards not included in standard protocols—such as central nervous system (CNS) damage.

Even when hazards have been identified, characterizing the risk in a manner that is useful in risk management can be made difficult by several methodologic problems.

For many of these risks, it is assumed that a threshold level exists below which effects do not occur. Efforts are made to establish a no observed or lowest observed effects level (NOEL or LOEL). These attempts meet with different degrees of success in the different cases.

Despite the extensive human data for lead, for example, difficulties exist in establishing NOELs or LOELs. Improved testing and measurement methods demonstrated effects at lower and lower levels. The ability to measure subtle CNS deficits kept improving leading some observers to conclude that human variability and continuing improvements in measurement and observation would move the threshold practicality to zero. However, as effects were noted at lower levels, some controversy emerged in deciding whether an observed effect was adverse or serious enough to merit corrective action. For example, while clinical anemia due to lead is usually observed only when blood lead levels exceed 100 μg/d, and these were considered adverse without reference to their severity (OSHA, 1978). In the case of noise, despite the fact that the levels of hearing loss at different noise levels were well established, there were substantial efforts to define the level of hearing that should be protected. The issue also emerged in the cotton dust case, where the comparative lack of severity of low-grade byssinosis and the uncertainty regarding the reversibility of the effects increased the difficulty of defining an appropriate level of protection.

The issue of setting standards where there are observed effects may involve defining the difference between material harm and lesser levels of harm or between changes that are viewed as morbid versus those that are merely abnormal or between those that are temporary versus those that are permanent. There is considerable debate over when change becomes an adverse health effect, with some willing to accept exposures that produce abnormalities not linked to clear morbidity and some not. One view is that abnormal physiologic test values represent a decreased reserve capacity for dealing with the next insult; while in another, the meaning of the changes is unclear and may either be totally reversible or have no impact on present or future disease states or subjective evaluations of well being (OSHA, 1978). More simply, one approach looks for proof of safety while the other looks for proof of harm. This issue is likely to become more important as the ability to identify smaller and smaller changes in function or structure improves.

Formaldehyde also illustrates this problem, because the risk assessments did not consider sensitization effects, nor did they factor in either worker or consumer problems with frequent headaches. It appears to be difficult to determine the relative weight that should be given to (and the degree of protection offered from) milder and reversible effects such as headaches or temporary hearing loss compared with effects such as cancer, permanent hearing loss or CNS damage.

One of the most striking features of these cases is the way that data on the effects and potency of lead, noise and cotton dust have been developed over time. Although these cases stand as examples of the ability to generate human data for risk assessment, they also underscore the high costs in sickness, time and other resources required for this data to become available.

Animal studies can generally provide information faster and at less cost than human studies. This seems especially important with respect to agents that are new or have affected relative few individuals. Animal studies of toxic and reproductive effects present several problems, however. One is simply whether there is an applicable animal model. In the case of cotton dust, for example, there was not. If there is an animal model, problems emerge in scaling dose and in determining whether humans are more or less sensitive than the test animals, and in characterizing variations in sensitivity within the human population. Historically, these problems have been dealt with by using LOEL/NOEL in the most sensitive species, assuming human sensitivity at least as great as in the animals, and applying higher safety factors when there is greater uncertainty about human sensitivity. However, where there is great variation among animal species, there is growing resistance to this methodology. The debate over dioxin is a good example of this.

There is considerable debate over the selection of safety factors. In principle, the choice of safety factor is a matter of risk management rather than risk assessment, but it should be based on such risk assessment concepts as the level of understanding of

the biologic processes at work, the quality of the data used to estimate the LOEL or NOEL and the severity of the effect associated with the agent. The NAS, for example, recommends the following safety factors:

- A factor of 10, if adequate human exposure data are available and supported by toxicity in other species.
- A factor of 100, if good toxicity data are available in some animal species.
- A factor of 1000, if toxicity data are limited or incomplete.

This scale has also been described as one order of magnitude for within-human variation; a second for the extrapolation from species to species; and a third for additional uncertainty.

The actual practices of risk assessors show substantial variation in the relation of safety factors. Both OSHA and New York State used safety factors explicitly in their analysis of the reproductive effects of EDB. Brown (1981) chose a safety factor of 10; New York, 1000 (New York State Dept. of Health, 1984). This may reflect differences in the perceived quality of the animal data and the confidence with which doses in animals could be extrapolated to people, or it may reflect differences in the populations being protected. In the regulation of occupational exposures, workers are rarely afforded the same degree of protection as is the general public in environmental regulation.

Presenting Methodologic Problems and Issues in Risk Assessment. Risk assessments often do not assist a decision maker, especially a nonscientist, in coming to terms with methodologic uncertainties in the risk assessments. The assessments are often weak in portraying the relative importance of various data, especially data not included in mathematical calculations. In the analysis of a carcinogen, for example, the principal element of the risk assessment is the dose-response modeling. A substantial amount of other information—results of in vitro tests, pharmacokinetic studies—may be examined. There may be a description of various factors that might increase or decrease the risk from the baseline estimated. Sometimes these are just listed; at other times they are presented with the direction of the effect identified but no indication of the likely order of magnitude of the effect. There is limited discussion of the effect of all errors together, or of the extent to which the different factors are likely to balance out. This may be because current methods do not allow these factors to be scaled in any way, but the effect is to leave the risk manager with a substantial amount of raw information without an adequate guide to its implications.

Another factor determining the presentation of information is that published risk assessments are meant to provide justification for risk management decisions as well as input into them. Thus, the agency is building a record for the courts, including all possible justifications of the actions taken and not restricting the record by defining in detail how the information is balanced and weighed.

Few risk assessments had graphs and other aids to the assimilation of the information. In one risk assessment, for example, there was a discussion of the dose-response curves generated by several different models. The text discussed how different models produced higher (more conservative) estimates of risk in different ranges of exposure. There was discussion of where the curves crossed. There was no graph of the curves, however. Risk managers may well want to examine detailed tabular displays of numbers, but these could be complemented with graphic and other displays that present the relationship of different numbers to one another. The tables, per se, seem adequate but their use and importance is not always explained clearly.

Many of these cases also failed to place risks or exposure into an "understandable extent" through intensity spectra, comparative models or examples. Some of the cases in which this was done illustrate how effective this could be. In the Tris case, exposure estimates were presented based on the number of pairs of pajamas a child would wear per year (Harris, 1977). In one EDB assessment, the following statement appeared: "Fumigated citrus stored at 40 degrees F. for 6 days contains EDB residues in excess of 1 ppm. For a middle sized California orange weighing approximately 200 grams, this would represent a daily intake of approximately 0.2 mg of EDB. By reference to the table [of the probability of cancer by dose as calculated at the upper confidence limit from the animal studies] this level ··· surpasses each of the eight estimates given for the "safe" daily doses corresponding to an upper limit on the risk of 1/1000." The assessment goes on to point out that if the fruit is held for a shorter period of time, the residues will be larger and the dose will be 50 to 70 times the "safe" dose (EPA, 1981).

The writing in risk assessments is of very uneven quality. Official government documents such as Rebuttable Presumption Against Registration (RPAR) position documents written to comply with FIFRA and *Federal Register* notices seemed to be the most clearly written, especially for the nonscientist. However, the design of these documents often makes them extremely repetitive and not particularly well organized. These documents were usually explicit in defining areas of controversy. Assessments written for scientific audiences (such as journal articles) contained more technical data with much less analysis of its policy implications.

The written record is only one of the vehicles for communicating risk assessments to decision makers. At many of the agencies studied there is a briefing on the findings. Interviewees did not discuss briefings in detail, but did provide some insight into the process. In addition, at least one briefing for CPSC on Tris was on the record and included in the Tris documentation. Briefings typically include a discussion of model used, limits, caveats and the range of estimates. If one model was preferred by the analytic staff, the reasons for this preference would be discussed. The briefing provides decision makers with an opportunity to ask questions. In the CPSC briefing, for example, the discussion of models led to the drawing of several charts at the meeting. There was substantial give and take about the quality of the evidence that Tris was in fact absorbed through the skin. This discussion led to subsequent radiolabeling experiments (CPSC, 1977).

What emerges from the Tris record are very active efforts by the commissioners, including the nonscientists, to understand the implications of the information. Discussions with staff at several agencies also suggested that they know from experience what questions are likely to be asked and that they prepare and present their material accordingly. Although these processes may involve some inefficiencies, they provide opportunities for decision makers to get the information they wish to have in a form that they can understand. Present processes may be best suited, however, for decision makers who work with relatively few hazards and programs. As the numbers expand, and as risk assessments are produced for a broad range of risks by a large number of offices, there is likely to be a greater need to standardize presentations for senior officials.

3. The Interpretation of Risk Assessments by Decision Makers

How do risk managers respond to the information they receive as part of the risk management process? Substantial differences were reported in the extent to which risk managers had confidence in the results of the risk assessment. Some accepted the basic operating assumptions of the risk assessment process. Others were completely cynical, with one stating about carcinogenesis, "· · · you can make the numbers come out any way you want." The attitudes of many fell between these two extremes and were quite complex.

Decision makers reviewing risk assessments will look for information that gives them confidence in the findings. A narrow range of estimates of risk may be one factor that provides such confidence. Some agency technical staff, for example, felt they were being pushed to produce a single number as the risk estimate with as few caveats and as little explanation as possible. Consistency in the evidence is another factor that risk managers, especially nonscientists, looked for. Thus, while there seemed to be general acceptance of the use of animal studies as the basis for judging carcinogenesis and other risks, some risk managers were troubled by negative findings in human epidemiologic studies, despite the acknowledged weakness of these studies to confirm positive animal studies. Given the methodologic limitations noted above, many risk assessments fall short of providing consistent, narrow estimates of risk.

To a large extent, the results of risk assessment are being interpreted qualitatively. The degree of hazard may be interpreted as high, medium or low, and in some cases risk managers will ask how the levels of risk compare with other things that have been regulated. Some risk managers appear to view quantitative risk assessment results as index numbers, not necessarily as a measure of the actual risk of exposed population. This would parallel a debate observed among the technical staffs about how much risk estimates from carcinogenesis analyses reflect risk in the real world. That is, does a projected risk of 7 in 1000 really mean that 7 of every 1000 people exposed at a particular level will get cancer or just that more individuals will probably get cancer than if the risk were 0.7 and fewer than if the risk were 70.

It is not clear to what extent decision makers, especially nonscientists, understand the specific methodologic issues that increase uncertainty in risk estimates. Peer review may reduce their perceived need to understand these issues. Managers are aware that most risk assessments have been subject to peer review or some other scientific review process. EPA, for example, regularly relies upon its Scientific Advisory Board, CPSC has used outside reviewers, and the PHS used scientific review procedures in developing its assessments of dietary fat, passive smoking and mammography. When the risk managers are scientists, they may participate in the reviews, as they regularly do in the PHS. When they are not scientists, they at least know that such review has taken place and that the assessments meet some minimum standards of scientific acceptability.

Policymakers appear to be sensitive to certain unresolved methodologic issues—such as less than lifetime exposures or the possibility of synergies—and may make informal and personal adjustments of risk estimates to account for these variables. Parts of the discussion in the CPSC hearing record on Tris and the OSHA record on EDB gives some hints about this process.

Policymakers come into office with different attitudes about how risk averse government policy should be. Personal experiences may have shaped their views

of what is dangerous and what is not, as well as what constitutes impairment or an adverse health effect. These attitudes will influence the weight given risk assessments, especially as the range of uncertainty in the assessments increases. Those who believe that policy should be especially risk averse may focus on synergies as a basis for acting on the lowest estimates. Those who are less risk averse may focus on less than lifetime exposures as a justification for lack of action in the face of high estimates of risk. Risk assessment will play a larger role when and if confidence in the conclusions and certainty of the estimates increases. This in turn is dependent on continued improvement in the methods.

PUBLIC PARTICIPATION IN THE RISK ASSESSMENT PROCESS

One phenomenon that was observed in each of these cases was substantial participation by nongovernment groups and individuals. The range of groups involved included corporations and industries with economic stakes in the agent being examined; organizations representing the exposed populations, such as unions and the Brown Lung Association; public interest groups involved with the subject, such as the EDF; professional societies; public officials, such as the health commissioners in Missouri and New York; individual citizens, such as those who complained about urea formaldehyde foam insulation; and the press.

While some of the participation we observed may be a result of choosing "big" cases, some can be anticipated in any case. Corporations affected by risk assessments (because their products, supplies, effluents, etc., are being reviewed) will inevitably participate in them. Public and other interest groups can be expected to track both research and agency activities with regard to potential workplace and community hazards. Citizens are likely to continue reporting and seeking action in cases of acute health problems. Of course, different risk assessments will attract different degrees of participation, and their roles will vary just as they did in the cases we studied. In some cases, such as formaldehyde, cotton dust and dioxin in Missouri, nongovernmental participation placed the item on agency agendas. Outside groups were also active in reviewing the analyses prepared by the agencies, disseminating them to the public and pushing for risk management decisions.

Opportunities for participation by outside groups are sometimes mandated by law. This is typically the case in regulation, where, once the regulatory processes have begun, both the timing and form of public participation may be specified. In some of the nonregulatory cases, private groups stayed involved with the agencies in their work. The creation of working groups or other study panels sometimes provided formal vehicles for involving outside parties.

As evidence in these cases, these outside groups have the resources to conduct risk analysis parallel to that done by government agencies. They have access to the same body of scientific literature, bioassays and computer models as the government. They have access to commercial or government laboratories that can measure the level of exposure to many of the agents of concern.

Public participation has several consequences. One is that the government does not have a monopoly on any phase of risk assessment and cannot control the risk assessment agenda. Risk assessments can be initiated through the actions of private groups, and relevant data may come from private groups. In the Tris case, for example, the Environmental Defense Fund brought in data and identified scientists doing relevant work. Agencies conducting risk assessments are likely to be forced to defend all their methodologic choices. There is probably no more stringent peer review conducted than the review of risk analyses conducted by outside parties in regulatory decisions. Even in advisory decisions, such as those on passive smoking and dietary fat, there is likely to be detailed outside review of the conclusions by affected parties.

In the past, some government agencies have hesitated to make statements on some issues because they were concerned that their statements and actions be well grounded in scientific knowledge and not be unnecessarily and inappropriately alarming to the public.

Outside groups, however, are also actively involved in disseminating risk analyses, and not just those emerging from the public risk management processes. In the area of passive smoking, for example, both the cigarette companies and antismoking groups have distributed information. Some of the antismoking groups have chosen to publicize a quantitative risk assessment by Repace and Lowery which, because of its technical limitations, the Office of Smoking and Health has not referenced in its report on the subject. Discussions of dietary fat and cancer, sparked by some of the research literature relied upon by NCI, were in the popular press long before NAS and NCI issued their guidelines. Reports on television news about EDB residues in grain products, coupled with the identification of EDB (based on the NCI bioassay) as a potent carcinogen, helped precipitate actions by state governments and large grain millers that in turn encouraged action by EPA. Such publicity reduces the ability of government agencies to not speak about suspected hazards and can create strong pressures to comment upon or react to these outside risk assessments.

The ability of outside groups to conduct and

disseminate risk assessment studies has encouraged public action to clear the air and provide an authoritative judgment of the risk. Outside involvement in risk assessment has also placed pressure on government analysis to do a better job and has brought additional resources to the tasks of identifying, analyzing and interpreting relevant information.

ACKNOWLEDGMENT

This paper was prepared based upon research funded by the United States Department of Health and Human Services under Contract 282-83-0069.

REFERENCES

BROWN, D. (1981). Quantitative cancer risk assessment for occupational exposure to ethylene dibromide. Unpublished data.

CONSUMER PRODUCT SAFETY COMMISSION (1977). Transcript of meeting with Environmental Defense Fund, Mar. 28.

DEPARTMENT OF HEALTH AND HUMAN SERVICES (1986). The health consequences of involuntary smoking: A report of the surgeon general.

ENVIRONMENTAL PROTECTION AGENCY (1980). Ethylene dibromide: Position document 2/3. NTIS PB81-157851.

ENVIRONMENTAL PROTECTION AGENCY (1981). Internal memo, Sept. 5.

ENVIRONMENTAL PROTECTION AGENCY (1984). Proposed guidelines for carcinogen risk assessment. *Federal Register* **49** 46294.

ESCH, A. F. (1978). Extrapolation of risk from animals to humans. Internal CPSC memorandum.

EXPOSURE TO LEAD: FINAL STANDARD (1978). *Federal Register* **43** Nov. 14, 1978.

GREENWALD, P., ERSHOW, A. G. and NOVELLI, W. D., EDS. (1985). *Cancer, Diet and Nutrition: A Comprehensive Sourcebook.* Marquis Who's Who, Chicago.

HARRIS, R. H. (1977). Estimating the cancer hazard to children from Tris-treated sleepwear. A memorandum to the Consumer Product Safety Commission in support of EDF's petition to ban the sale of Tris-treated wearing apparel. Mar. 8.

INTERAGENCY REGULATORY LIAISON GROUP (1979). Scientific bases for identification of potential carcinogens and estimation of risk. *Federal Register* **44** 39858.

JABLON, S. and KATO, H. (1971). Mortality among A-bomb survivors, 1950–1970. *J. Nat. Inst. Health-ABCC LifeSpan Study* Report 6, Technical Report.

MYRDEN, J. A. and HILTZ, J. E. (1969). Breast cancer following multiple fluoroscopies during artificial pneumothorax treatment of pulmonary tuberculosis. *Can. Med. Assoc. J.* **100** 1032–1034.

NATIONAL CANCER INSTITUTE (1984). Diet, nutrition, and cancer prevention: A guide to food choices. Washington.

NATIONAL RESEARCH COUNCIL (1972). The effects on populations of exposure to low levels of ionizing radiation. Report of the Advisory Committee on the Biological Effects of Ionizing Radiations (BEIR I). National Academy Press, Washington.

NATIONAL RESEARCH COUNCIL (1980). The effects on populations of exposure to low levels of ionizing radiation. Report of the Advisory Committee on the Biological Effects of Ionizing Radiations (BEIR III). National Academy Press, Washington.

NATIONAL RESEARCH COUNCIL (1984). Assigned share for radiation as a cause of cancer: Review of radioepidemiologic tables assigning probabilities of causation. Final Report of the Oversight Committee on Radioepidemiologic Tables. National Academy Press, Washington.

NEW YORK STATE DEPARTMENT OF HEALTH (1984). A risk assessment for ethylene dibromide. Feb. 21.

OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION (1981). Occupational noise exposure: Hearing conservation amendment: Final rule. *Federal Register* **46** 9738–9785.

OCCUPATIONAL SAFETY AND HEALTH ADMINISTRATION (1985). Occupational exposure to formaldehyde. Proposed rule making. *Federal Register* **50** Dec. 10, 1985.

REPOSE, J. L. and LOWREY, A. H. (1985). A quantitative estimate of nonsmokers' lung cancer risk from passive smoking. *Environ. Internat.* **11** 3–22.