REGULATORY TOXICOLOGY AND PHARMACOLOGY 20, 15–36 (1994)

Use of Probabilistic Expert Judgment in Uncertainty Analysis of Carcinogenic Potency

JOHN S. EVANS,*[,1] GEORGE M. GRAY,† ROBERT L. SIELKEN, JR.,‡ ANDREW E. SMITH,§ CIRIACO VALDEZ-FLORES,‡ AND JOHN D. GRAHAM†

Interdisciplinary Programs in Health and §Department of Environmental Health, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115; †Center for Risk Analysis, Harvard School of Public Health, 718 Huntington Avenue, Boston, Massachusetts 02115; and ‡Sielken Incorporated, 3833 Texas Avenue, Suite 230, Bryan, Texas 77802

Received November 29, 1993

A new approach to characterizing the state of knowledge about carcinogenic potency is described. In this approach, the carcinogenic risk posed by a specific dose is characterized by a probability distribution, indicating the relative likelihood of different risk estimates. The approach utilizes expert judgment and a probability tree and is illustrated in a case study of chloroform exposure. Experts in cancer biology/toxicology, pharmacokinetics, and dose–response modeling were identified by a panel of science-policy specialists. In a workshop, experts reviewed the chloroform data, received training in probability elicitation, and constructed a consensual probability tree based on biological theories of cancer causation. Distributions of carcinogenic risk were developed based on the probability tree, chloroform data, judgmental probabilities provided by the experts, and classical statistical techniques. Risk distributions varied considerably between experts, with some predicting essentially no risk from 100 ppb chloroform in drinking water while others have at least some probability on risks generally considered of regulatory significance. Estimated human risk was much lower when extrapolating from liver tumors in animals than from kidney tumors. Issues of scientific disagreement leading to different risk distributions between experts are discussed. The resulting risk distributions are compared to standard EPA risk calculations for the same exposure scenario as well as to the expert judgments of epidemiologists about cancer risks of chlorinated drinking water. Issues in combining expert judgments are discussed, and several alternative methods are presented. Strengths and weaknesses of the distributional approach are discussed.

INTRODUCTION

The carcinogenic potency of chemicals at low doses cannot be directly measured in humans or animals due to the limited statistical power of epidemiology and laboratory animal bioassays. Yet policymakers and the public need estimates of low-dose

[1] To whom correspondence should be addressed.

0273-2300/94 $6.00

cancer risks to establish regulatory priorities and to make risk management decisions. For example, both the U.S. and California Environmental Protection Agencies (EPAs) make widespread use of quantitative cancer risk estimates when making public health decisions. The standard methods, which are applied uniformly to all potentially carcinogenic chemicals, are intended to generate conservative, yet plausible upper-bound estimates of risk (U.S. EPA, 1986).

With many compounds the risk assessor is confronted with multiple sources of data from animal bioassays, pharmacokinetic investigations, mechanistic studies, and epidemiological analyses. Even within one of these areas (e.g., the animal bioassay), the risk assessor may encounter a multiplicity of inference options, such as use of tumor data from one site or several different sites. Furthermore, the risk assessor may need to consider data from several bioassays in which different test strains or species may have been used and/or the test compound may have been administered via different routes (i.e., inhalation, gavage, or drinking water).

Standard approaches to carcinogen risk assessment fail to adequately deal with the richness of such data and theory. They tend to focus on one theory of carcinogenesis, one measure of dose, one experimental data set, and so forth. While this simple approach may be acceptable in cases where the societal stakes in risk management are not large (i.e., regulation of compounds which neither produce large risks to public health nor entail large social costs to control), failure to incorporate the full body of scientific information on a compound has less justification when the stakes are large. Moreover, by ignoring the efforts of scientists to generate better data, the field of risk assessment may alienate the scientific community on which it depends for information and legitimacy.

How should standard risk assessment procedures be modified to allow use of more scientific information? A major controversy concerns how to determine when mechanistic data on a specific chemical is sufficiently compelling to justify departure from standard default approaches for computing low-dose cancer risk. Several committees of the National Academy of Sciences have struggled with this difficult issue and have been unable to reach a clear consensus on how to proceed (e.g., NRC, 1993, 1994). Both the U.S. EPA and the California EPA continue to struggle with this issue and have recently requested public comment on various criteria to be met for use of mechanistic data in proposed revisions to their carcinogen risk assessment guidelines (Society for Risk Analysis, 1993; U.S. EPA, 1992). However, these proposed approaches consider only "either/or" scenarios (i.e., use or do not use new science in lieu of default assumptions). It would be desirable to have methods to inform risk management with the full weight of the scientific evidence.

In this article we demonstrate a new risk assessment methodology, rooted in Bayesian concepts of subjective probability, that can be used to incorporate the full body of scientific evidence about a chemical, acknowledging uncertainty in both new mechanistic data and the default assumptions of standard risk assessment methods. This Bayesian method characterizes low-dose cancer risk as a probability distribution reflecting the judgments of qualified scientists about the weight of the evidence. The approach allows different, often mutually exclusive, methods of estimating low-dose cancer risk to be expressed, with the judgments of scientists determining the likelihood of alternative scenarios. If risk managers ultimately desire a single risk number for

regulatory purposes, a summary statistic can be derived from the reported probability distributions.

THE DISTRIBUTIONAL APPROACH

This case study builds on previous decision-analytic applications of subjective probability in engineering, medicine, and environmental health (Morgan, 1984; Weinstein and Fineberg, 1976; Finkel and Evans, 1987; Hammitt and Cave, 1991; Bonano *et al.*, 1990). One method of deriving a distribution of cancer potency is to directly elicit the different values and their subjective probabilities from experts. However, the direct-elicitation approach may fail to explicitly document the rationale behind alternative potency estimates and may not aid in identification of areas of controversy and agreement. Because the evaluation of carcinogenic potency is so complex, we instead developed a probability tree as a method to disaggregate the problem, allowing expert judgment on individual aspects of the problem. In a previous article, we illustrated how a probability tree and judgmental probabilities can be combined with standard statistical methods to perform a distributional analysis of cancer potency (Evans *et al.*, 1994). The result is a distribution of risk estimates weighted by their likelihood of being correct, as judged by scientists. Our first assessment, which used formaldehyde as a case study, was hypothetical in the sense that we used our own subjective weights in the analysis rather than weights provided by expert scientists (Evans *et al.*, 1994). This article extends this procedure by incorporating formal expert judgment of the components of the probability tree into the process.

Chloroform as the Illustrative Compound

Chloroform was chosen to illustrate the distributional method because it has a rich toxicological database and it is a compound with widespread human exposure. Along with industrial use, in the past it has had widespread public exposure from use as an anesthetic and sweetener for toothpaste. Most large water supplies in the United States utilize chlorine to disinfect drinking water supplies, and chloroform is formed during the process.

Evidence of the carcinogenic potential of chloroform in animals was first noted in 1945 (Eshenbrenner and Miller, 1945). Animal carcinogenicity has since been confirmed many times (NCI, 1976; Roe *et al.*, 1979; Palmer *et al.*, 1979; Heywood *et al.*, 1979; Jorgenson *et al.*, 1985). However, the results of these studies have not been consistent. For instance, bioassays conducted by the National Cancer Institute (NCI, 1976) observed a dramatic increase in liver tumor incidence in mice gavaged with chloroform in corn oil, while rats exhibited an increase in kidney tumors. In contrast, liver tumors were not seen in mice of the same strain in the Jorgenson *et al.* (1985) bioassay, in which chloroform was administered via drinking water. The Jorgenson *et al.* (1985) bioassay did confirm the NCI (1976) findings of an increase in kidney tumors in male rats. The fact that organ toxicity seems to accompany tumor formation in animals was first noted by Eshenbrenner and Miller and, along with a preponderance of negative assays for mutagenic potential, forms the basis for the hypothesis that high-dose toxicity and cell killing are necessary for chloroform carcinogenicity. Others

counter that many of the genotoxicity results are inconclusive because of inadequacies in the experimental protocols, and that no definitive conclusion on the mutagenic potential of chloroform can be reached (Rosenthal, 1987). Furthermore, they note that chloroform is metabolized to phosgene (a highly reactive molecule once used as a war gas) by the cytochrome P450 system. This reactive molecule might be expected to react with and alter DNA (Rosenthal, 1987).

Thus, some scientists believe that chloroform causes cancer in rodents at high doses through modes of action that may not be relevant to low-dose human risks, while others argue that the data are not yet convincing enough to abandon standard models of risk assessment. The large amount of scientific data and well-developed competing hypotheses of carcinogenic mechanism make chloroform an intriguing demonstration compound. Each of the hypotheses of tumor causation gives rise to notions of appropriate measures of dose for dose–response estimation and the proper form of the dose–response relationship. While our results on chloroform are not intended to be definitive, the case study provides sufficient technical detail to illustrate the mechanics, strengths, and limitations of the distributional method of carcinogen risk assessment.

Identifying and Selecting Expert Scientists

A crucial step in any risk assessment process is the recruitment of qualified scientists, since they are needed to assist in the selection and interpretation of data and in the process of mathematical modeling (Otway and von Winterfeldt, 1992; Cooke, 1993). Expert selection is particularly integral to the methodology proposed here because scientists are asked to evaluate the chloroform database, construct a probability tree in collaboration with the authors, and provide judgmental probabilities for critical branches in the probability tree.

We have previously discussed the desired properties of an expert selection process: recruitment of appropriate expertise, explicitness, reproducibility, ease of execution, absence of complete control by the risk assessor or manager, and balance of perspectives (Graham *et al.*, 1988). In contrast, informal selection of scientists by a risk assessor is readily executed but does not necessarily assure appropriate expertise or balance of perspective.

Since it is not known which expert selection procedure best satisfies all of these criteria, we are evaluating various selection procedures. In previous projects we chose experts based on counts of relevant publications in peer-reviewed scientific journals (Wolff *et al.*, 1990), peer nomination counts (Hawkins and Evans, 1989; Hawkins and Graham, 1988; Spedden, 1992), and membership on a specific committee of the National Academy of Sciences (Siegel *et al.*, 1990).

In this project, we tried a different selection procedure where a nomination panel was empowered to select experts. In particular, we designated a panel of five science-policy experts who we believed would be aware of the needs of the risk assessment process and knowledgeable of the relevant scientific communities. This panel of nominators included Thomas Burke (Johns Hopkins University School of Public Health), William Farland (U.S. Environmental Protection Agency), Carol Henry (Risk Science Institute), Robert Scheuplein (U.S. Food and Drug Administration), and Lauren Zeise (California Environmental Protection Agency).

The nominators concurred that our project required expertise in four key areas: cancer biology/toxicology, pharmacokinetics, epidemiology, and dose-response modeling. While prior knowledge of the chloroform database was considered preferable, it was not treated as a necessary condition for participation. The panel did not consider expertise in exposure assessment to be critical since only dose-response issues were to be addressed in the project. To assist each nominator in identifying and selecting experts, we supplied lists of names of scientists who had published relevant scientific papers on chloroform, chemical carcinogenesis, and related topics.

Based on discussions with the nomination panel, a group of scientists with expertise in the relevant disciplines was identified. Eleven scientists distributed among the four areas of expertise were invited to take part in the project. Two of the eleven were unable to participate, leaving us with a final group of nine scientists, who are identified in Table 1. The group possesses considerable diversity in scientific discipline and perspective on risk assessment issues.

Because the probability tree approach does not require scientists to choose a single "best guess" for each level of the tree, different scientists may put different weights on the alternatives. The approach allows the many areas of disagreement and consensus to be identified and explored. Thus, differences of opinion are encouraged rather than discouraged. The specific opinions of individual experts were left anonymous to facilitate candid discussion of the scientific issues. Throughout the paper, the experts are referred to with the letters A through F or as epidemiologist A, B, or C. This policy of anonymity is commonly used in expert judgment research (Morgan and Henrion, 1990).

The Chloroform Workshop

The Harvard Center for Risk Analysis convened the group of nine experts at a workshop that was held April 13–14, 1992 at the Belmont House in Elkridge, Maryland.

TABLE 1

PARTICIPATING EXPERTS

	Affiliation
Epidemiology	
Kenneth P. Cantor	National Cancer Institute
Philip Cole	University of Alabama, Birmingham, School of Public Health
Jack Siemiatycki	Institute Armand Frappier
Cancer biology/toxicology	
Michael A. Pereira	Environmental Health Research and Testing, Inc.
Byron E. Butterworth	Chemical Industry Institute of Toxicology
Pharmacokinetics	
Richard H. Reitz	The Dow Chemical Co.
Lorenz Rhomberg	U.S. Environmental Protection Agency
Dose-response modeling	
Dale Hattis	Clark University
Thomas B. Starr	Environ Corp.

The objectives of the workshop were to (1) review and evaluate the data on the carcinogenic potential of chloroform, (2) design a common probability tree that each expert would employ when assessing the carcinogenic potency of chloroform, and (3) review and familiarize the experts with the literature on the uses of subjective probabilities and with the difficulties in eliciting these judgments.

Prior to the workshop, each expert received a notebook including selected key original studies on chloroform, plus excerpts on chloroform from EPA's IRIS Database, EPA's Health Assessment Document on chloroform, the relevant IARC Monograph, and selected review articles. At the workshop, each expert made a presentation on the database in his domain of expertise. Some new data were presented, particularly in the area of chloroform epidemiology and the role of cell proliferation in chloroform carcinogenesis. The first day of the workshop, therefore, allowed the experts to thoroughly discuss the chloroform database.

The second day of the workshop was devoted to construction of a probability tree that was satisfactory to all of the experts. The tree provides a structured method for analyzing the individual components in risk estimation (Evans *et al.*, 1994). The three epidemiologists in the group decided that it would be useful if they performed a separate, direct assessment of relative risk from human studies that could later be compared to the distributional results based on a rodent-based probability tree. Hence, the probability tree was designed specifically to assist in using animal data to predict human risk.

Many biological scientists are unaware of the use of structured expert judgment in technical problems such as siting a high-level nuclear waste repository (Bonano *et al.*, 1990), nuclear power plant safety (NRC, 1975), and the health effects of childhood lead exposure (Whitfield and Wallsten, 1989) and sulfur air pollution (Morgan *et al.*, 1984). This history was discussed to assure our scientists, who were a little skeptical, that this is a serious analytic exercise (Keeney and von Winterfeldt, 1989). The experts were then familiarized with the techniques used to elicit probabilistic expert judgments. Behavioral decision analysts have demonstrated that many scientists and laypeople have difficulty comprehending and making probabilistic assessments (Bonano *et al.*, 1990). For example, many people exhibit overconfidence when asked to make probabilistic forecasts, *particularly* when the events are in their domain of expertise. Previous studies have demonstrated that certain difficulties in probability elicitation are reduced if scientists are exposed to the literature on probability elicitation, subjected to calibration exercises, and trained in tools to combat overconfidence (Alpert and Raiffa, 1982; Hawkins and Evans, 1989; Morgan and Henrion, 1990). Time was allotted at the workshop to address each of these concerns.

The Probability Tree for Chloroform

A probability tree is an analytical tool used to decompose a complex series of events into component parts. The tree is particularly useful in cancer potency assessment because the process of moving from external exposure to estimates of the induction of malignant tumors is complex. By decomposing the assessment of cancer potency into a series of more simple elements, the tree allows experts to think more clearly about analytical inferences that are closer to their area of expertise.

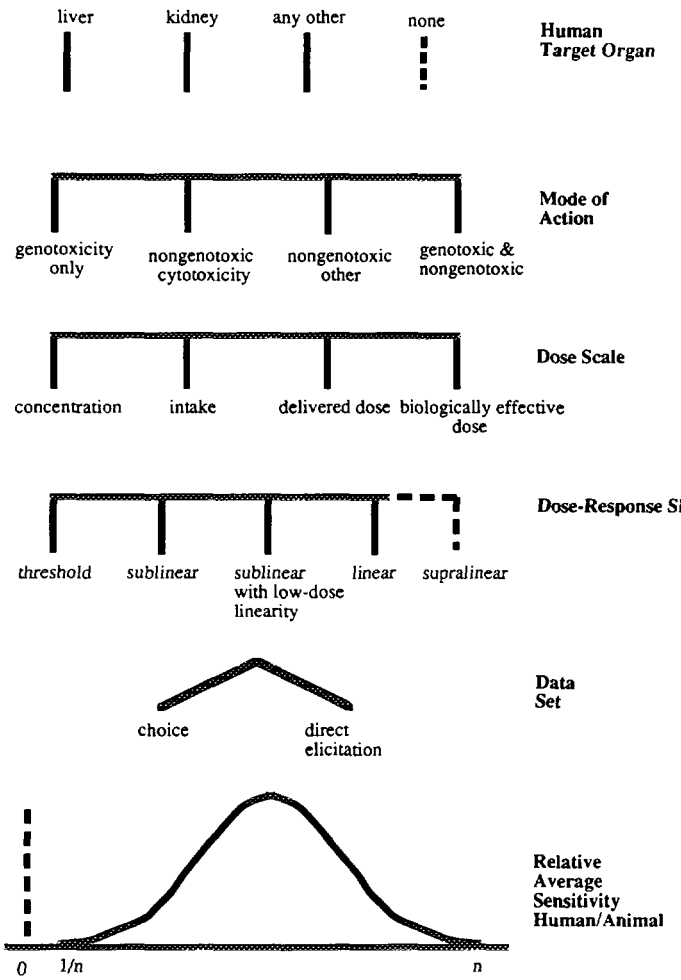

FIG. 1. Chloroform probability tree.

The structure of a probability tree refers to the elements represented at each level, the relationship among the levels (sequence), the number of levels, and the degree of resolution (number of branches) at each level. The structure is crucial to a distributional analysis. The design of a tree, as with any other model, reflects a tension between scientific completeness and pragmatism. The tree structure that is presented, in simplified form, in Fig. 1 represents the consensus structure by the group.

Several features of the probability tree are particularly important. Multiple human target organs were considered plausible and therefore a subtree was created for each organ. These subtrees each have five levels—mode of action (mechanism), dose scale, dose–response shape, data set, and animal to human extrapolation. At the first level of the tree—mechanism—our experts believed that the plausible mechanisms for chloroform's carcinogenicity were genotoxicity only, cytolethality alone, another nongenotoxic mechanism (such as mitogenesis) alone, or some combination of genotoxic

and nongenotoxic mechanisms. A variety of plausible measures of dose were considered in the second level of the tree, including concentration in the exposure medium, intake (applied dose adjusted for rate of entry to the body), pharmacokinetic measures (such as delivered dose to a particular organ), and pharmacodynamic measures of dose (such as estimates of cell killing in the target organ). The third level of the tree—shape of dose–response relationship—addresses the form of the dose–response function below the region of experimental observation. Here four categories of shape were considered as possibilities—a strict threshold, sublinearity, sublinearity with low-dose linearity, and strict proportionality. The choice of data set (i.e., specific animal bioassay), the fourth level of the tree, was approached in two different ways. In one, each expert was asked to identify the most relevant data set for risk assessment purposes. In the other, the experts were asked to first identify all data sets that they felt were useful for risk assessment and second, to indicate, by assigning probability weights to them, the relative emphasis that should be given to each data set. The fifth level of the tree—animal to human extrapolation—focused on the relative average sensitivity of humans in comparison with the species used in the bioassay(s) of interest.

The probability tree makes clear the conditional nature of an expert's beliefs about the relative plausibility of various hypotheses. For example, one's view about the best measure of dose depends strongly on which mechanism of carcinogenicity is considered. Pharmacodynamic measures of dose such as indices of cell killing would presumably be given little emphasis when an expert is considering the possibility that genotoxicity alone is responsible for chloroform's carcinogenicity. Similarly, one's view about the shape of the dose–response function might be strongly dependent upon which measure of dose is considered. While thinking about cytolethality as the potential mode of action of chloroform, and believing that an available measure of cytolethality (i.e., PTDEAD as defined in Corley *et al.*, 1990; Reitz *et al.*, 1990) was essentially a perfect measure of "biologically effective dose," an expert might believe that, on this dose scale, a proportional dose–response function would be appropriate. The same expert might believe that when a more "naive" measure of dose (such as concentration) was used, a sublinear curve or strict threshold would be appropriate. The actual probability tree used in the elicitation of expert judgment was more complex than Fig. 1 might suggest, in that each level of the tree was "nested" with the preceding levels of the tree to allow for this type of conditional belief structure.

Eliciting Weights from Individual Experts

During the summer and fall of 1992 (i.e., after the workshop), the authors interviewed each scientist to elicit probabilistic weights and obtain expertise ratings. The interviews lasted from a half day to 2 days in length and were typically performed by three of the authors at the office of the expert. Several experts had to be interviewed on more than 1 day and follow-up phone calls were sometimes necessary to clarify information that had been gathered in person.

Since scheduling conflicts caused some interviews to be conducted several months after the chloroform workshop, a synopsis of the chloroform science was prepared by the authors and discussed with each expert. A "refresher" on the difficulties with subjective probability assessment was also provided to each expert in order to combat

overconfidence. The probability tree (summarized in Fig. 1) was used with the six experts who relied on animal data in the prediction of human risk. Many key published papers and reports were present at the interviews, and experts were afforded the opportunity to look at any of this material. Each expert provided judgments for all levels of the tree. Interviewees were encouraged to proceed down a single path along the probability tree to completion before considering another path, rather than consider multiple paths concurrently at a given level of the tree. This approach was found helpful in reinforcing the conditionality inherent in the probability tree.

The process of eliciting judgmental probabilities from scientists is quite different from the process of eliciting opinions from laypeople in a standard opinion survey. Eliciting expert judgments follows a structured protocol but is inevitably interactive (Morgan *et al.*, 1979). Some scientists are comfortable moving directly to answering probabilistic questions while others are more comfortable discussing the critical qualitative issues before answering numerical questions. Some scientists were very reluctant to provide judgmental probabilities even though they were eager to offer incisive qualitative judgments.

The role of the interviewers was to identify the scientific rationale for judgments and to remind the interviewee about the conditionality of the probability tree. In some cases the interviewers acted as "devil's advocate" with an expert in order to pinpoint ambiguities, clarify the rationale for a claim, and highlight data that appeared to be contrary to a scientist's stated opinion. With the exception of 2 follow-up interviews, all others were performed by the same three authors.

The result of each interview was a set of expert judgments of the relative plausibility of each alternative for each level of the tree. It is important to note that this process does not require an expert to choose one "right" answer at a given level of the tree. Instead, an expert assigns probabilistic weight, which must be nonnegative and sum to 1.00, to each alternative based on his or her judgment of the scientific plausibility of each choice. For example, expert C, when considering the liver as the target organ, placed probability of 0.85 on cytolethality as the mode of action, 0.13 on some other nongenotoxic mechanism, and 0.02 on the possibility that both genotoxic and nongenotoxic mechanisms are responsible for liver tumor formation. At the level of the tree concerned with relative average sensitivity, the experts gave a distribution reflecting their uncertainty about the ratio of sensitivity to chloroform between the average human and average rodent. Note that the use of the average was intended to account for different degrees of heterogeneity among human and rodent populations.

Finally, as decided at the workshop, experts were asked both to judge which of the participating scientists was most knowledgeable at each level of the tree and to divide a weight of 1.00 over the levels of the tree to indicate their own level of expertise at each level. Weights could range from 1.00 on one level with no weight on the others, to a weight of 0.20 on all levels for an expert who considered himself equally knowledgeable at each level of the tree (the tumor site level of the tree was not included). These ratings are used later in the analysis to aggregate scientific opinions.

Constructing Risk Distributions

Each path through the probability tree leads to a single estimate of added cancer risk. The tree used for this analysis has a few hundred potential ways of calculating

potency. An estimate of added risk, derived from each potency estimate, is present in the final distribution in proportion to the probability assigned to the path leading to that estimate by each expert. These terminal subjective probabilities are derived by multiplying the elicited probabilities assigned to each alternative along the path to that estimate of potency. For example, for one of many paths through the kidney subtree, an expert may place a 0.20 probability on genotoxic and nongenotoxic as the mode of action, 0.25 on the AVEMMB (average macromolecular binding from the Corley *et al.* physiologically based pharmacokinetic model) delivered-dose measure, 0.50 on the sublinear with low-dose linearity dose–response shape, choose the Jorgenson *et al.* male rat kidney tumor data set for dose–response modeling, and believe in equal relative average sensitivity between humans and rats (for the delivered-dose measure). This analysis yields an estimate (MLE) of added risk from 100 ppb chloroform in drinking water of 2.2×10^{-7}. This risk estimate would appear in the final distribution with a weight of 0.025 (i.e., 0.20*0.25*0.50). The final distribution for each expert is a summary of the hundreds of estimates of risk weighted by the experts' subjective probability that the path through the tree is correct. Different distributions for the experts, therefore, arise from differences in the relative plausibility that they assign to various alternatives (and therefore paths through the tree).

In this way, risk distributions are constructed for each organ subtree and for total risk combined across sites. Total risk distributions are calculated assuming that risk estimated from different animal tumor sites is independent, i.e., $P_{\text{human cancer}}(\text{any organ}) = 1 - [1 - P_{\text{human cancer}}(\text{liver})][1 - P_{\text{human cancer}}(\text{kidney})][1 - P_{\text{human cancer}}(\text{any other organ})]$.

RESULTS AND DISCUSSION

This analysis presents distributions of potential chloroform carcinogenicity, based on expert judgment, for lifetime residential exposure to 100 ppb of chloroform in drinking water.

Results for Individual Experts

Figure 2 presents the six distributions for total human cancer risk derived from the interviews with the experts in toxicology/cancer biology, pharmacokinetics, and dose–response modeling. The results of our work with the epidemiologists are presented in a subsequent section.

Notably, for all experts there was some probability that the cancer risk from daily consumption of 2 liters of water containing 100 ppb chloroform could be very small ($<10^{-10}$). An incremental risk $<1 \times 10^{-10}$ can, and in this case does, arise in several ways. If an expert assigned weight to a pharmacodynamic measure of dose which has a threshold, and if 100 ppb of chloroform in drinking water does not exceed the threshold, the added risk is zero. Alternatively, an expert may place some weight on relative animal-to-human sensitivity of zero, indicating some possibility that the carcinogenic response in animals has no relevance to humans. Finally, very small risk estimates arise from certain combinations of dose scale and dose–response shape (hereinafter any risk $<10^{-10}$ will be referred to as zero risk, although it is clear that all

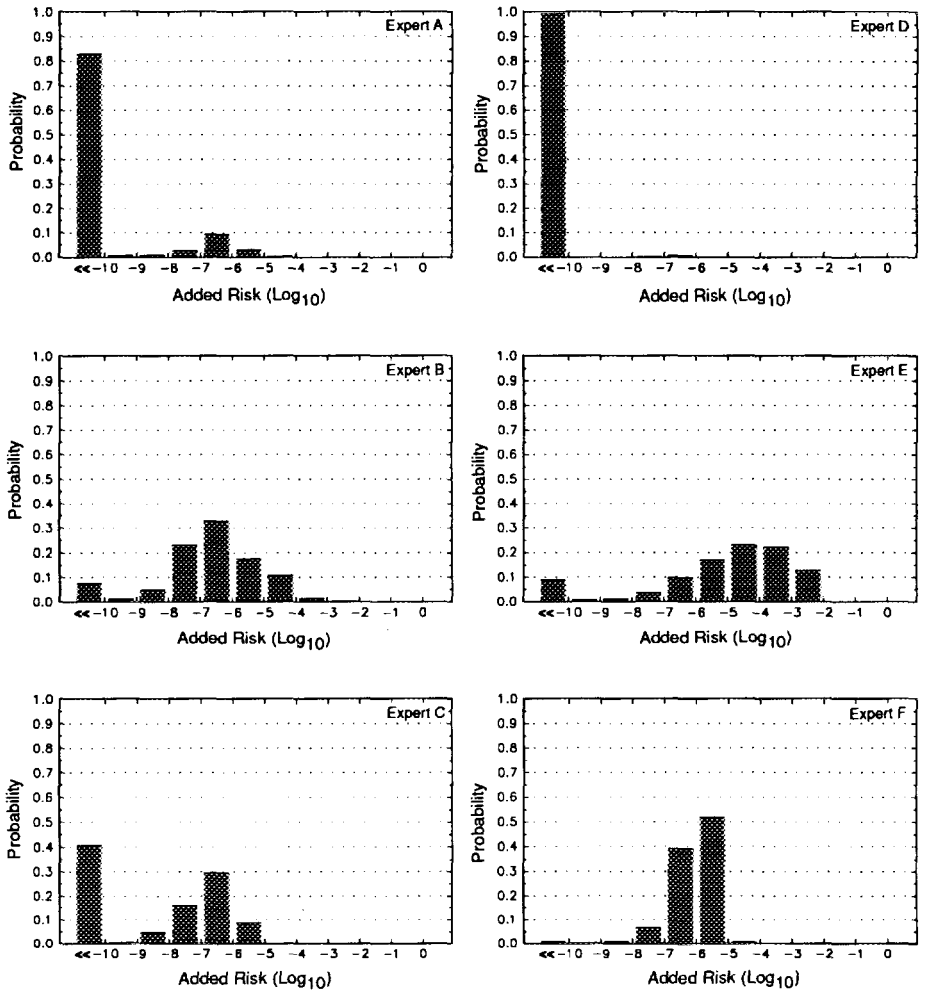

FIG. 2. Risk distributions for exposure to 100 ppb chloroform in drinking water (any organ). Risk distributions calculated all paths down the tree judged relevant by each expert, weighted by the judged probability of being correct. Experts are identified only by letters A through F. Risks are calculated for any organ using the relationship $P_{\text{human cancer}}(\text{any organ}) = 1 - [1 - P_{\text{human cancer}}(\text{liver})][1 - P_{\text{human cancer}}(\text{kidney})][1 - P_{\text{human cancer}}(\text{any other organ})]$, which combines the risks from the subtrees constructed for different target organs.

of the risks are not strictly zero). The chance of zero risk varied considerably among experts, ranging from 1 to more than 99%. With the exception of expert E, little probability was associated with risks greater than 10^{-3}. For four of the experts, there was considerable probability (>60%) that risks would fall between 10^{-8} and 10^{-3}.

Inspection of Fig. 2 suggests that the judgments of the experts lead to distributions of two sorts. Experts A, C, and D had probability distributions with considerable chance that risks could be zero. Their distributions have expected values of 2×10^{-7}, 3×10^{-7}, and 1×10^{-9}, respectively. The probability distributions obtained from experts B, E, and F had little probability of zero risk and considerable chance that risk could exceed 10^{-6}. The expected values for these distributions were 1×10^{-5}, $4 \times$

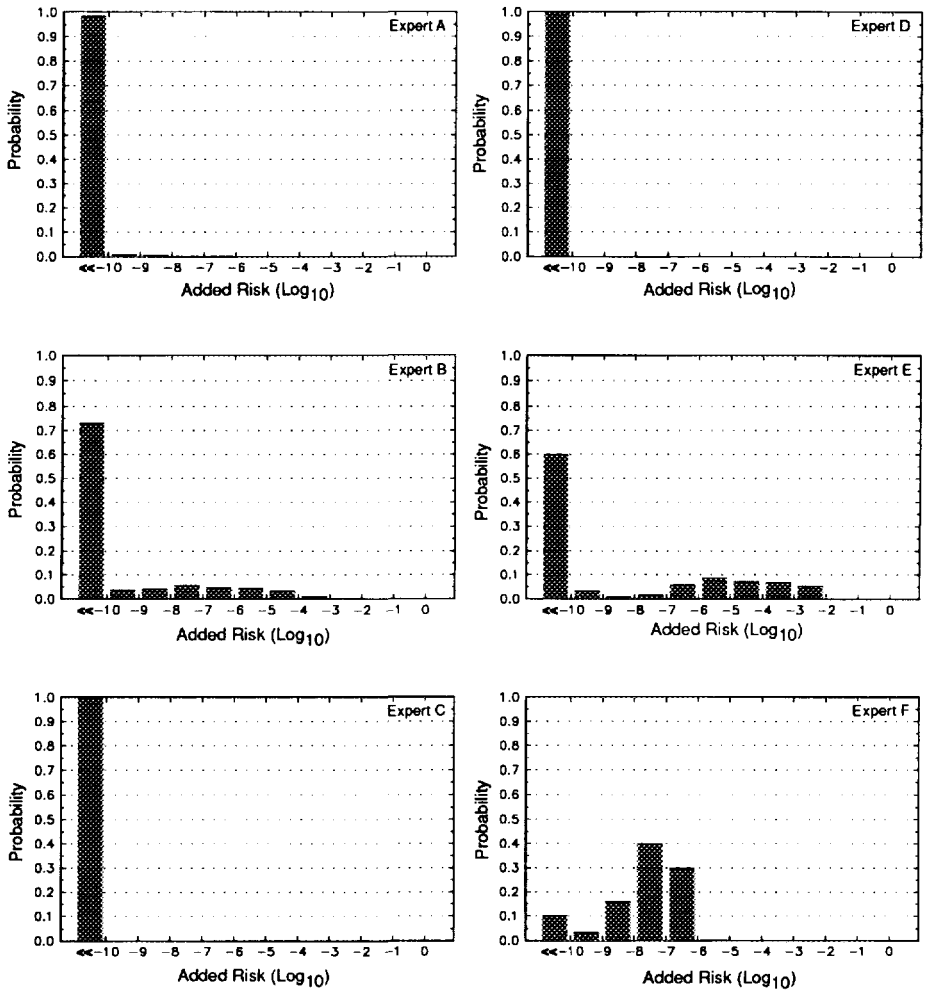

FIG. 3. Risk distributions for exposure to 100 ppb chloroform in drinking water (liver). Risk distributions calculated for each expert from the liver subtree.

10^{-4}, and 2×10^{-6}, respectively. Clearly, there was no strong consensus among the experts about the level of risk posed by drinking water with 100 ppm of chloroform.

Given such a wide range in expected values, it is of interest to identify the major factors contributing to the different estimates of risk. It is interesting to examine the influence of the animal liver tumor data (Fig. 3) and the animal kidney tumor data (Fig. 4) on these estimates. For cancer risk based on liver tumors in the animal bioassays, the experts' distributions fell into two groups (Fig. 3): three indicated more than 98% of cumulative probability on zero and expected values of about 10^{-10} and three with 25% or more chance of risks greater than 10^{-10} and expected values between 10^{-4} and 10^{-6}. A major difference between these two groups was judgment about the best data set for dose–response modeling. One group believed that the Jorgenson *et al.* (1985) female mouse liver tumor data set should be used because it had a better study design (more dose groups and larger numbers of animals in the lower dose groups) and the

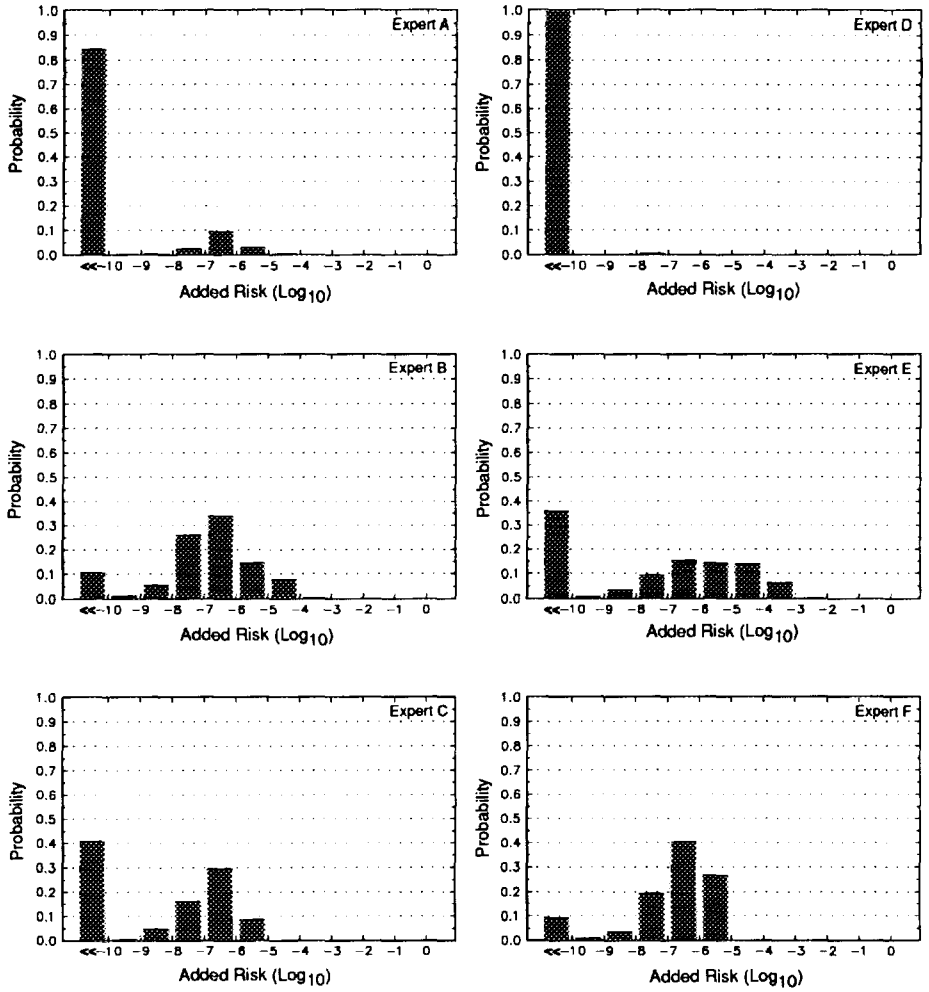

FIG. 4. Risk distributions for exposure to 100 ppb chloroform in drinking water (kidney). Risk distributions calculated for each expert from the kidney subtree.

animals were exposed by drinking water, the same route by which humans would be exposed. In this bioassay there was no increase in liver tumors over background in either rats or mice. Two of the remaining three experts believed the NCI (1976) data sets for male and female mice were best for computing risk. This judgment was based in part on concern about mortality in the high-dose groups in the Jorgenson study, a problem not encountered in the NCI bioassay. In the NCI experiments, in which chloroform was administered by gavage, a strong dose–response relationship was evident. The final expert believed that the Jorgenson *et al.* (1985) male kidney tumor data set was best.

Inspection of the distributions based on kidney tumor data (Fig. 4) reveals essentially the same shapes seen for *any* organ, indicating that overall estimates of risk were heavily influenced by risk computed from the kidney tumor subtree. Here, all experts judged the Jorgenson *et al.* (1985) drinking water bioassay (male rat kidney tumors)

to be best. There was little difference in an experts' judgment about the mode of action for these two organs. For the kidney, however, no data were available to estimate a biologically effective dose (based on cytolethality) so most experts assigned most weight to delivered dose instead. Another important difference between the analyses based on the liver tumor data and the kidney tumor data is that the kidney tumor data (from male rats in the Jorgenson *et al.* study) appear to be essentially linear in the experimental region. The differences in the risk distributions of various experts were due largely to differences in their views about the relative plausibility of threshold versus nonthreshold dose–response shapes. The two experts whose risk distributions included little probability of risks greater than zero (experts A and D) assigned large weights to threshold models of delivered dose and risk. Experts C and E, whose distributions include appreciable probability that risk is above 10^{-5}, were uncertain about dose–response shape and spread their weights among threshold and nonthreshold dose–response models. Expert B assigned the greatest weight to a sublinear dose–response model. Expert F was somewhat of an exception to this generalization, in that his results were strongly influenced by both choice of dose scale, a very sensitive measure which was based on enzymatic markers of cellular damage, and dose–response shape (strict threshold).

Another influential judgment, responsible for the rather high estimates of risk ($>10^{-4}$) in both liver and kidney risk distributions of experts B and E, concerned the relative average sensitivity of humans and rodents. Experts B and E, who were considered most knowledgeable about interspecies extrapolation by the other experts, expressed considerable uncertainty about the relative average sensitivity, primarily motivated by concerns about our limited knowledge of the extent of heterogeneity in human sensitivity to chloroform. Both experts assigned some probability to the proposition that the average human may be as much as 400 times more sensitive to the carcinogenic effects of chloroform than the rodents tested in the bioassay. Interestingly, expert E also assigned some probability to the proposition that chloroform is not a human carcinogen (the relative average sensitivity is zero).

In the process of computing these estimates, we learned that two experts can be in firm qualitative agreement about a critical issue (e.g., mode of action), yet still exhibit significant quantitative differences. For example, experts A and E both believe that cytolethality alone is the most likely mode of action for chloroform carcinogenicity. However, expert A is far more confident about the role of cytolethality (assigning a probability of 0.95) than is expert E (who assigned a probability of 0.55). This example emphasizes the importance of going beyond qualitative discussions when assessing the relative plausibility of alternative modes of action.

Distributions Combined across Experts

When scientists do not offer similar probabilities for the same event, the process of aggregating expert opinions becomes extremely important and can be controversial. We had recruited scientists with expertise relevant to specific levels of the probability tree. Although all of the scientists completed all levels of the tree, it was clear that all were more comfortable with some levels than others. A strength of the probability tree approach is that we can use the knowledge of individual experts in their area of expertise and then combine the different levels to obtain an overall picture of risk.

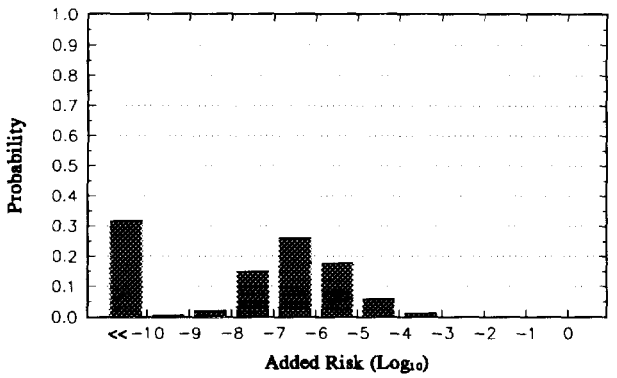

FIG. 5. Combined expert assessment based on peer nomination (any organ). Risk distribution combined across experts with the expert judged by the other experts in the group to be most knowledgeable about each level of the tree providing weights for each branch.

A variety of approaches could be used to aggregate the results across experts. Perhaps the most simple would be to average the final risk distributions for the six experts. While this would be simple, it might not take full advantage of the diversity of scientific background (and knowledge) among the group of experts. Recall that experts from several distinct areas of specialization—cancer biology/toxicology, pharmacokinetics, and dose–response modeling—were involved in this exercise. An alternative to the simple averaging approach would involve a weighted combination of expert judgments. Below we present two such weighted combinations. The first combines the judgments of the experts considered most knowledgeable at each level by the other experts (Fig. 5). The second combines the experts' judgments weighted by their own assessment of their relative expertise (Fig. 6). Both of these approaches rely on the experts' opinions of their areas of relative strength and weakness.

These composite distributions are based on calculations using the "best" animal bioassay data set (as selected by the experts), weights in the probability tree (as provided

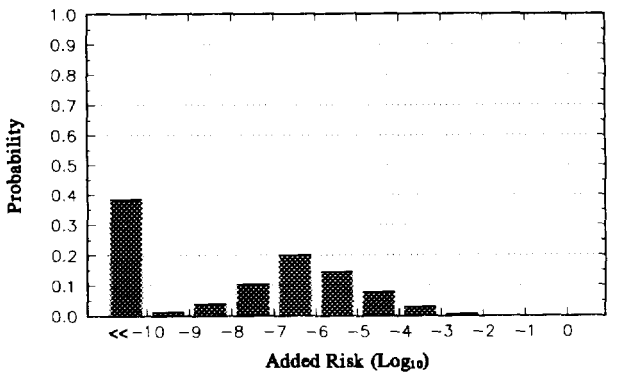

FIG. 6. Combined expert assessment based on self-weighting by experts (any organ). Risk distribution combined across experts with each expert weighted according to self-assessed expertise at each level of the tree.

by the appropriate experts), and bootstrap estimates of parameter uncertainty (calculated by the authors using standard statistical procedures (Sielken, 1989)).

The distribution in Fig. 5 (which we will call the "most knowledgeable" distribution) has a substantial spike on very small risks ($<1 \times 10^{-10}$) with a long right-hand tail that extends considerably above 1 chance in 1,000,000, a point that is sometimes considered a *de minimus* value by regulators (Rosenthal *et al.*, 1992). Interesting summary statistics from the most knowledgeable distribution include the median (1.1×10^{-7}), the expected value (4.2×10^{-6}), and the 95th percentile (1.8×10^{-5}). This risk distribution is heavily influenced by risks estimated using kidney tumor data (the most knowledgeable probability distribution based only on the liver tumor data has an expected value 3 orders of magnitude smaller than the kidney tumor data, with a 95th percentile that is more than 10 orders of magnitude smaller).

The summary statistics for the combined risk distribution based on self-weighting of expertise (the "self-weighted" distribution) are median (4.8×10^{-8}), the expected value (3×10^{-5}), and the upper 95th percentile (6×10^{-5}). It is apparent that this distribution is qualitatively similar to the most knowledgeable distribution but has an even longer right-hand tail. Again risks based on kidney tumor data are very influential, although the differences between liver and kidney self-weighted distributions are smaller (expected values differ by a factor of 2, the 95th percentiles by 16-fold).

It is instructive to compare these combined distributions with the plausible upper-bound (1.7×10^{-5}) that is generated by using EPA's default assumptions and linearized multistage model. EPA states that the true value of the risk could be as low as zero, but gives no indication of the relative plausibility of values between zero and the upper bound. The EPA plausible upper bound falls at about the 95th percentile of Fig. 5 and above the 95th percentile of Fig. 6. The distributions from our experts indicate that it is highly likely that the risk is less than 1×10^{-6} with a substantial chance that it is zero. This information might be important to a risk manager making risk comparisons or setting priorities for resource allocation.

Results from the Epidemiological Assessments

There are no published epidemiological studies of exposure to chloroform per se. All epidemiologically based inferences about human cancer risk are based on studies of chlorinated drinking water. Chlorination leads to the formation of many halogenated compounds in water, but chloroform is predominant. The amount of halogenated material, as well as the levels of different compounds, varies by location, water source, and degree of chlorination. As an indication, an EPA survey of 100 water supplies found chloroform levels of 0.1–1.0 ppb in 39%, 1–10 ppb in 49%, 10–100 ppb in 12%, and fewer than 1% had levels between 100 and 1000 ppb (Perwak *et al.*, 1980).

The epidemiologists in our study had varying degrees of familiarity with the studies of chlorinated drinking water. Ken Cantor has an active research program on chlorinated drinking water and has conducted some of the most highly regarded studies of drinking water and cancer. Jack Siemiatycki had recently served on an International Agency for Research on Cancer (IARC) panel which had examined the available studies and found "inadequate evidence for the carcinogenicity of chlorinated drinking water in humans" (IARC, 1991). Philip Cole had little direct knowledge of the epidemiological

data for chlorinated drinking water but extensive experience in the epidemiology of cancer.

The personal interview with each epidemiologist proceeded very much like that for the biological scientists. First, the goals of the project and the workshop were reviewed. We then discussed the relevant epidemiological literature, including a meta-analysis (Morris *et al.*, 1992) which had been published between the workshop and the interviews. Fallibilities in expert judgment were then reviewed, again with an emphasis on expert overconfidence.

Each epidemiologist provided a direct elicitation of a probability distribution of cancer risk from chlorinated drinking water. The basic question posed to the experts was "What is the risk due to lifetime consumption of chlorinated drinking water with a level of trihalomethanes equivalent to 100 ppb of chloroform?" None of the experts was comfortable with this framing of the question and all instead chose to give risks for a less specific scenario of lifetime residential exposure to chlorinated drinking water. The epidemiologists were quite uncertain about what role chloroform might be playing in the potential cancer risk. Cole and Siemiatycki felt that there was insufficient information to evaluate chloroform's role, while Cantor believed that it was unlikely that chloroform played a large role in risk from chlorinated drinking water. Because of this, the distributions from the epidemiologists are not directly comparable to the distributions based on animal data for chloroform.

Rather than giving information directly on the impact of chlorination on overall cancer risk, the epidemiologists preferred to give information about the relative risks of developing specific cancers. All three epidemiologists discussed the bladder, colon, and rectum; two provided estimates of lung cancer risk and one considered stomach cancer.

The final risk distributions for the epidemiologists, which are shown in Fig. 7, were calculated using background incidence data from NCI (1981) and the relative risk distributions directly elicited from the experts. The expected values for the experts' overall distributions differed only by a factor of about 5 (from 1.5×10^{-4} to 7.8×10^{-4}). Likewise, the 95th percentiles only differed by a factor of about 7. The main differences among the epidemiologists' distributions are in the lower fractiles and depend on whether an expert was willing to assign any weight to the hypothesis that chloroform has a protective effect against cancer (a relative risk less than 1). Interestingly, some animal experiments suggest that chloroform may protect against tumors initiated with powerful genotoxic agents (Daniel *et al.*, 1989). Overall, the epidemiologists' risk distributions for chlorinated drinking water are wide enough (i.e., express enough uncertainty) to encompass all of the values in the distributions constructed by the biological scientists.

CONCLUSION

Utilizing all relevant information is the key to any complete characterization of risk. The distributional approach described here allows all relevant scientific information to be considered and used in a comprehensive potency assessment. It encourages quantitative consideration of both statistical and model uncertainty without artificially forcing one to make the controversial decision of when science is complete enough to abandon default methods of risk calculation.

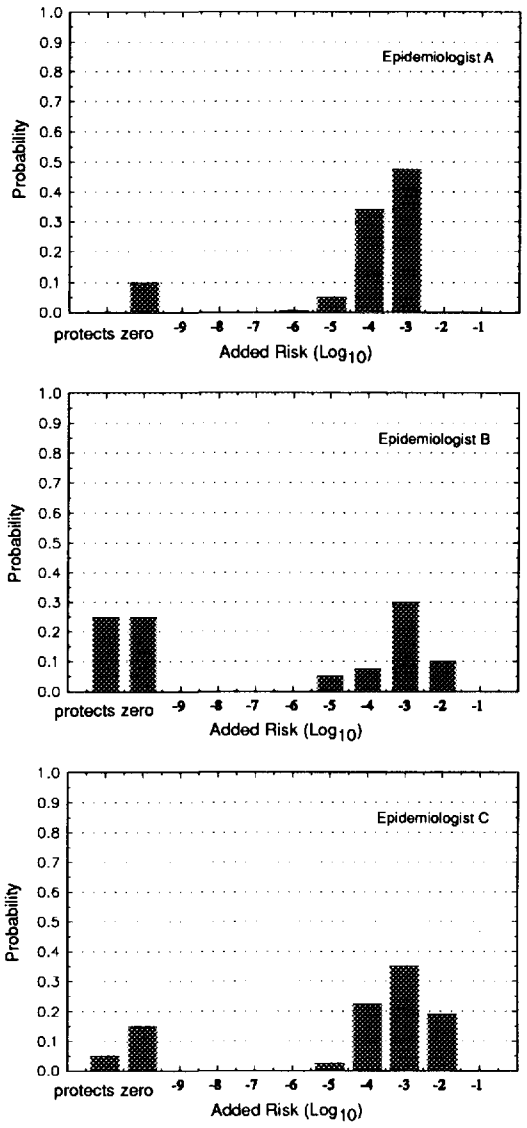

FIG. 7. Epidemiologist's risk distributions from chlorinated drinking water. Risk distributions calculated for lifetime exposure to chlorinated drinking water as judged by epidemiologists. Risks were calculated using incidence data from SEER, 1977. Epidemiologists are identified only by letter A–C. Risk distributions were calculated as follows. Distributions for the added risk at each target organ considered (bladder, colon, rectum, lung, and stomach) were computed by multiplying the estimated human background tumor rate (NCI, 1981) by the distribution on relative risk obtained from each expert. These distributions were then combined under the assumption of independent tumor processes.

Naturally, in order to incorporate all relevant scientific information and capture the current state of knowledge in the expert community, the distributional method requires more time and resources to perform than does a standard potency assessment. For example, the tasks of developing a probability tree, probability training, and eliciting judgmental probabilities from the experts are not elements of a standard potency assessment. Similar to recent recommendations (e.g., National Research Council, 1994), one can envision the distributional approach used in an iterative approach, applied to high-stakes decisions in which screening or default risk assessment is judged to be inadequate. Consequently, it is useful to consider what the chloroform case study tells us about the practicality, strengths, and weaknesses of the method.

A virtue of the method is that it provides a starting point for determining which research strategies are most likely to reduce uncertainty about the carcinogenic potency of chloroform. For compounds with little baseline data, value-of-information tools can be used to make default potency assessments that are then refined as more data are generated. For an example of recent work illustrating this approach, see Taylor *et al.* (1993). In this study we identified the choice of appropriate data set as the critical source of difference for risk distributions based on liver tumors and the selection of dose–response model as the key uncertainty in using kidney tumor data. Both of these issues may be resolved through further research. Better understanding of mechanisms of carcinogenesis and factors influencing pharmacokinetics and pharmacodynamics in different species will help in reconciling various animal data sets and how risk should be extrapolated from high to low dose. Interestingly, in the area which has received the most research attention, liver tumors caused by chloroform, the experts' distributions indicated that human risks would be quite low if a risk exists at all. For the kidney, which has not been studied nearly as much, fewer mechanistic and pharmacokinetic studies were available and risk estimates tended to be much higher.

A strength of the distributional method is the assessor's ability to incorporate new mechanistic data while taking into account technical reservations about the quality, relevance, and uncertainty in such information. In this case, all of the biological scientists' judgments were influenced by the data on the relationship between cell proliferation and tumor formation, especially relating to liver tumors. There were quantitative differences in how much confidence was assigned to the cell proliferation hypothesis. As long as multiple expert scientists are consulted, this method allows risk managers to learn about how differences in scientific judgment translate into different quantitative risk estimates.

Current risk characterization frequently presents a point estimate of cancer risk along with a qualitative statement about uncertainty in the estimate, including the fact that there could be no risk at all. Point estimates convey a false degree of precision about risk to the risk manager and narrative statements of uncertainty are rarely considered by risk managers, journalists, or the public (Gray and Graham, 1991). Some have suggested that risk managers be presented with multiple risk estimates based on different assumptions and data (Habicht, 1992; American Industrial Health Council, 1989; American Industrial Health Council, 1992). However, it is unclear how a risk manager would evaluate the relative likelihood of these many risk estimates. Although some would say that a potential weakness of this method is that it does not provide a single, unambiguous estimate of risk that can be used by risk managers, we believe that this is actually a strength. Rarely is there a single value with complete

scientific support. When the science is uncertain and scientists have differing opinions, that should be reflected in what risk assessors tell risk managers (Graham *et al.*, 1988). The probability distribution produced by this method conveys both the range of possible risk values and the relative likelihood that each is correct. While a single summary statistic can be reported from the probability distributions, there is no strong technical rationale for reporting one summary statistic as opposed to another. More experience with the distributional method, including collaboration with risk managers on how to interpret and use the results, will help to convince agencies of the value of quantitative characterization of uncertainty in estimation of risk using a weight-of-the-evidence approach.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support provided by the American Industrial Health Council. John Evans and Andrew Smith also received partial support from the U.S. Environmental Protection Agency's Cooperative Agreement with Harvard on Environmental Health and Public Policy.

REFERENCES

ALPERT, M., AND RAIFFA, H. (1982). A Progress Report on the Training of Probability Assessors. In *Judgment Under Uncertainty: Heuristics and Biases* (D. Kahneman, P. Slovic, and A. Tversky, eds.), pp. 294–305. Cambridge Univ. Press, Cambridge, England.

American Industrial Health Council (1989). *Presentation of Risk Assessments of Carcinogens. Report of an Ad Hoc Study Group on Risk Assessment Presentation.* American Industrial Health Council, Washington, DC.

American Industrial Health Council (1992). *Improving Risk Characterization. Summary of a Workshop Held in Washington D.C. on September 26 & 27, 1991.* American Industrial Health Council, Washington, DC.

BONANO, E. J., HORA, S. C., KEENEY, R. L., AND VON WINTERFELDT, D. (1990). *Elicitation and Use of Expert Judgment in Performance Assessment for High-Level Radioactive Waste Repositories.* U.S. Nuclear Regulatory Commission, NUREG/CR-5411.

COOKE, R. (1993). *Experts in Uncertainty: Opinion and Subjective Probability in Science.* Oxford Univ. Press, New York.

CORLEY, R. A., MENDRALA, A. L., SMITH, F. A., STAATS, D. A., GARGAS, M. L., CONOLLY, R. B., ANDERSEN, M. E., AND REITZ, R. H. (1990). Development of a physiologically based pharmacokinetic model for chloroform. *Toxicol. Appl. Pharmacol.* **103,** 512–527.

DANIEL, F. B., DEANGELO, A. B., STOBER, J. A., PEREIRA, M. A., AND OLSON, G. R. (1989). Chloroform inhibition of 1,2-dimethylhydrazine-induced gastrointestinal tract tumors in the Fisher 344 rat. *Fundam. Appl. Toxicol.* **13,** 40–45.

ESHENBRENNER, A. B., AND MILLER, E. (1945). Induction of hepatomas in mice by repeated oral administration of chloroform with observations on sex differences. *J. Natl. Cancer Inst.* **2,** 251–255.

EVANS, J. S., GRAHAM, J. D., GRAY, G. M., AND SIELKEN, R. L., JR. (1993). A distributional characterization of cancer risk from formaldehyde exposure. *Risk Anal.* **14,** 25–34.

FINKEL, A. M., AND EVANS, J. S. (1987). Evaluating the benefits of uncertainty reduction in environmental health risk management. *J. Air Poll. Control Assoc.* **37,** 1164–1171.

GRAHAM, J. D., GREEN, L. C., AND ROBERTS, M. J. (1988). *In Search of Safety: Chemicals and Cancer Risk.* Harvard Univ. Press, Cambridge, MA.

GRAHAM, J. D., HAWKINS, N. C., AND ROBERTS, M. J. (1988). *Expert Scientific Judgment in Quantitative Risk Assessment. Banbury Report 31: Carcinogen Risk Assessment. New Directions in the Qualitative and Quantitative Aspects.* Cold Spring Harbor Laboratory, New York.

GRAY, G. M., COHEN, J. T., AND GRAHAM, J. D. (1993). The challenge of risk characterization: current practice and future directions. *Environ. Health Perspect. Suppl.* **101,** 203–208.

GRAY, G. M., AND GRAHAM, J. D. (1991). Risk assessment and clean air policy. *J. Policy Anal. Manage.* **10,** 286–295.

HABICHT, F. HENRY, II (1992). *Guidance on Risk Characterization for Risk Managers and Risk Assessors.* U.S. Environmental Protection Agency, February 26.

HAMMITT, J. K., AND CAVE, J. A. K. (1991). *Research Planning for Food Safety: A Value of Information Approach.* RAND R-3946-ASPE/NCTR, Santa Monica, CA.

HAWKINS, N. C., AND EVANS, J. S. (1989). Subjective estimation of toluene exposures: A calibration study of industrial hygienists. *Appl. Ind. Hyg. J.* **4,** 61–68.

HAWKINS, N. C., AND GRAHAM, J. D. (1988). Expert scientific judgment and cancer risk assessment: A pilot study of pharmacokinetic data. *Risk Anal.* **8,** 615–625.

HEYWOOD, R., SORTWELL, R. J., NOEL, P. R. B., STREET, A. E., PRENTICE, D. E., ROE, F. J. C., WADSWORTH, P. F., WORDEN, A. N., AND VAN ABBÉ, N. J. (1979). Safety evaluation of toothpaste containing chloroform III. Long-term study in beagle dogs. *J. Environ. Pathol. Toxicol.* **2,** 835–851.

International Agency for Research on Cancer (IARC) (1991). Chlorinated Drinking-Water; Chlorination By-Products; Some Other Halogenated Compounds; Cobalt and Cobalt Compounds. In *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans,* Vol. 52. IARC, Lyon, France.

JORGENSON, T. A., MEIERHENRY, E. F., RUSHBROOK, C. J., BULL, R. J., AND ROBINSON, M. (1985). Carcinogenicity of chloroform in drinking water to male Osborne–Mendel rats and female B6C3F1 mice. *Fundam. Appl. Toxicol.* **5,** 760–769.

KEENEY, R. L., AND VON WINTERFELDT, D. (1989). On the uses of expert judgment on complex technical problems. *IEEE Trans. Engin. Manage.* **18,** 83–86.

MORGAN, M. G., AND HENRION, M. (1990). *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge Univ. Press, New York.

MORGAN, M. G., HENRION, M., AND MORRIS, S. C. (1979). *Expert Judgment for Policy Analysis.* Brookhaven National Laboratory, BNL 51358 UC-13.

MORGAN, M. G., MORRIS, S. C., HENRION, M., AMARAL, D. A. L., AND RISH, W. R. (1984). Technical uncertainty in quantitative policy analysis: A sulfur air pollution example. *Risk Anal.* **4,** 201–216.

MORRIS, R. D., AUDET, A., ANGELILLO, I. F., CHALMERS, T. C., AND MOSTELLER, F. (1992). Chlorination, chlorination by-products, and cancer: A meta-analysis. *Am. J. Public Health* **82,** 955–963.

National Cancer Institute (NCI) (1976). *Report on Carcinogenesis Bioassay of Chloroform.* Carcinogenesis Program PB264 018. National Cancer Institute, Bethesda, MD.

National Cancer Institute (NCI) (1981). *Cancer Incidence and Mortality in the United States, 1973–1977.* NCI Monograph No. 57, Surveillance, Epidemiology, and End Result (SEER) Program. National Institutes of Health, Bethesda, MD.

National Research Council (NRC) (1993). *Issues in Risk Assessment.* National Academy Press, Washington, DC.

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

Nuclear Regulatory Commission (NRC) (1975). *Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants.* WASH 1400 NUREG-75/014, Nuclear Regulatory Commission, Washington, DC.

OTWAY, H., AND VON WINTERFELDT, D. (1992). Expert judgment in risk analysis and management: Process, context, and pitfalls. *Risk Anal.* **12,** 83–93.

PALMER, A. K., STREET, A. E., ROE, F. J. C., WORDEN, A. N., AND VAN ABBÉ, N. J. (1979). Safety evaluation of toothpaste containing chloroform II. Long-term study in rats. *J. Environ. Pathol. Toxicol.* **2,** 821–833.

PERWAK, J., GOYER, M., HARRIS, J., SHIMKE, G., SCOW, K., WALLACE, D., AND SLIMAK, M. (1980). *An Exposure and Risk Assessment for Trihalomethanes: Chloroform, Bromoform, Bromodichloromethane, Dibromochloromethane.* EPA-440/4-81-018; NTIS PB 85-211977. U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, DC.

REITZ, R. H., MENDRALA, A. L., CORLEY, R. A., QUAST, J. F., GARGAS, M. L., ANDERSEN, M. L., STAATS, D. A., AND CONOLLY, R. B. (1990). Estimating the risk of liver cancer associated with human exposures to chloroform using physiologically based pharmacokinetic modeling. *Toxicol. Appl. Pharmacol.* **105,** 443–459.

ROE, F. J. C., PALMER, A. K., WORDEN, A. N., AND VAN ABBÉ, N. J. (1979). Safety evaluation of toothpaste containing chloroform I. Long-term study in mice. *J. Environ. Pathol. Toxicol.* **2,** 799–819.

ROSENTHAL, A., GRAY, G. M., AND GRAHAM, J. D. (1992). Legislating acceptable cancer risk from exposure to toxic chemicals. *Ecol. Law Q.* **19,** 269–362.

ROSENTHAL, S. L. (1987). A review of the mutagenicity of chloroform. *Environ. Mol. Mutagen.* **10,** 211–226.

SIEGEL, J. E., GRAHAM, J. D., AND STOTO, M. A. (1990). Allocating resources among AIDS research strategies. *Policy Sci.* **23,** 1–23.

SIELKEN, R. L., JR. (1989). Useful tools for evaluating and presenting more science in quantitative cancer risk assessments. *Toxic Subst. J.* **9,** 353–404.

Society for Risk Analysis (SRA) (1993). *Workshop Report: Key Issues in Carcinogen Risk Assessment Guidelines.* Society for Risk Analysis, McLean, VA.

SPEDDEN, S. E. (1992). *Judgments and Perceptions of Human Risk from Chemical Carcinogens.* Doctoral dissertation, Harvard University, Cambridge, MA.

TAYLOR, A. C., EVANS, J. S., AND MCKONE, T. E. (1993). The value of animal test information in environmental control decisions. *Risk Anal.* **13,** 403–412.

U.S. EPA (1986). *Guidelines for Carcinogen Risk Assessment.* 51 FR 33992.

U.S. EPA (1992). *Working Paper for Considering Draft Revisions to the U.S. EPA Guidelines for Cancer Risk Assessment.* EPA/600/AP-92/003.

WEINSTEIN, M. C., AND FINEBERG, H. V. (1976). *Clinical Decision Analysis.* Saunders, Philadelphia.

WHITFIELD, R. G., AND WALLSTEN, T. S. (1989). A risk assessment for selected lead-induced health effects: An example of a general methodology. *Risk Anal.* **9,** 197–207.

WINKLER, R. L., AND POSES, R. M. (1993). Evaluating and Combining Physicians' Probabilities of Survival in an Intensive Care Unit, *Manage. Sci.* **39,** 1526–1543.

WOLFF, S. K., HAWKINS, N. C., KENNEDY, S. M., AND GRAHAM, J. D. (1990). Selecting experimental data for use in quantitative risk assessment: An expert-judgment approach. *Toxicol. Ind. Health* **6,** 275–291.