Risk Analysis, Vol. 8, No. 4, 1988

Expert Scientific Judgment and Cancer Risk Assessment: A Pilot Study of Pharmacokinetic Data

Neil C. Hawkins[1] and John D. Graham[2]

Received December 22, 1987; revised April 14, 1988

When high-dose tumor data are extrapolated to low doses, it is typically assumed that the dose of a carcinogen delivered to target cells is proportional to the dose administered to test animals, even at exposure levels below the experimental range. Since pharmacokinetic data are becoming available that in some cases question the validity of this assumption, risk assessors must decide whether to maintain the standard assumption. A pilot study of formaldehyde is reported that was undertaken to demonstrate how expert scientific judgment can help guide a controversial risk assessment where pharmacokinetic data are considered inconclusive. Eight experts on pharmacokinetic data were selected by a formal procedure, and each was interviewed personally using a structured interview protocol. The results suggest that expert scientific opinion is polarized in this case, a situation that risk assessors can respond to with a range of risk characterizations considered biologically plausible by the experts. Convergence of expert opinion is likely in this case if several specific research strategies are executed in a competent fashion. Elicitation of expert scientific judgment is a promising vehicle for evaluating the quality of pharmacokinetic data, expressing uncertainty in risk assessment, and fashioning a research agenda that offers possible forging of scientific consensus.

KEY WORDS: Bayesian; cancer risk assessment; delivered dose; pharmacokinetics; expert judgment.

1. INTRODUCTION

Since the Supreme Court's 1980 decision in the famous benzene case,[1] the practice of quantitative risk assessment has accelerated rapidly in federal agencies. Often faced with limited toxicology data and inconclusive epidemiology, such agencies have been forced to develop standard assumptions for use in cancer risk assessment. The U.S. Environmental Protection Agency has been a leader in this area through the efforts of the Carcinogen Assessment Group (CAG).

CAG's typical risk assessment assumptions[2] in using animal data have been published in EPA guidelines[3] and are widely employed in the public and private sectors. The use of the CAG's standard approach has several strengths: the approach can easily be applied to many chemicals; the upper confidence limit (UCL) of the multistage model is thought to represent an upper bound on risk; the approach is

[1]Department of Environmental Science and Physiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115. Present address: Dow Chemical Company, Health and Environmental Sciences, Industrial Hygiene Laboratory, Bldg. 1803, Midland, Michigan 48674.

[2]Department of Health Policy and Management, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115.

0272-4332/88/1200-0615$06.00/1

predictable when applied consistently; and, lastly, the assumptions of the standard approach are explicit and are becoming more widely understood by the broader community of risk assessors.[4]

Although the CAG approach has strengths, it also suffers from weaknesses.[4] By focusing primarily on producing an upper bound on risk, the approach leaves unstated both the degree of uncertainty about risk and the best characterization of risk. Reporting true uncertainty is necessary for honestly communicating risk both to the risk manager and the general public. Informed with genuine uncertainty estimates, the risk manager can make more thoughtful decisions regarding what margins of safety are needed in light of the relative social costs (health and economic) of incorrect decisions. The value of gathering additional scientific data for the risk manager can also be considered if genuine uncertainty is reported.

Another weakness of the standard approach is a consequence of its simplicity. Molecular biology has been advancing rapidly and CAG analysts have struggled to incorporate new kinds of data. Although such efforts are underway, EPA has been resistant to modifying its standard approach because it is considered conservative. Further, there is confusion about when pharmacokinetic data are compelling enough to depart from the standard approach for a given chemical and how these data should actually be used. Advocates of pharmacokinetic models fear that federal agencies will forgo accuracy for the sake of conservatism and simplicity, thereby discouraging the generation of new pharmacokinetic data.

The CAG approach has been criticized frequently for insufficient biological basis, and EPA risk assessors are now at a crossroad. They generally want to bring more biology into risk assessments, but they are uncertain about when and how data should be used to modify guidelines. Bridging this gap between a simple and inflexible approach to risk assessment and the more complex biologic data available for a given chemical requires careful use of *expert scientific judgment.* We believe that the risk assessor should indeed be reluctant to depart from the standard assumptions because he or she often does not have the requisite training and laboratory experience to evaluate the biological complexities and validity of available data. Today's molecular biology is evolving at such a rapid rate that only practicing scientists are likely to appreciate the subtleties of experimental procedures and the reliability of results. Bridging the gap between the practicing scientist and the risk assessor is, in our opinion, the key to improving risk assessment practice in the near term. The difficult challenge is to persuade the best practicing scientists to give their time and energy to an admittedly speculative exercise that to some extent runs against the culture of science.

2. PILOT STUDY: FORMALDEHYDE AND CANCER RISK

Formaldehyde is a widely used chemical with significant occupational, residential, and environmental exposures. Positive animal bioassay results, first published in 1981, led to concern about potential human risk arising from formaldehyde exposures. Rule-makings are now in progress. From the perspective of quantitative risk assessment, formaldehyde seems like a typical case: definitive animal bioassay results are accompanied by suggestive, yet inconclusive, epidemiology.

What sets formaldehyde apart, however, is the availability of additional "pharmacokinetic" or "delivered-dose" data. As we shall see, these data have sufficient credibility to be considered seriously, while at the same time are not so *definitive* that a risk assessor's decision to use them would be straightforward and noncontroversial. The stakes in the formaldehyde risk assessment are perceived to be large both because of the widespread use of the chemical and because of the possible precedent-setting character of the deliberations. How to utilize the data, if at all, remains an open question for risk assessors. Hence these data, with their biochemical complexities and controversy, present a challenging case for illustrating how expert scientific judgment can help risk assessors.

2.1. CIIT Experiments

The original inhalation bioassay of formaldehyde commissioned by the Chemical Industry Institute of Toxicology (CIIT) included both rats and mice as subjects.[5] The results for rats are summarized in Table I. Squamous cell carcinomas of the nasal cavity were observed in both species, though only the results in rats are considered statisically significant. The dose–response data for rats, which are considered highly nonlinear, have been chosen

Table I. CIIT Bioassay Results for Formaldehyde[a]

Administered formaldehyde concentration (ppm in air)	Total Fischer 344 rats (male and female)	Squamous cell carcinomas	Crude observed incidence
0.0	232	0	0.00
2.0	236	0	0.00
5.6	235	2	0.01
14.3	232	103	0.44

[a]Kerns *et al.* also reported results for B6C3F1 mice. For the same exposure regimen, only two squamous cell carcinomas were reported for the 14.3 ppm exposure group, a result not considered statistically significant.

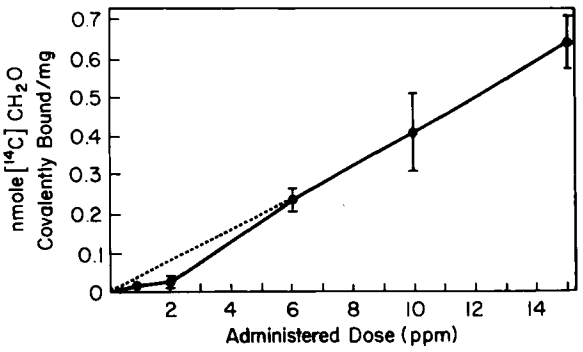

Fig. 1. Results of the delivered-dose experiment.

repeatedly as the primary data for use in cancer risk assessment.[6-13]

CIIT followed up this bioassay with a research program concentrated on mechanisms of formaldehyde toxicity. One component of this work focused on measuring the amount of covalent binding of formaldehyde to rat nasal mucosal DNA.[14] It was presumed that this binding would be related in some way to the observed nasal tumors. The experimental technique used to measure gross binding to DNA was based on incorporation of radiolabeled formaldehyde, a traditional first-step approach for investigating the biologic fate of inhaled chemicals. Formaldehyde presents a difficult measurement problem due to its rapid metabolism and incorporation into the body's carbon pool. Because of this, Casanova-Schmitz *et al.* had to develop a new method using two radioisotopes and a complex mathematical analysis based upon isotope activity ratios in order to differentiate formaldehyde covalently bound to nasal DNA (delivered dose) from metabolically incorporated formaldehyde. This duoisotope method was somewhat new to the cancer research community, although it had been applied to some extent in toxicology. No direct chemical evidence of covalently bound formaldehyde was provided by these experiments, although such evidence has been provided for other substances such as alkylating agents.[15] For practical considerations, the covalent binding was assessed for acute exposure (6-hr formaldehyde exposures on two successive days), even though humans and the rats in the bioassay experienced chronic exposures. Chronic studies using radioactivity incorporation were considered impossible by the CIIT investigators.

The data from the delivered-dose experiments show that the amount of radioactive incorporation in nasal mucosal DNA is not linearly related to administered concentration. The results are displayed in Fig. 1. Starr and Buck of CIIT have attempted to incorporate these data into a modified risk assessment.[16] Their results, which are partly summarized in Table II, indicate a very large reduction in cancer risk compared to that predicted by the standard approach used when no pharmacokinetic data are available. Starr and Buck explicitly lay out the assumptions that must be made in order to use these data in risk assessment. Several scientists and risk assessors in federal agencies, led by Dr. Murray Cohn of CPSC, questioned the validity of the delivered-dose data on methodological grounds and also on the applicability of acute exposures to the chronic exposure situation. A heated letter correspondence through the journal editor ensued with no apparent resolution.[17]

Table II. Starr and Buck's Formaldehyde Risk Assessment[a]

Exposure level	Dose metric	Maximum likelihood estimate of risk	Upper 95% confidence limit
0.1 ppm	Administered	2.5 E-7	1.6 E-4
	Delivered	4.7 E-9	6.2 E-5
1.0 ppm	Administered	2.5 E-4	1.8 E-3
	Delivered	4.7 E-6	6.2 E-4

[a]Three-stage multistage model fitted to CIIT rat bioassay data using either administered or delivered dose as the dose metric. Risk is the lifetime probability of a malignant tumor for chronic exposure at the stated level.

Casanova and Heck followed up the original delivered-dose paper with additional research: a pa-

per on glutathione depletion[18] attempts to explain mechanisms for the incorporation of formaldehyde, while another paper on the "isotope effect"[19] contends that the nonlinear relationship between administered and delivered dose is even greater when this effect is taken into account. The subsequent research using the same measurement method provided an independent verification that the sublinearity result was a real effect. Casanova-Schmitz *et al.* recommend in their published papers that their pharmacokinetic data should be utilized by risk assessors.

The Science Advisory Board of EPA reviewed the 1985 EPA risk assessment and encouraged the agency to consider using the pharmacokinetic data in the final EPA risk assessment. EPA contracted with a consulting firm to arrange an independent panel review of the data to determine their utility for cancer risk assessment. The panel formed a concensus that the data were premature for use in risk assessment, although the research was an "important first step."[20] Casanova-Schmitz *et al.* disagreed with the panel's report and wrote a point-by-point refutation of its conclusions.[21] Ultimately, however, in EPA's most recent 1987 risk assessment, the data were not used and the panel review is cited as the basis for not incorporating the data.[22]

If one examines the charge given the review panel, it seems clear that no other consensus was possible. The panel was asked: Are the data "good enough?" The question to the panel could have been posed differently: What are the strengths and weaknesses of the data relative to the traditional assumption and what influence (if any) do the delivered-dose data have on your subjective beliefs about the relationship between administered and delivered doses of formaldehyde? Our hypothesis is that this approach would provide more insight and guidance to risk assessors without compelling either consensus or divergence of opinion.

3. DESCRIPTION OF PILOT STUDY

The critical assumption in question is whether one should assume that the amount of active substance at the target site is proportional to administered concentration. This traditional linearity assumption has been challenged by Casanova-Schmitz *et al.* on the basis of pharmacokinetic data that are said to demonstrate low-dose (less than 6 ppm) nonlinearity between administered concentration and formaldehyde dose delivered to nasal DNA. The purpose of this pilot study was to demonstrate how experts can be recruited and their judgments elicited and incorporated into cancer risk assessment, using the formaldehyde delivered-dose issue as illustrative of potential cases where expert judgment could be applied. The total cost of the study was less than $25,000, involving two persons part time for a calendar year.

3.1. Selection of Experts

There are many approaches that might be employed to select experts for such a study. Obviously, the ideal approach would be the one that ultimately selects experts who provide the correct quantitative relationship between administered and delivered doses of formaldehyde. Since no one is in a position today to judge which approach would accomplish that, more process-oriented criteria must govern the selection process.

We believe that approaches to expert selection should be evaluated according to at least the following criteria: relevant expertise, explicitness, reproducibility, ease of execution, level of sponsor control, and balance of expert opinions.[4] In the context of regulatory decision making, we believe that, other things equal, approaches should be used that achieve *more* relevant expertise, *more* explicitness, *greater* potential for reproducibility, *more* ease of execution, *less* sponsor control, and *more* balance than would result from a purely informal approach.

One goal of the pilot project was to develop and implement a formal method for assembling scientific experts that is responsive to the criteria given above. This need has become more critical in the last few years since federal agencies have begun to use expert panels more frequently for reviewing scientific data. The formal approach we used relied heavily on the scientific community of interest and its ability and willingness to nominate peers with expertise on specific scientific questions.

Our approach proceeded as follows: Study of the specific scientific question (pharmacokinetics) was undertaken by the authors in order to acquire a sufficient understanding for accurately using the scientific concepts and language associated with the disciplines in question. A computerized literature search was performed using appropriate key words and concepts. Identified papers were examined and a list of authors was assembled. All authors on identified papers were contacted by mail and asked for

names of experts on the specific scientific question. Finally, the list of nominated experts was assembled and the desired number of experts was selected sequentially from a list ordered by nomination counts.

For the formaldehyde pharmacokinetics question, this procedure was a formidable task. The burgeoning literature on pharmacokinetics and delivered dose in chemical carcinogenesis was studied. A computerized literature search for the TOXLINE[23] and MEDLINE[24] databases was performed using key words including "delivered dose," "molecular target," "molecular binding," "covalent binding," and "biomarkers." Several papers relevant to delivered dose were also followed using SCISEARCH.[15] A total of 339 authors was identified in the United States and abroad. Addresses for these 339 were gathered through listings such as *American Men and Women of Science,*[26] *National Faculty Directory,*[27] and the original listings in the scientific papers. A letter requesting the nomination of up to three individuals with expertise on delivered-dose and pharmacokinetic issues was sent to each person. Formaldehyde expertise *per se* was not sought. A random sample of 30 nonrespondents in the United States were selected and phone contact attempted in order to increase the response rate. The overall response rate for those receiving letters was 30%—with the domestic rate at 35% and the international rate at 19%.

We recognize that formaldehyde expertise may have been desirable due to formaldehyde's biochemistry; however, we believed that a preferable group of experts (in terms of criteria given earlier) could be assembled if our exmphasis was on recruiting expertise on carcinogen pharmacokinetics. Moreover, the pool of formaldehyde pharmacokinetics experts is probably quite small and may consist primarily of those involved with the experiments in question.

Rather than using raw counts of nominations, a method for controlling overnominating within research groups was used. In particular, the raw counts were adjusted so that each research "group" (e.g., CIIT, MIT, and NIEHS) contributed at most only one count for a nominated expert. The total number of experts nominated by two or more groups was 29 (25 domestic, 4 international). A subset of these 29 who constitute the top five tiers of nominees (by adjusted counts) is given in Table III. A substantial convergence in nominated experts developed, making selection relatively straightforward.

The eight experts eventually selected (Table IV) were chosen sequentially from the first five tiers in the following manner. Each scientist in the first two tiers was selected outright (Andersen and Swenberg). From the third tier, Tannenbaum was selected, but Starr was not selected because CIIT already had participation by Swenberg. In the fourth tier, Perera and Reitz were selected, but Neumann was not because funding for European travel was unavailable. In the fifth tier, Beland, Dedrick, and Poirier (each representing a different government laboratory) were selected. Ehrenberg was not selected due to European location, while Gibson, Heck, and Watanabe were not selected because their research groups already had representation among the selected experts.

Table III. The Top Five Tiers of Nominated Experts Grouped by Number of Peer Nominations

Melvin Andersen	U.S. Air Force
James Swenberg	Chemical Industry Institute of Toxicology
Thomas Starr	Chemical Industry Institute of Toxicology
Steve Tannenbaum	Massachusetts Institute of Technology
Hans Neumann	University of Wurzburg
Frederica Perera	Columbia University
Richard Reitz	Dow Chemical Company
Fred Beland	National Center for Toxicological Research
Robert Dedrick	National Institutes of Health
Lars Ehrenberg	University of Stockholm
James Gibson	Chemical Industry Institute of Toxicology
Henry Heck	Chemical Industry Institute of Toxicology
Miriam Poirier	National Cancer Institute
Phil Watanabe	Dow Chemical Company

Table IV. Participating Experts in Alphabetical Order

Dr. Melvin E. Andersen	United States Air Force
Dr. Fred Beland	National Center for Toxicological Research
Dr. Robert L. Dedrick	National Institutes of Health
Dr. Frederica Perera	Columbia University School of Public Health
Dr. Miriam Poirier	National Cancer Institute
Dr. Richard Reitz	Dow Chemical Company
Dr. James Swenberg	Chemical Industry Institute of Toxicology
Dr. Steven Tannenbaum	Massachusetts Institute of Technology

Table V. EPA Expert Panel for Review of Pharmacokinetic Data for Formaldehyde

Dr. Edward Bresnick	University of Nebraska
Dr. Peter Bloomfield	North Carolina State University
Dr. Richard Cornell	University of Michigan
Dr. Helen Evans	Case Western Reserve University
Dr. Robert Hanzlick	University of Kansas
Dr. Christopher Wilkinson	Cornell University
Dr. Lemone Yielding	University of South Alabama

We recognize that the definition of a "research group" is somewhat arbitrary and that the exclusions designed to achieve institutional balance and accommodate only domestic travel may have diminished to some extent the relevant expertise in the final group.

Once the eight experts were selected, each was contacted by letter and invited to participate. Much to our surprise, all agreed to be interviewed. Possible reasons for the 100% acceptance rate include: no travel was required of the experts; the interviews were to be personal and anonymous; no consensus was to be developed, and no writing on the part of experts was required; the approach used to select them was explicitly presented as well as a listing of all experts to be interviewed; a $500 stipend for participation was made available for those experts who wished to receive it; and formaldehyde is a scientifically interesting chemical of current regulatory concern.

For comparison, the expert panel assembled by EPA through an informal method for the purpose of reviewing these same data is given in Table V. There is no overlap between the two groups of experts. In terms of the several criteria outlined earlier, the formal approach appears superior because it is more explicit and open in how and why experts are selected and has the potential for being reproduced. Moreover, the power of the sponsor is much lower in our approach than in the informal approach.

3.2. Elicitation Strategy

Elicitation of expert judgment can proceed by several methods: personal interviews, groups or panels, or a combination thereof. When expert panels are convened to review scientific data, they are often asked to develop a consensus about the quality and suitability of data for use in the regulatory environment. In this framework, dominant personalities can guide the formation of consensus, and opinions of less vocal members may not emerge. One version of this group approach, the Delphi method, entails a gradual revision of written opinion until all members of a group can agree to a final consensus statement.[28]

The personal interview approach, which was selected for this study, attempts to preserve the unique perspective of each expert. Preservation of individual opinion was considered to be very important so that a full range of expert judgments could be reported. This approach accommodates consensus where it emerges but does not strive to foster or manufacture it. Personal interviews also conform to the Bayesian, decision-analytic tradition of personal degrees of belief being a valid measure of plausibility of events and true states of nature.[29]

3.2.1. Application of the Personal Interview Technique

Prior to the interviews, correspondence with each expert was conducted. This correspondence served several purposes: to answer questions about the project, to build credibility with each expert, and to provide background materials preparatory for the interviews. A memo with illustrative questions was also included to provide a common foundation for each expert interview. Experts reported that they spent from a half day to three working days reviewing the packet of materials.

A uniform protocol (set of written questions and summary of experiments) was devised prior to the interviews in accordance with a framework designed to permit analysis and comparison of the eight expert judgments. The questions were nonetheless somewhat open-ended to allow for interaction between the interviewer and the expert. The interview protocol was pretested on several in-house scientists and then revised. The purpose of the questions was to draw attention both to general issues in risk assessment and pharmacokinetics and to specific concerns about the delivered-dose data for formaldehyde. The interview protocol contained several sections: an overall description of the project, general questions on pharmacokinetic issues and research directions in chemical carcinogenesis, general questions on current cancer risk assessment practice, and questions about the delivered-dose data on formaldehyde. The questions on the delivered-dose data had two components: qualitative questions about the validity of the experimental procedure and toxicological assumptions, and

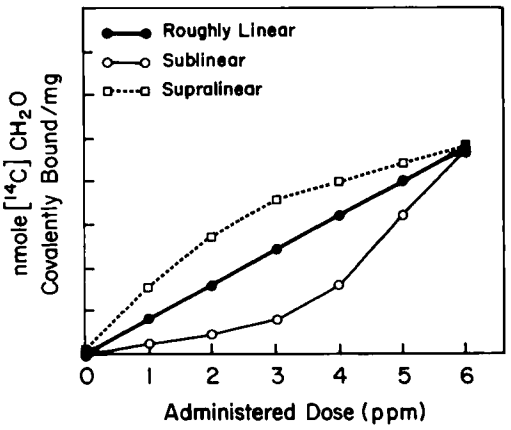

Fig. 2. Hypothetical relationships between administered and delivered dose.

an elicitation of judgmental probabilities about the true relationship between delivered and administered doses. The interview protocol is available upon request from the authors.

The hypothetical states of nature presented in the protocol are summarized in Fig. 2. The Casanova-Schmitz *et al.* interpretation of the data led them to assert that the true relationship between administered and delivered doses at low doses was sublinear, in contrast to the linearity typically assumed by agency risk assessors. The three possible states of nature are roughly linear (agency assumption), sublinear (the Casanova-Schmitz *et al.* hypothesis), and supralinear (greater than linear expectation).

In the subjective probability section, the experts were asked to assume that covalently bound formaldehyde was the measure of delivered dose. They were then asked to give prior probabilities for the three shapes (states of nature) based only on their knowledge and intuition. Then, they were asked for posterior probabilities taking into account the delivered-dose results. The change in prior and posterior probabilities essentially gives a measure of the internal validity of the experiment (i.e., how well did Casanova-Schmitz *et al.* measure covalently bound formaldehyde?). We recognize that the "prior" probabilities are somewhat contrived because they were elicited after publication of the delivered dose data.

The interview protocol closed with a discussion of the relevance of the Casanova-Schmitz *et al.* data for extrapolation to delivered doses in humans.

3.2.2. Interview Process

All interviews were performed by Neil Hawkins at the offices of the experts and lasted three hours on average. Some experts were reticent at the outset of the interviews, but all seemed comfortable by the end. After the interviews, a verification letter was sent to each expert regarding the prior and posterior probabilities provided during the interview. One expert wished to make a slight change in probabilities during this verification step and that change is reflected in the reported results. The interviewer made no attempt to play "devil's advocate" with the experts, although some discussions led naturally to commentary on the plausibility of alternative perspectives. The interviewer tried to understand each expert's particular judgments about formaldehyde within the context of his or her broader beliefs about risk assessment and carcinogenesis.

4. RESULTS OF THE PILOT STUDY

In keeping with the goal of preserving the distinctive judgments of all eight experts, each expert's views are summarized below for several key questions. Their prior and posterior probabilities are also reported separately. No attempt has been made to "weight" the experts. Each expert has been randomly assigned a letter (A to H) to ensure anonymity of opinions. This procedure was adopted to facilitate frank expression of beliefs.

The qualitative protocol questions have been summarized into three main topics that seem to cover the fundamental issues: evaluation of the delivered-dose research performed by Casanova-Schmitz *et al.*; consideration of how well the acute experiments simulate the chronic exposures faced by animals in the original bioassay; and the problems with extrapolating these data to humans.

4.1. Evaluative Comments on CIIT's Delivered-Dose Data

In this section we summarize and paraphrase each expert's overall evaluation of the delivered-dose research performed by Casanova-Schmitz *et al.*

Expert A was troubled by measurement of radioactive incorporation rather than measurement of a defined chemical entity. Expert A felt that the

analysis presented certain logical yet untested assumptions, but that direct chemical evidence of cross-link formation was needed. Expert A felt that the research provided evidence of uptake and distribution effects, which was useful, but in addition wanted to see more chemistry research directly proving the cross-link and its biological significance. Measurement of specific DNA adducts would be ideal. *Expert B* said that the principal hypothesis of sublinearity would be believable if and only if some direct chemical analysis were performed. Expert B also felt troubled that the basis for nonlinearity consisted of one data point, in that expert's opinion (see 2.0 ppm in Fig. 1). *Expert C* felt that radioisotope incorporation was an inadequate method, not based in chemistry, and that an elegant mathematical analysis was wasted on shaky data. Expert C had no confidence that cross-links had been measured and would never use these particular delivered-dose data for risk assessment. *Expert D* was also bothered that the claim of nonlinearity was primarily based on the 2.0 ppm point in that expert's opinion. Expert D thought that a heroic yet plausible effort in logic was made but that the empirical results were overinterpreted. Expert D would prefer a direct measure of covalently bound formaldehyde and more time-dependent data. *Expert E* felt that the existence of cross-links was proven to his or her satisfaction by two or three lines of evidence. The methods used were state-of-the-art for formaldehyde in 1983. *Expert F* thought it was a sensible first step that required much follow-up in chemistry research. Expert F felt that the work was too primitive to be used in risk assessment since no direct chemical evidence was provided. *Expert G* felt the finding of nonlinearity was an unusual and real phenomenon and that the work on glutathione depletion and isotope effect significantly strengthened the claim of nonlinearity. This work challenges the traditional thinking on linearity between delivered and administered doses. *Expert H* had no doubt that the experiment measured covalently bound formaldehyde. Covalent binding is the lowest level of evidence for delivered dose since it does not provide direct chemical evidence; however, this is pathbreaking work for formaldehyde due to its complex biochemistry. The nonlinearity is a real effect, which has been repeated in independent experiments.

All experts agreed that covalent binding of formaldehyde to DNA had theoretical plausibility for playing a role in formaldehyde carcinogenesis. Also, all experts agreed that formaldehyde poses a particularly difficult measurement problem due to its metabolism. Radioactivity incorporation may have been the only viable approach available in 1983 when the work was initiated. Some felt it is still the only method possible today, while others insisted that alternative methods could and should be explored (e.g., P-32 postlabeling and direct chemical analysis with an instrument such as a mass spectrometer).

4.2. Acute Versus Chronic Exposures

This section considers the applicability of the Casanova-Schmitz *et al.* delivered-dose data to chronic exposure conditions. Since most human exposures of concern have a chronic component, it is important to consider the applicability of these data to this exposure scenario.

Expert A had little confidence that the acute exposures relate to chronic delivered doses because acute exposure does not address the long-term effect of exposure on biological processes such as defenses. *Expert B* believes that the pulse radioactivity incorporation experiments had little relevance to the chronic case since a plateau in binding is usually observed with other substances. *Expert C* had no confidence that acute delivered-dose models the behavior of chronic delivered dose with any degree of precision. *Expert D* thought that the acute data give no information about time scale, repair, and chronic effects. *Expert E* felt that the Casanova-Schmitz *et al.* research was a good first approximation to the chronic case and that the next step was to do the chronic experiments and remove the residual uncertainty. *Expert F* felt that this chronic extrapolation was the main downfall of this study, in that the chronic extrapolation ignored likely differences in cell proliferation, DNA access, formaldehyde metabolism, and adduct repair between acute and chronic exposure situations. *Expert G* felt that the data should be used in risk assessment just as outlined by Starr and Buck, but the actual chronic effect would not be known with confidence until such experiments are actually performed. *Expert H* had confidence in the extrapolation of these data to chronic exposure in the low-dose range, citing short repair half-times in the low-dose range and steady-state conditions within two or three days of exposure.

4.3. Relevance of the CIIT Data to Humans

Experts A, B, C, D, E, and G wanted more comparative physiology information before extrapolating from the rat model to humans. *Expert F* was concerned that results on a limited number of animals exposed for six hours might be used to model effects of chronic human exposures in light of variations in human susceptibility, possible interactions of formaldehyde with other chemicals, and physiological differences. *Expert H* felt that cancer mechanisms in humans would be the same but that the location of target sites may be different, and that DNA would be a relevant target in both animals and humans.

4.4. Probabilistic Characterizations of Judgments

In order to facilitate a concise comparison of expert judgments, we elicited subjective probabilities from each expert about the true relationship between administered and delivered doses of formaldehyde in rats. Experts were asked in particular to consider whether their subjective probabilities are influenced by the delivered-dose data.

The prior and posterior probabilities are reported in Table VI. The movement (or lack thereof) in prior and posterior probabilities of each expert seemed to correspond well with qualitative evaluations of the data. The probabilities revealed several patterns. Four of the eight experts (B, C, F, H) did not have their probabilities moved: Three of them (B, C, F) because they felt the data were not internally valid and the fourth (H) because the expert was firmly convinced that the relationship was sublinear to begin with and that the data simply confirmed the prior belief. The other four experts' probabilities (A, D, E, G) were moved by varying degrees by the data. Posterior probabilities indicate that four of the eight experts (B, C, D, F) still believe that the traditional assumption of linearity may be most appropriate in this case. The other four experts (A, E, G, H) put more probability on the sublinear category, which corresponds with the Casanova-Schmitz *et al.* hypothesis. The results suggest that among these eight experts the biologically plausible assumptions range from roughly linear to sublinear. Since only eight experts were interviewed, little can be said about the statistical distribution of the population of expert beliefs.

Table VI. Prior and Posterior Probabilities (for the Relationship Between Administered and Delivered Dose)[a]

Person	Supralinear	Roughly Linear	Sublinear
Expert A			
Prior	0.00	0.50	0.50
Posterior	0.00	0.30	0.70
Expert B			
Prior	0.05	0.80	0.15
Posterior	0.05	0.80	0.15
Expert C			
Prior	0.33	0.33	0.33
Posterior	0.33	0.33	0.33
Expert D			
Prior	0.15	0.60	0.25
Posterior	0.10	0.50	0.40
Expert E			
Prior	0.00	0.67	0.33
Posterior	0.00	0.05	0.95
Expert F			
Prior	< 0.25	> 0.50	< 0.25
Posterior	< 0.25	> 0.50	< 0.25
Expert G			
Prior	0.10	0.30	0.60
Posterior	0.00	0.10	0.90
Expert H			
Prior	0.00	0.00	1.0
Posterior	0.00	0.00	1.0

[a] Delivered dose = formaldehyde covalently bound to nasal mucosal DNA.

4.5. Implications for Formaldehyde Risk Assessment

The qualitative and quantitative results of expert interviews have been presented, and in the real world, a risk assessor would be forced to decide whether and how the data should be used in a quantitative risk assessment. In light of the responses, we believe it would be inappropriate to abandon the traditional assumption because there is no scientific consensus in support of sublinearity. We also believe that it would be inappropriate to rely solely on the traditional linearity assumption since it is considered implausible by four of eight experts (A, E, G, H). It seems, therefore, most appropriate to report two risk characterizations: one based on the traditional assumption and the other with the modification using the Casanova-Schmitz *et al.* data as a basis to specify a nonlinear relationship between administered and delivered doses.

Reporting of both assessments is appropriate because of the qualitative and quantitative schism in expert scientific opinion about these data. In cancer

risk assessments, both hypotheses about delivered dose (and implicitly those in between) should be considered to have biological plausibility. The range would not be arbitrary since it is expert judgment that provides a basis for expressing a range of uncertainty about the risk.

Armed with an extramural review of the data, an administrator of a regulatory agency would have objectively gathered opinions of experts to guide his or her risk assessors. The administrator can then decide in an informed way which risk characterization to rely upon and also how much confidence expert scientists have in the alternative characterizations. Selection of a single characterization is not necessarily required for decision purposes. If a standard must be set to protect the public health, it must be set in light of the risk characterizations and other authorized statutory factors. Although expression of risk as a plausible range is not wholly satisfactory, it does explicitly and honestly place the burden of decision about uncertainty where it should be: on the risk manager. The scientific community will have done its part by opining on the data and expressing its views on the biological uncertainty of the risk assessment.

4.6. Prospects of Convergence of Judgments with New Evidence

A pervasive theme in the qualitative comments of the experts was that direct chemical evidence and analysis are needed to characterize the radiolabeled entity accurately. It seemed clear to us that if this line of research were undertaken and performed competently, the judgments of the eight experts would not remain as polarized as they currently are and might converge radically, regardless of which hypothesis the new results supported.

We were encouraged to learn that CIIT scientists are already on this trail. We believe, however, that it would be desirable for a non-CIIT group to perform some of the future work as well. A similar result by CIIT and non-CIIT scientists would probably do more to create scientific consensus than would work by CIIT scientists alone. Ultimately, some uncertainty will remain until direct delivered-dose data on chronic exposures are available and some credible procedure for predicting delivered doses in humans is devised.

5. GENERAL IMPLICATIONS FOR RISK ASSESSMENT

Several conclusions and inferences can be drawn from this pilot study. The most obvious is that expert scientific judgment offers a feasible and promising avenue to assist cancer risk assessment practice in the short run. Scientific judgment can guide risk assessors on when to use pharmacokinetic data, how to use them in characterizing risk, and how to describe their strengths, limitations, and uncertainties. When scientific data and judgment provide a plausible basis for challenging the standard assumptions, risk assessors then have a basis for reporting a nonarbitrary range of risk characterizations, moving away from exclusive use of an upper bound potency estimate. Only a breakthrough in pharmacokinetic research that generates scientific consensus in favor of an alternative hypothesis is likely to justify complete abandonment of the standard delivered-dose assumption for a given chemical.

While more reliance on scientific judgment may be desirable, further research is also needed to develop methods for selecting appropriate experts and incorporating their judgments into risk assessments. The personal interview approach preserves opinions of individual experts rather than seeking a consensus, which is preferable for expressing biological uncertainty. Also, the experts in the study were, for the most part, comfortable providing subjective probabilities; therefore, Bayesian approaches deserve further investigation as one means for elicitng and conveying scientific judgment. Formal approaches to expert selection seem superior to an informal approach in terms of criteria already defined, and therefore, these methods deserve additional attention.

One important strength of using expert judgment is that the conservatism of the standard approach need not be forgone. The advantages of the standard approach can be maintained while at the same time a mechanism would become available for incorporating new biological hypotheses and estimating ranges of risk based on them. As demonstrated in this pilot project, expert scientific judgment can be incorporated into cancer risk assessment inexpensively, rapidly, openly, and credibly.

ACKNOWLEDGMENTS

This work was funded by the Risk Science Institute, the Richard King Mellon Foundation, and the

Alfred P. Sloan Foundation. Dr. Hawkins also received support from the Occupational Safety and Health Educational Resource Center NIOSH grant award No. 5 T15 OHO7096 at the Harvard School of Public Health. Dr. Susan Kennedy and Mr. Scott Wolff provided essential technical support during initial stages of this research. The authors also thank an informal advisory committee of colleagues consisting of Dr. John Bailar, Dr. Joseph Brain, and Dr. John Evans. Most importantly, the authors thank the eight experts for their candidness in interviews and for their exceptional cooperation. All experts had the opportunity to review this manuscript and specific views attributed to them in the text, and any suggested change was fully incorporated; however, all general opinions about use of expert judgment in risk assessment are strictly those of the authors.

REFERENCES

1. Industrial Union Department versus American Petroleum Institute, 448 U.S. (1980).
2. E. L. Anderson and the Carcinogen Assessment Group of the U.S. Environmental Protection Agency, "Quantitative Approaches in Use to Assess Cancer Risk," *Risk Analysis* **3**, 277–295 (1983).
3. U.S. E.P.A., "Guidelines for Carcinogen Risk Assessment," *Federal Register* **51**, 33991 (1986).
4. J. D. Graham, N. C. Hawkins, and M. J. Roberts, "Expert Scientific Judgment in Quantitative Risk Assessment," Prepared for the Banbury Center Conference on New Directions in the Qualitative and Quantitative Aspects of Carcinogen Risk Assessment, Cold Spring Harbor Laboratory, New York, October 11–14, 1987, Conference Proceedings in press.
5. W. D. Kerns *et al.*, "Carcinogenicity of Formaldehyde in Rats and Mice After Long-Term Inhalation Exposure," *Cancer Research* **43**, 4382–4392 (1983).
6. K. G. Brown, "Risk Assessment of Laboratory Rats and Mice Chronically Exposed to Formaldehyde Vapors," *Risk Analysis* **5**, 171 (1985).
7. P. W. Crockett and K. S. Crump, "Quantitative Assessment of Carcinogenic Risk from Formaldehyde Based Upon Time-to-Response Data from the CIIT Study of Long-Term Exposure in Rats," Report to EPA, 1984.
8. R. L. Sielken, "A Quantitative Look at the Low-Dose Cancer Risk in Rats Due to Formaldehyde Inhalation," submitted to OSHA, January, 1984.
9. Consumer Product Safety Commission, "Revised Carcinogenic Risk Assessment for Urea Formaldehyde Foam Insulation: Estimates of Cancer Risk Due to Inhalation of Formaldehyde Released by UFFI," 1981.
10. Consumer Product Safety Commission, "Formaldehyde Carcinogenic Risk Assessment: Exposure in Manufactured Housing," Report submitted to Department of Housing and Urban Development, 1984.
11. Occupational Safety and Health Administration, "Preliminary Assessment of the Health Effects of Formaldehyde," Office of Risk Assessment, November 5, 1984.
12. Environmental Protection Agency, Memorandum from Gary Grindstaff to Rich Hefter, "Quantitative Cancer Risk Assessment: Formaldehyde," Office of Pesticides and Toxic Substances, March 27, 1984.
13. Environmental Protection Agency, "Preliminary Risk Assessment of Health Risks to Garment Workers and Certain Home Residents from Exposure to Formaldehyde," Office of Toxic Substances, draft, May, 1985.
14. M. Casanova-Schmitz, T. B. Starr, and H. d'A. Heck, "Differentiation Between Metabolic Incorporation and Covalent Binding in the Labeling of Macromolecules in the Rat Nasal Mucosa and Bone Marrow by Inhaled Carbon-14 and Tritiated Formaldehyde," *Toxicology and Applied Pharmacology* **76**, 26–44 (1984).
15. J. A. Boucheron, F. C. Richardson, P. H. Morgan, and J. A. Swenberg, "Molecular Dosimetry of O^4-Ethyldeoxythymidine in Rats Continuously Exposed to Diethylnitrosamine," *Cancer Research* **47**, 1577 (1987).
16. T. B. Starr and R. D. Buck, "The Importance of Delivered Dose in Estimating Low-Dose Cancer Risk from Inhalation Exposure to Formaldehyde," *Fundamental and Applied Toxicology* **4**, 740–753 (1984).
17. M. S. Cohn, F. J. DiCarlo, A. Turturro, and A. G. Ulsamer, "Letter to the Editor," and M. Casanova-Schmitz, T. B. Starr, and H. d'A. Heck. "Reply," *Toxicology and Applied Pharmacology* **77**, 363–368 (1985).
18. M. Casanova and H. d'A. Heck, "Further Studies of the Metabolic Incorporation and Covalent Binding of Inhaled Tritiated and Carbon-14 Formaldehyde in Fischer-344 Rats: Effects of Glutathione Depletion," *Toxicology and Applied Pharmacology* **89**, 105–121 (1987).
19. H. d'A. Heck and M. Casanova, "Isotope Effects and Their Implications for the Covalent Binding of Inhaled Tritiated and Carbon-14 Formaldehyde in the Rat Nasal Mucosa," *Toxicology and Applied Pharmacology* **89**, 122–134 (1987).
20. Life Systems, "Expert Review of Pharmacokinetic Data: Formaldehyde," Final Evaluation Report submitted to U.S. E.P.A., contract No. 68-02-4228, January 2, 1986.
21. M. Casanova, T. B. Starr, and H. d'A. Heck. "Comments on the Final Report of the Panel Reviewing the CIIT Pharmacokinetic Data on Formaldehyde," submitted by CIIT to U.S. E.P.A., February 4, 1986.
22. Environmental Protection Agency, "Assessment of Health Risks to Garment Workers and Certain Home Residents from Exposure to Formaldehyde," Office of Pesticides and Toxic Substances, April, 1987.
23. *TOXINE, Toxicology Information on Line (Database)* (National Library of Medicine, Bethesda, Maryland, 1987).
24. *MEDLINE, Medical Literature Analysis and Retrieval System (Database)* (National Library of Medicine, Bethesda, Maryland, 1987).
25. *SCISEARCH, Science Citation Index (Database)* (Institute for Scientific Information, Philadelphia, 1987).
26. J. C. Press (ed.), *American Men and Women of Science,* Physical Sciences and Biology, 16th edition (R. R. Bowker Company, New York, 1986).
27. *National Faculty Directory,* 17th edition (Gale Research Company, Detroit, 1986).
28. H. A. Linston and M. Turoff (eds.), *The Delphi Method — Techniques and Applications* (Addison-Wesley, Reading, Massachusetts, 1975).
29. H. Raiffa, *Decision Analysis* (Addison-Wesley, Reading, Massachusetts, 1968).