TOXICOLOGICAL SCIENCES **61,** 18–31 (2001)

FORUM

Harmonization of Cancer and Noncancer Risk Assessment: Proceedings of a Consensus-Building Workshop

Matthew S. Bogdanffy,*[,1] George Daston,† Elaine M. Faustman,‡ Carole A. Kimmel,§ Gary L. Kimmel,§ Jennifer Seed,¶ and Vanessa Vu§

**DuPont Haskell Laboratory for Toxicology and Industrial Medicine, P. O. Box 50, Newark, Delaware 19714; †Procter and Gamble, Miami Valley Laboratories, P. O. Box 538707, Cincinnati, Ohio 45253-8707; ‡University of Washington, Department of Environmental Health, 4225 Roosevelt Avenue, Suite 100, Seattle, Washington 98105–6099; and §National Center for Environmental Assessment, and ¶Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency, Ariel Rios Building, 1200 Pennsylvania Ave, NW, Washington, DC 20460*

Received September 29, 2000; accepted January 3, 2001

Significant advancements have been made toward the use of all relevant scientific information in health risk assessments. This principle has been set forth in risk-assessment guidance documents of international agencies including those of the World Health Organization's International Programme on Chemical Safety, the U.S. Environmental Protection Agency, and Health Canada. Improving the scientific basis of risk assessment is a leading strategic goal of the Society of Toxicology. In recent years, there has been a plethora of mechanistic research on modes of chemical toxicity that establishes mechanistic links between noncancer responses to toxic agents and subsequent overt manifestations of toxicity such as cancer. The research suggests that differences in approaches to assessing risk of cancer and noncancer toxicity need to be resolved and a common broad paradigm for dose-response assessments developed for all toxicity endpoints. In November 1999, a workshop entitled "Harmonization of Cancer and Noncancer Risk Assessment" was held to discuss the most critical issues involved in developing a more consistent and unified approach to risk assessment for all endpoints. Invited participants from government, industry, and academia discussed focus questions in the areas of mode of action as the basis for harmonization, common levels of adverse effect across toxicities for use in dose-response assessments, and scaling and uncertainty factors. This report summarizes the results of those discussions. There was broad agreement, albeit not unanimous, that current science supports the development of a harmonized set of principles that guide risk assessments for all toxic endpoints. There was an acceptance among the participants that understanding the mode of action of a chemical is ultimately critical for nondefault risk assessments, that common modes of action for different toxicities can be defined, and that our approach to assessing toxicity should be biologically consistent.

Key Words: **risk assessment; harmonization; cancer; noncancer; mode of action; dose response; uncertainty factors; interspecies extrapolation.**

Introduction

Significant advances have been made in the use of all relevant biological information in the risk-assessment process. For example, EPA's proposed revisions to the carcinogen risk-assessment guideline have emphasized the use of mode-of-action information in the characterization of hazard and selection of the dose-response approach to estimate risk at low exposure levels (U.S. EPA, 1996). EPA has proposed to use the margin-of-exposure approach as a default method for carcinogenic agents that are judged to have a mode(s) of action likely to exhibit a nonlinear dose-response curve at low doses. This approach is similar to that in use by Health Canada (Meek *et al.*, 1994).

EPA's proposed carcinogen risk-assessment guidelines are precisely in line with one of the major goals in the Society of Toxicology's long-range plan to improve the scientific basis of risk assessment. However, the publication of these guidelines highlights the significant differences that have evolved between the assessments of risks for developing cancer versus that for any other manifestation of toxicity, even when they all arise from the same mode of action. Current science, which establishes mechanistic links between noncancer responses to toxic agents and subsequent overt manifestations of toxicity such as cancer, suggests that these differences need to be resolved and a common broad paradigm for dose-response assessments developed for all toxicity endpoints.

Separate approaches to the assessment of cancer and noncancer health risks can be traced back to origins of the cancer risk-assessment guideline in which the process of chemical carcinogenesis was thought to be similar to that of radiation carcinogenesis (IRLG, 1979). That is, risk assessment for chemical carcinogens was based on the assumption that any exposure carries with it a risk of cancer. For noncancer risk assessment, Lehman and Fitzhugh (1954) applied the more traditional toxicology principle of dose thresholds. In the past

[1] To whom correspondence should be addressed. Fax: (302) 366-5003. E-mail: matthew.s.bogdanffy@usa.dupont.com.

TABLE 1
Definitions of Mode of Action

Group	Definition
American Industrial Health Council	A mode of action is a category or class of toxic mechanisms for which the major (but not all) biochemical steps are understood (Schlosser and Bogdanffy, 1999).
United States Environmental Protection Agency	A mode of action is composed of key events and processes starting from the interaction of an agent with a cell, and through operational and anatomical changes, resulting in cancer formation. "Mode" is contrasted to "mechanism" of action, which implies a more detailed molecular description of events than is meant by mode of action (U.S. EPA, 1999).
World Health Organization-International Programme on Chemical Safety	A supported mode of action would have sufficient evidence to establish a biologically plausible explanation. Mechanism of action, in contrast, relates to a rigorous proof of causality (IPCS, 1999).
Cancer-Noncancer Risk Assessment Methods Harmonization Workshop	Mode of action is a series of key events supported by a body of scientific knowledge that provide a biologically plausible explanation of causality for a given toxic effect within a context of dose and duration of exposure and susceptibility of target tissues. In contrast, "mechanism" of action refers to a complete understanding and demonstration of all biological steps leading to toxicity.

20 years, however, there has been an explosion of basic research into molecular mechanisms of toxicity. Much of this work has identified many common toxicological responses following exposure to carcinogens and noncarcinogens, and these findings beg the question of whether distinctly different philosophical approaches to cancer and noncancer risk assessments are appropriate (Butterworth and Bogdanffy, 1999). Furthermore, many of these mechanisms are likely not to be unique to a particular manifestation of toxicity, but rather parts of the pathogenic pathway that results in a number of toxic responses. These responses may depend on a number of factors, such as the type of exposure, the sensitivity of the individual, etc.

The goal of the workshop reported here was to provide a forum for the exchange of scientific views on the most critical risk assessment issues involved in developing a more consistent and unified approach to risk assessment for all toxic endpoints. The intent of the workshop was to build consensus where it could be achieved and to identify the range of opinions for those areas where consensus was not yet possible. The aim of this manuscript is to summarize these discussions. Thus, the content reflects the discussions of the workshop and individual breakout groups and does not necessarily reflect the views of any individual author.

Meeting Participants and Charge to the Participants

The workshop participants included those attending by invitation of the steering committee and other registered observers. Invited participants (see Acknowledgments) were selected with two primary objectives in mind: engage those scientists most active in the field, and strive for balance in the representation of government, academic, and industry scientists. Invited participants were allowed full participation in the discussions of the focus questions; observers were provided the opportunity to participate as time allowed.

Prior to the meeting, the participants received background material and a list of focus questions that were used to guide discussions of the breakout groups and the plenary sessions. Also included were three examples of case studies and relevant scientific literature.[2] These studies were used as resources for the discussions. The case studies included information on mammalian toxicology and risk assessments for ethylene oxide, ethylene thiourea, and trichloroethylene, and were chosen to provide a range of toxic responses and potential modes of action for the discussions. The focus questions assigned to each group were developed by the steering committee and are presented below in the individual breakout group reports.

Following an introductory plenary session in which Dr. Vu (U.S. EPA) presented an overview of various international definitions of the term "mode of action" (Table 1), Dr. Conolly (Chemical Industry Institute of Toxicology) provided his perspectives on an integrated approach to risk assessment for cancer and noncancer endpoints (Conolly, 1995). The participants then convened in their assigned breakout groups and began discussions of the charge questions.

Report of Breakout Group 1: Mode of Action as the Basis for Harmonization

This breakout group felt that, prior to discussion of the focus questions, it was necessary to define the term "mode of action" and to develop a common understanding of its meaning. The group reviewed the various definitions of mode-of-action developed by the American Industrial Health Council (AIHC) (Schlosser and Bogdanffy, 1999), the U.S. Environmental Protection Agency (U.S. EPA, 1996), and the International Programme on Chemical Safety (IPCS, 1999) (Table 1).

Group 1 adopted the following working definition, which is applicable to all toxic manifestations:

Mode of action is a series of key events supported by a body of scientific knowledge that provide a biologically plausible explanation of causality for a given toxic effect within a

[2] The case studies are available at www.toxicology.org or by written request to the corresponding author.

context of dose and duration of exposure and susceptibility of target tissues. In contrast, "mechanism" of action refers to a complete understanding and demonstration of all biological steps leading to toxicity.

Question 1: Are There Similar Modes of Action Identified or Suspected for a Variety of Toxic Manifestations?

In light of current knowledge and given the working definition of mode of action as described above, the Breakout Group 2 generally agreed that there are common modes of action for different toxicities. Examples of such modes of action are cytotoxicity, mutagenesis, endocrine modulation, and immune suppression. However, the group was not comfortable with talking in generalities and felt a need for linking mode of action with dose and duration of exposure, and with levels of response (molecular, cellular, tissue cell, physiological).

The group also reviewed the IPCS framework for analyzing and evaluating a postulated mode of action for cancer risk assessment (IPCS, 1999). The IPCS framework is a tool providing a structured approach to assessing the overall weight of evidence for a postulated mode of action of a carcinogenic agent. The framework was developed by an international review group and was built on concepts discussed in the revised EPA cancer guidelines (U.S. EPA, 1999). The framework utilized a modification of the Bradford Hill criteria (Hill, 1965) for causality for human epidemiological studies that had been modified for noncancer endpoints (Faustman *et al.*, 1996). The framework begins with a summary description of the postulated mode of action and is followed by an evaluation of available data pertaining to the following issues or topics:

- description of the postulated mode of action;
- identification of measurable key events that are critical to induction of tumors, as hypothesized in the postulated mode of action;
- dose-response relationships of key events and tumors;
- temporal relationships among key events and tumors;
- strength, consistency, and specificity of association of tumors with key events;
- biological plausibility and coherence of the postulated mode of action in light of current knowledge;
- discussion of other modes of action that are supported by available data; and
- a conclusion about whether the postulated mode of action is supported along with a description of uncertainties, inconsistencies, and data gaps.

The framework is designed to bring transparency to the analysis of a postulated mode of action and thereby promote confidence in the conclusions reached; it is not designed to provide criteria for what constitutes sufficient evidence to establish a particular postulated mode of action.

Group 1 generally endorsed the utility of the IPCS analytical framework. It was recommended, however, that the section on dose-response relationship of the framework should be expanded to include other toxicity endpoints besides cancer, and an explicit and careful evaluation of the dose-response relationships for key events, precursor lesions, and more frank manifestations of toxicity. The discussion should include a qualitative evaluation of the effect and the shape of dose-response curves.

Questions 2 and 3: How Much Evidence Is Needed to Show That a Substance Acts Via a Particular Mode/Mechanism? How Much Evidence Is Needed to Show that Two Toxic Manifestations Caused by the Same Substance Were Produced by Different Modes of Action?

Overall, this breakout group expressed great difficulty in prescribing how much information or evidence is needed, both to establish a plausible mode of action and to judge whether different modes of action are involved (i.e., data-sufficiency criteria). The group felt that they could cite examples for use of mode-of-action information to conclude qualitatively that a particular mode of action was not relevant to humans (e.g., renal tumors in male rats associated with the alpha-2u-globulin). However, more quantitative dose-response information is usually needed to make judgments that sufficient evidence is available to support the use of mode-of-action data in low-dose extrapolation. The group recommended the need for providing criteria for "sufficiency of evidence," although no specific criteria were presented. They felt that the use of a peer review process alone is not a sufficient means for accepting or rejecting a postulated mode of action. Peer review will be important for judging technical integrity of data presented to support a mode of action.

In order to facilitate the discussion of "sufficiency of evidence," the group reviewed the information provided as part of the workshop background materials on two of the case studies: ethylene thiourea (ETU) and ethylene oxide (ETO). With regard to ETU, the group focused the discussion on the three major organ targets: thyroid (rats more susceptible than mice), pituitary (mice only), and liver (mice only). The group concluded that there was sufficient evidence to support the postulated mode of action for ETU-induced thyroid effects in the rat (i.e., via inhibition of the enzyme thyroid peroxidase, resulting in decreased serum T3 and T4 levels, increased thyroid-stimulating hormone, thyroid follicular-cell hyperplasia, and subsequently, thyroid follicular-cell adenomas and carcinoma). Thus, this group supported a threshold approach for both ETU-induced thyroid toxicity and carcinogenicity in the rat. However, they felt that there was not enough evidence provided in the case studies they reviewed to support a mode of action involving disruption of thyroid hormone homeostatsis (i.e., via antithyroid action) that was common to thyroid, liver, and pituitary tumors.

The discussion of the ETO case study focused on cancer and genetic, developmental, reproductive, and neurotoxic end-

points. The group concluded that although a common mode of action, DNA and protein alkylation, is plausible for cancer and noncancer endpoints, the sufficiency of evidence for this mode of action for all endpoints varies greatly. Also, it was felt that this particular mode of action (i.e., via DNA and protein alkylation) does not necessarily imply that the dose response is linear for cancer and noncancer endpoints, as these effects arise from many events subsequent to macromolecular binding.

Taken together, these discussions illustrated that, in establishing a mode of action for a given endpoint in a single tissue or organ system, it is inappropriate to assume the same mode of action for other endpoints or organ systems without sufficient data to bridge the mode of action between endpoint or organ systems.

In summary, Group 1 participants were able to make constructive modifications to the mode-of-action framework (adding additional emphasis on dose-response and *in vivo* contexts) and were able to apply these points in evaluating the mode-of-action data for two case studies. Although the committee was unable to prescribe a set of common criteria for "sufficiency of evidence" *a priori*, within the context of these specific case studies, the committee was able to identify when they felt comfortable with proposed modes of action within a given endpoint and across endpoints.

Question 4: What Should Be the Dose-Response Approach for a Chemical that Produces Multiple Manifestations, but through a Similar Mode of Action?

The group generally supported the use of similar dose response approaches for low dose extrapolation regardless of endpoints (i.e., cancer or noncancer endpoints) when there is sufficient evidence for a common mode of action. The group had considerable discussion on EPA's current default approaches for low-dose extrapolation, which stipulate that in the absence of knowledge in support of a particular mode of action, nonthreshold linear approaches are to be used for carcinogenic effects and threshold approaches for noncancer endpoints. In response to Question 4, more than half of the group felt that in light of current scientific knowledge, it is a plausible default that both carcinogenic and noncarcinogenic responses follow biological threshold-like responses. Thus, a majority of the participants held the view that there should be no difference between genotoxic and nongenotoxic carcinogens with regard to thresholds or points of departure for low-dose considerations. Some expressed the view that all responses should follow nonthreshold responses, while other felt that nonthreshold responses should be considered for certain types of genotoxic mechanisms. In summary, the Breakout Group 1 discussions concluded that more options are needed for evaluating dose-response relationships. These considerations should not be driven by a distinction of cancer versus noncancer but rather by considerations of mode of action.

Report of Breakout Group 2: Common Levels of Adverse Effect across Toxicities for Use in Dose-Response Assessment

Question 1: Should Statistical Power or Degree of Change Be the Basis for Defining Adverse Effect Levels?

Although Breakout Group 2 did not reach a consensus answer to this question, there was productive discussion of many related issues and several potential solutions. At the beginning of the discussion, the group agreed that the goal of this question was not to define the term "adverse." They assumed, for the purposes of this question, that the "adverse" effect was previously defined. Rather, the group decided that this question was addressing the issue of the varying resolving powers (i.e., statistical robustness) of different toxicology study designs. Given this clarification, Group 2 decided to address the question with regard to quantal versus continuous endpoints. In addition, they discussed how knowledge of mode of action would influence such decisions.

Initially, the group focused on quantal endpoints. Standard toxicology studies have been designed to characterize the potential hazard for specific toxicities. As a consequence, the studies may have different sample sizes, and the power may vary between studies as well as among endpoints within a specific study. This can result in difficulties when one attempts to compare the results of various toxicology studies. The breakout group discussed two examples that are often encountered. The first example was how to compare a neuropathology evaluation with a sample size of 5, a subchronic toxicity study with a sample size of 10, a prenatal developmental toxicity study with a sample size of 20, and a cancer bioassay with a sample size of 50. Some of these studies can detect a 5% change in some endpoints, whereas others may only detect a 40% change. The second example discussed was the situation where two chemicals are being compared for a given endpoint. One of the chemicals has been examined in a large study with a 1% detection level, while the second chemical has been examined in a small study with a 30% detection level. Just by the nature of the study design for the two chemicals, one of them will appear to be more toxic than the other.

The group agreed that there is no immediate solution to the problem and offered several available options. One option is simply to acknowledge that there are differences in power among studies and to select a standard level of change like an ED_{10}, LED_{10}, or ED_{50}. However, caution needs to be exercised in the selection of the specific level of change. The group stressed that modeling should only be done within, or very near, the detectable range; modeling below that range is problematic. For this reason, selection of an ED_{10} would not be useful for studies with a resolving power of 30 or 50%. On the other hand, selection of an ED_{50} would not make full use of dose-response data for studies with statistical power to resolve a 5% change.

Epidemiology data were also discussed in the context of

statistical power and limits of detection. Using epidemiological data, it is possible to identify a BMD that is much smaller (e.g. $ED_{0.01}$), than that based on typical chronic bioassay data. There is, however, less extrapolation to environmentally relevant exposures. Proposed criteria for the selection of a BMD based on epidemiological data are as follows: BMD is within the observational range, and the BMD estimate is consistent across models.

A second option to providing consistency in dose-response assessments of various types of data used in risk assessment is to acknowledge the bounds of the studies and utilize all the available data in a sliding-scale approach. The sliding scale would acknowledge that the resolving power of different types of studies is not comparable. It would permit the risk assessor to use a starting point for risk estimation that is within the experimental dose range for the type of study being used, with that starting point being dependent on study design. However, because study designs are different, more consideration of a study's resolving power would need to go into the consideration of the margin of exposure necessary to be protective. Criteria would need to be established for the sliding scale. The group acknowledged that one drawback to this approach is resistance by risk managers. Finally, the group stressed that in order to resolve this issue and make standard toxicology studies more comparable, some may need to be redesigned. Many current studies are designed for hazard identification, not dose-response assessment; they provide qualitative information, but not necessarily useful quantitative information.

Breakout Group 2 also discussed continuous endpoints and offered several options, one of which was to quantalize the data based on a justifiable cut-off point (Kavlock *et al*., 1995). For example, one could decide that changes of 10% or greater from control means are "adverse," and that changes of less than 10% are acceptable (Gaylor and Slikker, 1990). Another option is simply to model the continuous endpoint as a continuous variable. Finally, the group stressed that continuous endpoints should be viewed within the context of what is known about the mode of action and how a specific continuous endpoint may relate to a particular quantal endpoint. For example, a continuous endpoint may represent a precursor event to a particular quantal endpoint, and therefore provide information about the shape of the dose-response curve within a lower dose range. One would then have to apply knowledge about the biology and mode of action to determine a response level for the precursor that is considered acceptable.

Question 2: Should One Use a Consistent Response Level across Toxicities (e.g., 5%, 10%) for Comparing Adverse Effects; for Selecting the Point of Departure? How Does Mode of Action Impact the Point of Departure Selection?

The breakout group recognized several limitations in trying to use a consistent response level for comparing adverse effects, and cautioned that this may lead to similar problems encountered in the past when comparing NOAELs, RfDs, or cancer potencies. The breakout group stressed that some endpoints/studies should not be compared (e.g., 5% change in fetal body weight [continuous endpoint] vs. 5% increase in fetal death [dichotomous endpoint]) and that knowledge of the shape of the dose-response curve is requisite to making any comparison. Some study designs are useful for hazard identification but not well designed for quantitative risk assessment (e.g., the functional observation battery in the EPA neurotoxicity guideline). If the shape of the dose-response curve is not known, then the curves could cross, leading to errors in the comparison. In general, Group 2 did not recommend making such comparisons, but acknowledged that if this had to be done, it should be restricted to endpoints with similar response metrics, and should be based on a response level within the detectable limit of all the studies concerned (e.g., ED_{50}). The breakout group stressed that it is not appropriate to prescribe, *a priori*, a consistent response level as the point of departure for all endpoints of toxicity.

Knowledge of the mode of action can impact the point of departure selection. If there is a common mode of action for different toxicities associated with a chemical, then one should select the study with the greatest resolving power, or base the point of departure on a precursor event that would require support for the conclusion that the event is linked to an accepted adverse effect. Use of a common precursor effect would increase confidence that the assessment was protective for all toxicities. The breakout group also discussed how mode-of-action information could influence decisions about which toxicology studies should be conducted for a given chemical. Knowledge of mode of action could allow for limited testing on chemicals with an assumed common mode of action. However, the breakout group also stressed that knowledge of mode of action could lead to false conclusions about potential effects. For example, a chemical may exhibit weak estrogenic activity, but not be associated with clinically adverse effects.

Question 3: How Can Severity of Response within an Endpoint Be Considered in Dose-Response Assessment?

Modern toxicology studies involve numerous measurements of the structure and function of tissues, organs, and individuals. For many of these, there appear to be progressive, dose-related increases, not only in magnitude of response but also in the severity of the effect elicited. That is, at low-dose levels, effects may be limited to pretoxic effects such as altered clinical chemistry parameters, xenobiotic enzyme induction, or subtle histological changes, while larger doses evoke responses that are frankly adverse. As toxicity measurements become more sophisticated, it will become possible to determine that many of these effects are mechanistically related. Whereas more subtle changes at the molecular or biochemical level may be evident at low doses, toxic manifestations may become more severe and involve cellular or whole-organ levels of

biological organization at higher doses or longer duration of exposure.

It seems both a waste of information and an unreasonable approach to simply choose the most sensitive effect as the starting point for dose-response assessment, particularly given that studies designed to identify different toxic responses and conducted under different guidelines, or simply with less care, will likely be limited to the more fulminant manifestations of toxicity. Breakout Group 2 explored the available options for modeling effects of graded severity, and also issues around choosing the most sensitive effect as a point of departure.

It was noted that statistical models already exist that allow the use of all of the effects along a mechanistic continuum, from precursor events to frank toxicity. Categorical regression models were cited as having the longest history of use (Simpson *et al*., 1996). Other models have also been presented in the literature, such as one used to account for the relatively lesser severity of rib variations vs. frank rib malformations in rodents after oral exposure to high levels of boric acid (Allen *et al*., 1996). The group felt that the use of such models and development of biologically based models in the future would improve risk assessment. However, the development and use of models must be coupled to the collection of richer data sets from toxicity studies.

The discussion then moved to a consideration of the problems that are inherent in choosing the most sensitive response (e.g., a precursor event). If chosen as the critical effect for one agent, and the same uncertainty factors are applied as might be applied to a more severe effect for a different agent, the reference dose for the former may be unduly conservative (or insufficient) for the latter. The breakout group briefly examined the traditional uncertainty factors to determine whether it would be appropriate to decrease the magnitude, or to eliminate certain factors, when basing an assessment on subtle precursor effects. While the group was able to identify some examples where uncertainty factors may not be relevant, little progress was made. The group recommended that this issue be discussed separately by Group 3. Nevertheless, the group did urge that risk assessments explicitly distinguish between precursor events and frankly adverse effects. It was further recommended that, with enough information, mode of action-based models might be a viable alternative to uncertainty factor-based approaches for setting RfDs.

Question 4: How Should Slope of the Dose-Response Curve Be Considered in Dose-Response Assessment?

There was a great deal of discussion about whether steep or shallow dose response curves were of greater concern. If a dose-response curve has a shallow slope in the observable range, then the expectation is that it would continue to have a shallow slope at lower dosages, and that the default factors applied to arrive at a RfD may not be sufficient to achieve the desired risk reduction. Steep curves would be expected to require smaller margins between the NOAEL and RfD to achieve comparable risk reduction.

There was some concern in the committee that the combination of a steep slope and small margin of safety is imprudent, because errors in estimation may have significant health consequences. The discussion on this point was far ranging, with some members of the group expressing their opinion that the level of their concern was contingent on the nature of the effect or the population that may be affected (e.g., children). No consensus was reached on this problem, but the nature of the concerns suggested that this is more an issue of risk management than risk assessment.

There are always numerous dose-response curves from the same study, even for different effects that are produced by the same mode of action. The question was raised of which dose-response curve to use in determining slope, particularly given that precursor events or lesser manifestations of toxicity would be present at more dose levels and might, therefore, have dose-response curves with better resolution than frank effects. It was decided that, if precursor effects were used as the basis for determining slope magnitude, the causal, rate-limiting precursor effect was the appropriate one to use. Limitations to this approach might be experienced when more severe effects of higher doses mask precursor effects, or if there is uncertainty concerning the "true" causal events.

The group discussed whether the slope of the dose-response curve from an animal study had predictive value for the shape of the dose-response curve in humans; if not, then there would be little reason to adjust for slope from animal studies. It was felt that mode-of-action information, including interspecies studies of pharmacokinetics, would be useful in addressing this concern. If it can be shown that the mode of action driving the dose-response relationship is operable in humans, then it would be more likely that the slopes of the dose-response curves would be similar.

The group considered the use of adjustment factors to account for slope in a RfD determination. One idea from the proposed cancer risk assessment guidelines is an adjustment factor that would decrease as the slope increases (Fig. 1). It was also pointed out that confidence intervals on the dose-response curve partially address the shallow/steep controversy, because shallow curves have wider confidence intervals. In plenary discussion, it was noted that such use of adjustment factors to account for slope was in contrast to other recommendations calling for greater direct use of the dose-response relationships.

Question 5: Risk Reduction: A Discrepancy between Cancer and Noncancer Risk Assessment.

One of the major issues that the breakout group struggled with was the apparent discrepancy between cancer and noncancer risk assessment approaches in reducing risks. In traditional cancer risk assessment approaches, the dose-response curve is extrapolated far below the experimental dose range to

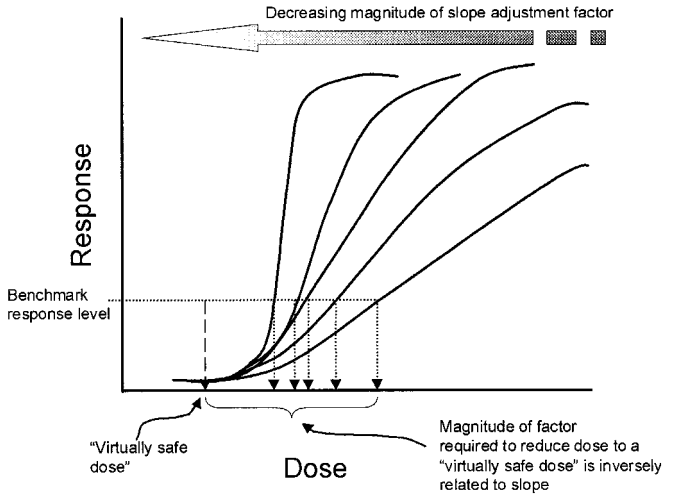

FIG. 1. Illustration of the concept of an adjustable factor to account for the slope of the dose-response curve. For steep dose-response curves, a small factor might adequately reduce risk to negligible levels ("virtually safe dose") whereas a large factor might be required for shallow-slope dose-response curves. Such a factor does not exist in current approaches to dose-response assessment and no specific factor or mechanism for determining the magnitude of the factor was discussed.

determine a dose that carries virtually no risk. Nominally at least, the exercise is very much one of risk reduction, with other factors, such as potential interspecies or intrahuman variability, addressed only indirectly by the use of a lower confidence limit on dose-response modeling analysis.

On the other hand, noncancer risk assessment involves the application of factors that account for the uncertainty in extrapolating across species or across the human population, but with no explicit factor for risk reduction other than an uncertainty factor for extrapolating a NOAEL from a LOAEL. In other words, if one determines experimentally that there is no more than a 10% risk at a certain dosage in the tested animal species, then applies 10-fold uncertainty factors to cover the possibility that humans are 10 times more sensitive than animals, and assumes the most sensitive human subpopulation is 10 times more sensitive than the average human, it can be argued that the risk of toxicity to that subpopulation of exposure at the RfD is still around 10%, at least in theory (Sheehan *et al*., 1989).

The group found that maintaining these two different approaches is likely to be an impediment to harmonization of risk-assessment practices. The divergent approaches currently used for cancer and noncancer risk assessment are also an impediment to the acceptance of the benchmark dose as an alternative to the no-observed-adverse-effect level (NOAEL) as a starting point for risk estimation. The benchmark dose is a defined, nonzero level of effect, whereas the NOAEL represents an indeterminate level of response that may be zero or greater and that is highly influenced by study design issues affecting the statistical power of the study. One solution that was suggested for the benchmark dose/NOAEL problem is better communication of the true meaning of NOAEL, particularly in showing that there may be risk at the NOAEL. For studies of low statistical power, it may be that risks of adverse effects at the NOAEL are significant and may be much higher than expected at the benchmark response level. However, estimating doses associated with lower risks requires extrapolation outside of the region of experimental observation, and this process is also fraught with uncertainty. The closer the extrapolation is to the region of observation, the less significant these uncertainties become.

Report of Breakout Group 3: Scaling and Uncertainty Factors

Question 1: How Should Interspecies Adjustments for Dose Be Made?

The current approaches of the U.S. EPA to interspecies adjustments for dose are different for noncancer and cancer dose-response assessments for ingested chemicals. This is not the case for risk assessments of inhaled chemicals in which the reference concentration (RfC) approach is used regardless of type of toxic endpoint (Jarabek, 1995; U.S. EPA, 1996). Thus, dosimetric adjustments for inhalation exposures to toxicants are already harmonized, at least within the EPA. There currently exists no guidance regarding dosimetric adjustments for dermal exposures.

For ingestion exposures, the current practice advocated by most regulatory agencies worldwide is to make adjustments for dose when extrapolating from data in test animals to humans by assuming that pharmacokinetic differences impart greater internal exposure to parent chemical or its toxic metabolites in humans, relative to test species. Traditionally, dosimetry adjustment has been accomplished using categorical defaults (Appendix 1); i.e., by dividing the NOAEL or LOAEL by factors of 10 to allow for interspecies differences and human variability (Dourson and Stara, 1983). Recently, it has been recognized that each factor of 10 has to allow for pharmacokinetics (PK) and pharmacodynamics (PD); both of which contribute to the overall response (Andersen *et al*., 1995; Renwick, 1991, 1993). It has been suggested that the inter- and intraspecies extrapolation factors of 10 are each comprised of two log-normally distributed factors; $10^{0.5}$ or approximately three for each. Thus, it has been proposed that each 10-fold factor be subdivided into two numerical values, the product of which is the original value of 10 (Renwick, 1993; IPCS, 1994). This scheme allows chemical-specific data to replace either the PK or PD components of the default factors.

Evaluations of data on PK and PD differences between rodents and humans and among different human individuals gave rise to the proposal that the interspecies factor be split (4.0 and 2.5 for PK and PD, respectively), and that the 10-fold factor for human variability be evenly divided into PK and PD values equal to 3.16 each (Appendix 2, Fig. 2) (IPCS, 1994; Renwick, 1993; Renwick and Lazarus, 1998). Segregating the

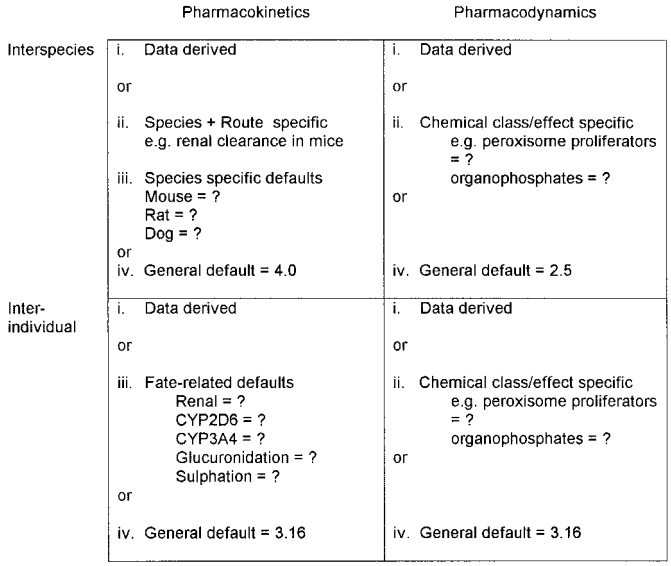

FIG. 2. Subdivision of the 100-fold uncertainty factor and possible future refinements as proposed by Renwick and Lazarus (1998). ? = a value which is currently being developed on the basis of a comprehensive review of existing published data. In each box is the option to use a data-derived (chemical-specific) value or various default values (ii, iii, and iv), which can be selected depending on the extent of knowledge of the compound under evaluation. In the absence of a BBDR model, the factor selected within any box would be at the highest level (i, highest level; iv, lowest) that could be justified scientifically. The total uncertainty factor would be the product derived by multiplying the value in each box. The general default values (which collapse back to 100) would be the fallback position in the absence of chemical-specific/relevant data.

interspecies factor of 10 into separate (PK and PD) components allows investigation of these components for a given chemical risk assessment through further directed experimental research and replacement of a PK or PD default by chemical-specific data (Renwick, 1993). The product of a chemical-specific value, and remaining default factors for which data are not available, give rise to a chemical-specific factor termed a "data-derived uncertainty factor." Appendix 2 and Figure 2 establish a framework whereby the results of chemical-specific investigations conducted to inform these components can be easily integrated into risk assessment.

The RfC interspecies dosimetric adjustment thereby provides partial accounting for the processes that affect dosimetric differences between rodents and humans and justifies reduction of the composite interspecies factor of 10 to a residual factor of three. The residual factor of three represents the uncertainty remaining in interspecies differences in PD. The RfC dosimetry adjustment practiced by the U.S. EPA for inhaled chemicals is similar to the process described by Renwick, in that the adjustment factor is recognized as having PK and PD components. These factors are believed to be log-normally distributed and therefore are each equivalent to approximately three. The RfC methodology replaces the PK factor of three with factors specific for site-of-contact respiratory toxicants, inhaled systemic toxicants, or inhaled particulates (Jarabek, 1995).

The introduction of chemical-specific data by the application of data-derived factors for dosimetry (PK) adjustment is gaining wider acceptance internationally (IPCS, 1994, 1999; Meek, *et al.*, 1994; Renwick, 1993) for both cancer and noncancer dose-response assessments. However, when conducting cancer risk assessments, the EPA uses a different approach to effect a similar accounting for dosimetric differences between rodents and humans. For cancer risk assessments only, EPA applies a default approach whereby interspecies dosimetry adjustment is accomplished through allometric scaling. In this approach, dose is adjusted by multiplying the daily dose, in mg, by the ratio of body weights raised to the $\frac{3}{4}$ power.[3] This approach effectively reduces the daily dose by a factor of approximately 4 or 7, if extrapolating from rat or mouse data, respectively. This adjustment, based largely on empirical observations for metabolism, clearance, and cancer potencies across species, is said to account for both PD as well as PK differences between rodents and humans (U.S. EPA, 1992).

The Group 3 participants discussed these approaches to dosimetric adjustment and clarified differences between the default uncertainty factors being considered here versus those additional factors that are more appropriately referred to as "safety factors." The workgroup agreed that, for the purposes of this workshop, the term "safety factors," such as that proposed in the context of the Food Quality Protection Act to protect children from risk of pesticide exposure, is appropriate when applied strictly to provide additional precaution or comfort. The necessity and magnitude of these factors are the purview of risk management. As such, safety factors were not discussed by the group. It should be made clear, however, that in some countries and advisory bodies, the default uncertainty factors for inter- and intraspecies differences are referred to as "safety factors."

The following points of consensus were reached by Breakout Group 3:

(1) There is no scientific basis for considering dosimetry adjustments differently for different toxicity endpoints (cancer or noncancer) except to choose dosimeters that are most closely associated with mode of action. A harmonized approach to dosimetry adjustment is appropriate and desirable. Use of dosimetric correction factors, whether they are values derived by subdivision of the original 10-fold factors (e.g., the general defaults presented in Fig. 2) or derived through chemical-specific direct research on interspecies differences in PK and PD (e.g., through the development of biologically based PK and/or PD models), should be guided by the most plausible hypothesis of mode of toxic action. Likewise, the animal model chosen as the basis for dose extrapolation should be that which

[3] In practice, test species dose in mg/kg/day is multiplied by the ratio of test species body weight$^{1/4}$:human body weight$^{1/4}$.

physiologically and biochemically most resembles humans and which shows the toxic effect of concern.

(2) When developing an approach to dosimetry adjustment factors, all available data should be considered before resorting to defaults.

(3) PK and PD should be assessed explicitly in any approach to inter- and intraspecies dosimetry adjustment.

(4) Biologically based dose response models are the preferable means for assessing interspecies differences in PK and PD. As a general rule, this requires sufficient information to support a hypotheses of mode of action that is consistent with the weight of available data, and quantitative descriptions of the critical events that capture the biological data better than the default. However, uncertainty associated with the assumptions and data underlying their use should be considered in the risk assessment. In general, as more chemical-specific information is used in the risk assessment, uncertainty also transitions from generic to specific.

(5) In the absence of a biologically based PK/PD model, data-derived factors should be used when possible. An example of data derived factors is shown in Fig. 2. A data-derived factor is a composite compound-specific adjustment factor derived as the product of chemical-specific data and default values for undefined areas of uncertainty. The concepts supporting data-derived factors, described in the literature, are being adopted by health agencies internationally (IPCS, 1994; Health Canada, 1994), and examples of their application are available (e.g., Moore, 1995; Zhao *et al.*, 1998). Describing default factors as probability distributions might expand this approach further (Hattis *et al*., 1999). Where a composite data-derived factor cannot be determined from available data on a chemical, less quantitative data such as the pathway of metabolism may be useful to determine a suitable categorical default (Fig. 2). Useful categorical defaults would be considered only in the absence of any other more appropriate quantitative chemical-specific PK or PD data.

Question 2: How Should Exposure-Duration Relationships Be Taken into Account?

Currently, many aspects of adjusting for exposure-duration are dealt with for both cancer and noncancer health effects using defaults based on the C × T concept. This concept, associated with Haber (1924), assumes that equal effects are observed when the product of concentration and time is equal, regardless of the value of either parameter individually. For example, in setting a chronic RfD for noncancer health effects, C × T is the default assumption, and adjustments are made on that basis. Examples of such adjustments include extrapolation from 5 days per week to 7 days per week for an oral exposure, or from 6 to 24 h/day for inhalation exposure. However, when extrapolating from a subchronic (approximately 90-day study) to a chronic (usually 2-year rodent) study, an uncertainty factor (default 10-fold factor) is applied to account for uncertainties associated with the lack of chronic data. For shorter-term reference values, for example an acute inhalation reference value, the C × T adjustment is made when extrapolating from shorter-term to longer-term exposures (e.g., 6 to 24 h), but not when extrapolating from longer-term to shorter-term exposure duration. In the latter instance, the longer-term reference value is typically used for both exposure duration, as a more conservative estimate of the potentially toxic dose. This approach to duration adjustment is a conservative use of Haber's Law, as shown in several studies on neurotoxicity and developmental toxicity (e.g., Bushnell, 1997; Crofton and Zhao, 1997; Weller *et al*., 1999). Modeling of data from these studies indicates that using C × T for adjusting from shorter to longer duration tends to overestimate risk, while similar adjustments from longer to shorter duration tends to underestimate risk.

For cancer, most data are from near-lifetime (2-year bioassay) studies in rodents, although less-than-lifetime data are sometimes used, and data from humans may be available. The C × T concept is also used in cancer dose-response assessment to calculate the lifetime average daily dose (LADD) as follows:

$$\text{LADD} = \frac{\text{mg/kg/day dose rate in animal study} \times \text{duration of animal study (days)}}{\text{Lifetime (days)}}$$

Where interim sacrifice data are available, the validity of the C × T assumption can be assessed. Alternatively, these data may be useful for deriving more appropriate methods of relating exposure duration to response. Group 3 agreed that studies should be designed with greater attention to the utility of interim and postexposure sacrifice data for addressing dose/temporal toxicity relationships.

Within the framework proposed (Appendix 2), there were a number of issues pointed out by the group that should be considered in adjusting for exposure-duration relationships:

- The need for and approach to dose-duration adjustments is dependent on the mode of action, e.g., whether the effects are the result of cumulative exposures or short duration exposures, as well as whether the chemical itself accumulates and/or the damage accumulates.
- Exposure-duration adjustments should be based on pharmacokinetic considerations such as the extent of bioaccumulation and the percentage of steady state attained in the experimental study (Renwick, 1999). PBPK models are particularly useful for exposure duration extrapolations and enable consideration of the appropriate dose metric such as area under the curve (AUC) versus peak plasma levels.
- Exposure duration should be closely related to the timing of the event, particularly for different ages, physiological states, and during development.
- Adjustments should be based on physiological time versus chronological time.

• Intermittent exposure patterns and how those patterns affect overall internal dose should be considered.

• If a chemical acts at a time of susceptibility during life (e.g., lack of detoxication or repair mechanisms) then risk assessments should address that time of vulnerability. But this same approach should be used for all endpoints for which this is the relevant mode of action.

The workgroup agreed that the approaches used for exposure-duration adjustment should be the same for cancer and noncancer endpoints. These approaches should be harmonized under the concept that mode of action provides the rationale used for exposure-duration adjustments.

Question 3: How Can Interspecies and Intraspecies Variability Be Treated Consistently for All Endpoints?

Variability in responses to toxic agents is a combined function of variability in individual physiology and biochemistry that affect PK and PD. The group was very clear in endorsing the need for explicit and consistent accounting for variability within the human population. Thus, the framework presented in Appendix 2 and Figure 2 allow for data-derived categorical expressions of variability or the use of biologically based PK/PD models. These approaches are appropriate for both cancer and noncancer endpoints.

The group also agreed that where data are available, distributional approaches to individual susceptibility among humans for toxic effects (governed by PK and PD differences) might replace default PK and PD components that account for the within-human-population variability (Hattis *et al*., 1999). The challenge will be to adequately separate and characterize the variability of the PK and PD components. Thus, it might be possible, with further research, to replace both interindividual factors of 3.16 ($10^{0.5}$) proposed by IPCS (1994) with chemical-specific population distributions for kinetics and dynamics (e.g., population distributions of the key PK and PD parameters) (Fig. 3). While these approaches are very much in their infancy, they offer the opportunity to inform and increase the level of objectivity of our assessments of human variability. The limited work to date, evaluating the PK and specific PD responses to therapeutic agents, suggests that the product of the two factors of 3.16 is adequate to cover the variability in all but approximately 2 in 10,000 individuals (assuming that variability is log-normally distributed) or 3 in 10^6 individuals (assuming that variability is normally distributed) (Renwick and Lazarus, 1998).

Question 4: How Does the Type of Adverse Effect Influence the Choice and Magnitude of Uncertainty Factors?

Two types of effects can be considered in noncancer risk assessment as possibly having an influence on the size of uncertainty factors: (1) the severity of the effect, and (2) whether it is a precursor or the adverse effect itself. Severity has been characterized in a variety of ways; for example, the degree of change within a response (mild, moderate, severe), the type of response within an organ system (liver hypertrophy vs. liver necrosis), or the type of response among organ systems (liver hypertrophy vs. renal pathology). In some cases, point scores have been given to various degrees or types of severity and have been used in ranking severity of effects. Severity also has been characterized as the degree of change with dose (NOAEL, LOAEL, frank effect, lethality), and this characterization has been used as the basis for categorizing effects for a wide range of outcomes and duration of exposure in categorical regression modeling (Hertzberg, 1989; Hertzberg and Miller, 1985). Precursor effects have been considered in dose-response assessment for noncancer health effects for some time. For example, liver hyperplasia or cellular foci of altered enzyme expression are considered possible precursors to liver pathology at higher doses or with longer exposure times. However, the type of such precursor effects has rarely, if ever, influenced the size of the uncertainty factors applied for RfD calculation. Recently, EPA proposed the use of precursor effects for cancer risk assessment, particularly for those that are thought to work through a nongenotoxic mode of action (EPA, 1996). For example, thyroid hyperplasia has been characterized as a precursor of thyroid carcinogenesis (U.S.

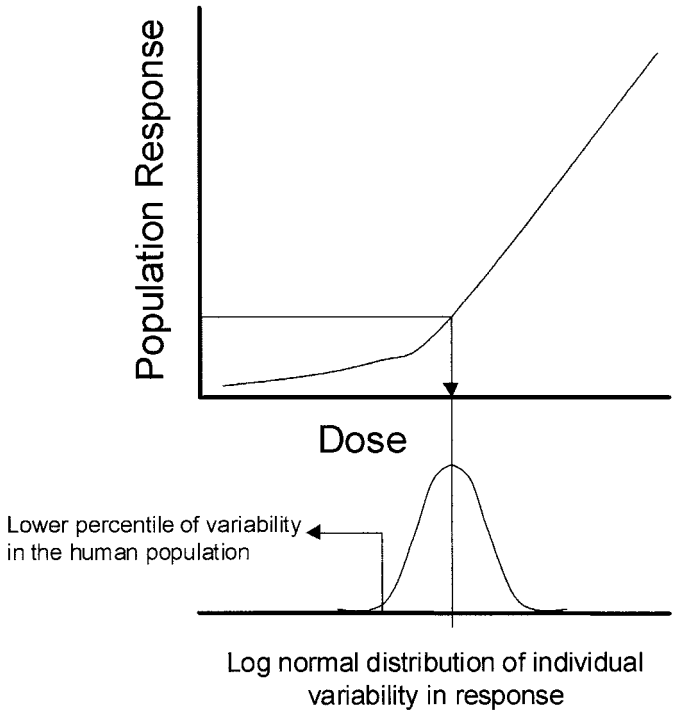

FIG. 3. Variability in the human population response to a target organ or cellular dose of a chemical may be described by probability distributions. The degree of heterogeneity (or spread of distribution) of responses in the population can be accounted for by using Z-scores to calculate the proportion of a population that would fall more than, for example, 3.16-fold away from the mean. Phenotype profiling of the population for gene expression distributions or allelic variants could be applied in this manner.

EPA, 1998). For these types of carcinogenic effects that may have a nonlinear dose-response relationship, the use of a precursor effect might be a more appropriate means for assessing the margin of exposure (MOE) considered acceptable, or the uncertainty factor applied for interspecies variability. This approach is generally consistent with noncancer dose-response assessment.

The workgroup stressed that the type of factor applied in dose-response assessment depends on the adequacy of both the pharmacokinetic and pharmacodynamic information available. All data for both cancer and noncancer health effects must be considered in making judgements about severity and its influence on the size of the uncertainty factor applied. Certainly, the move toward using mode of action information in risk assessment will include consideration of the nature of an effect. One major concern expressed about using severity as a basis for comparison is that the levels of effect for different outcomes need to be comparable, particularly when comparing across organ systems. However, this requires scientific judgement, which can be very difficult. From a public health point of view, one might ask questions about how much harm will be done if the choice of the critical effect is wrong, and whether the effect is reversible, a common or rare effect, and/or associated with a common exposure. It was stated that dose-response evaluations should be conducted for all endpoints, including calculation of the human equivalent doses and the application of uncertainty factors in order to determine the most appropriate endpoint to use for risk-assessment purposes.

Group 3 briefly discussed the issue of adversity of responses (e.g., developmental toxicity vs. neurotoxicity) and recognized the magnitude of this issue as too important to be resolved in that workshop. It was apparent from the discussion that there are several ways to approach the problem, and that it should be pursued in a separate forum.

Question 5: What Aggregate Uncertainty Factor (i.e., Product of Individual Uncertainty Factors) Is Appropriate to Determine the RfD and RfC, or to Evaluate the MOE? How Does Mode of Action or Spectrum of Toxic Manifestations Influence the Answer to this Question?

The group refused to specify a default aggregate uncertainty factor, feeling that this was not a science-based approach to the problem, and emphasized that the choice of the total uncertainty factor depends on the database available as well as the mode of action. The group strongly agreed that approaches to cancer and noncancer assessment should be harmonized, particularly abandoning the cancer/noncancer dichotomy, and considering instead the mode of action in guiding decisions on methods of low-dose and interspecies extrapolation. The group discussed the different assumptions that are made as the basis for linear (genotoxic, no-threshold) and nonlinear (nongenotoxic, possible threshold) low-dose approaches. In the case of linear low-dose assessment, the resulting metric is a dose associated with a specific risk (a cancer potency or slope factor) that can be used to derive a "virtually safe dose" (e.g., a dose associated with a 1×10^{-6} cancer risk). These expressions of potency can impart more precision than is warranted. In contrast, for nonlinear assessments, the result is a "virtually safe dose" with no associated estimate of risk. Some members of the group questioned the linear low-dose assumption, and stated that much of the early work leading to this assumption was based on radiation data, which may be very different from chemical reactions in the body. Others raised the issue of additivity to background, especially for agents that occur endogenously, or those that add to endogenous mechanisms, advocating the use of linear low-dose assessment in these cases. Ultimately, there was consensus that biological thresholds are possible for both cancer and noncancer endpoints. Most advocated the use of uncertainty factors for both cancer and noncancer risk assessments and of calculating only a "virtually safe dose" or its equivalent without a slope factor. A minority of participants did not want to exclude the possibility of low-dose linearity for any carcinogenic mechanism and expressed the need for slope factors that enable population risk estimates.

Question 6: How Should One Account for a Database Deficiency?

Currently, risk assessment approaches used by various regulatory agencies account for database insufficiency by further dividing the NOAEL by an uncertainty factor. For example, the U.S. EPA uses a database uncertainty factor of up to 10 (generally either 3 or 10) if key toxicity studies are not available or are unreliable. The Group 6 discussion concluded with the following major points:

- The minimal amount of data that should be required is that which is necessary to make a plausible argument for mode of action or to enable a default risk assessment.
- The adequacy of the database depends on the purpose of the risk assessment. For example, assessing risk for an exposure scenario involving only short-term high concentration inhalation exposures does not require a chronic inhalation study. Applying an uncertainty factor for the lack of chronic inhalation data in this case would not be appropriate. Similarly, there may be no need for an uncertainty factor to account for missing endpoint-specific systemic toxicity data when the database overwhelmingly indicates that effects are limited to the site of contact and these effects are the basis for an RfD/RfC. In cases where key data are missing and necessary, an uncertainty factor should be applied.
- The risk assessment process should be structured such that there are incentives for conducting toxicity research and testing beyond regulatory mandates in order to further inform qualitative and quantitative extrapolations, to further reduce risk-assessment uncertainties, and to yield information that might be more broadly useful to other assessments (e.g., structural class-toxicity assessments).

- A guiding question in determining the need for additional data is, how much more improved will a decision be with additional data?
- The precision of a risk assessment outcome is generally recognized as no more than one arithmetic digit. Thus, the precision reported in a risk assessment must be tempered by the uncertainties inherent in the assessment.

Summary and Conclusions

Traditional approaches to risk assessment are being challenged, and this is especially true in the separation of cancer and noncancer risk assessment. *Science and Judgment in Risk Assessment* (National Research Council, 1994) noted the importance of an approach that is less fragmented, more consistent in application of similar concepts, and more holistic than centering on specific endpoints. The report questions the application of a nonthreshold quantitative approach as a default in all cancer risk assessments, as well as the use of a threshold concept as a default for agents that cause other types of toxicity. The U.S. EPA has begun to address this changing philosophy, and its revised cancer risk-assessment guidance has proposed departing from the assumption that all cancer effects show linear, nonthreshold dose-response relationships (U.S. EPA, 1996).

Accomplishments of the Workshop

A major accomplishment of the workshop was providing an opportunity for experts to discuss issues important to harmonizing approaches to risk assessment. Both the plenary and work group sessions allowed for considerable discussion of definitions, terminology, opinions, and positions. In many cases, consensus was achieved, but where it was not, the range of opinions was identified. In general, there was a growing acceptance that the field of risk assessment must move away from the traditional dichotomies: cancer/noncancer, linear/nonlinear. Participants referred to carcinogenesis as an endpoint of toxicity and not as an entity separate from other forms of toxicity.

The workshop developed specific definitions that will prove useful in the future. One was the definition of "harmonization," as it was used for the purposes of this workshop:

Harmonization refers to developing a consistent set of principles and guidelines for drawing inferences from scientific information. It does not mean that a single method should be used for the assessment of all toxicities and chemicals.

In addition, as noted above, Group 1 developed a working definition of mode of action that could be considered applicable to all toxic manifestations. There was an acceptance that understanding the mode of action of a chemical is ultimately critical for nondefault risk assessment, that common modes of action for different toxicities can be defined, and that our approach to assessing toxicity data should be biologically consistent. However, it was also recognized that a toxic response to an environmental exposure should not be viewed simplistically, and that there is a need for linking mode-of-action with the exposure and response at different levels of biological organization. Particular importance was given to assessing toxicity within the context of the exposure-response relationship. This pertains not only to the ultimate manifestation of toxicity, but to the entire process of pathogenesis. This will include effects at the molecular, cellular, and tissue/organ level, as well as the whole organism level, requiring a better understanding of how various biological events are related, both qualitatively and quantitatively, to the overall effect on an organism. There was consensus that, to the extent possible, all data should be considered in assessing the potential toxicity of an exposure, that the data should drive the choice of analysis and not vice versa, and that the basis for any assessment should be biological and not statistical. Choosing the "most sensitive endpoint" can severely limit the data available for consideration, and will not provide a profile of the overall toxicity that may be associated with an exposure. Limiting the data also squanders information that may be important when considering other exposure scenarios or regulatory uses of the data.

The Future

The future will bring a growing number of challenges to the field of risk assessment. One of the most immediate will be in shifting our mind set from the traditional dichotomies to a more holistic view. The workshop provided an opportunity to begin that process. The field will have to develop a comfort level with treating all endpoints of toxicity with a consistent set of principles and guidelines for drawing inferences from scientific information. This was not entirely realized at the workshop. There still seemed to be a lingering willingness to accept certain assumptions for justifying our approach to cancer, but not accepting identical or similar assumptions for approaches to other endpoints. This may be due in part to a natural reluctance to approach what we should be doing in a way that is not predicated on what we have done before. But it is also due to the growing complexity of the field of risk assessment, and the need to be able to develop, communicate, and understand a wide range of scientific and analytical areas. This complexity will only increase as new technologies provide us with considerably more data for inclusion in risk assessments. We will have to reach out to other disciplines and incorporate the broader components in toxicology, but not at the expense of maintaining a strong, in depth expertise in specific areas.

This workshop was the culmination of over a year of planning by the cosponsors and organizers. While much was accomplished, the workshop should be considered a significant first step toward harmonizing approaches to cancer and noncancer risk assessment. Certainly, the need for continued work and discussion on the harmonization of risk-assessment approaches is clear. Hopefully, these workshop proceedings can serve as a jump start for future work in this area.

APPENDIX 1

Descriptions of Categorical Defaults (from A. G. Renwick and endorsed by Breakout Group 3)

Adequacy of database-related defaults. A minimum database is recognized for any risk assessment, although the requirements may vary, depending on the risk-assessment scenario. Additional categorical default factors may be added to allow for deficiencies in the database, for example:

- *LOAEL to NOAEL*. The LOAEL may be divided by a factor if a NOAEL has not been defined.
- *Subchronic to chronic.* A factor may be used if there is no chronic study to match chronic human exposures (refer to discussion of question on exposure-duration relationships).
- *Database deficiencies. A* factor might be used if there is a part of the human life cycle during which exposure occurs, and which has not been tested in animal studies.
- *Severe toxicity*. A factor might be applied when the only toxicity data relates to doses causing severe toxicity.

All of these factors could be removed by developing the appropriate test data. Further analysis of existing databases, for example by distribution analysis, may help to provide more scientifically based default values, or allow estimation of uncertainty distribution.

Interspecies defaults (adjustment factors). The selection of the default factor depends on the route of exposure and the site of uptake/deposition. Different factors may be used for oral toxicity, causing systemic effects (typically 10) and inhalation (with dosimetric correction for deposition), causing local toxicity (typically 3). Interspecies differences arise from both kinetic and dynamic differences. Alternative defaults are possible to account for kinetic aspects, such as the ratio of body weight$^{0.75}$ or body weight$^{0.66}$ or a generic kinetic default of 4.0 to allow for species differences in parent compound after oral dosage (see Appendix 2). These kinetic defaults would be multiplied by the generic default for dynamics of 2.5. Future developments could include pathway-related defaults (for kinetics) and process or mode of action-related defaults (for dynamics) to allow fine-tuning of the factor to the chemical in the absence of detailed chemical-specific data (see Fig. 2).

Intraspecies (interindividual) defaul. A categorical default of 10 is usually applied for both oral and inhalation to account for kinetic and dynamic variability in the human population. Future developments will include pathway-related defaults (for kinetics) and process or mode of action-related defaults (for dynamics), to allow fine-tuning of the factor to the chemical in the absence of detailed chemical-specific data.

APPENDIX 2

A Framework Used by Breakout Group 3 in Discussion of How to Derive Interspecies Adjustments for Dose (from A. G. Renwick, as endorsed by Breakout Group 3)

Categorical default factors. Factors applied on the basis of general properties: values derived from databases for compounds sharing (wherever possible) similar general properties, used in the absence of chemical-specific data. Current standard default factors (such as 100 when the NOAEL is for systemic toxicity after chronic oral administration to animals, or such as 10 for the LOAEL to NOAEL factor) are considered to be categorical, because they relate to simple categories, such as oral/inhalation, systemic/local effects or database deficiencies. Current standard defaults were consistent with data on compounds available at the time of their development. Categorical default factors include factors related to adequacy of database, interspecies differences, and interindividual variability (see Appendix 1).

Chemical-specific (adjustment) factors. Factors in which one or more of the categorical defaults for inter- and intraspecies differences has been modified by the incorporation of chemical-specific data regarding kinetics or mode of action.

Data-derived factors. Factors in which physiologically based kinetic parameters or target organ sensitivity data are used to replace part of a default (requiring information on mode of action). (Factors may be required to consider uncertainty in chemical-specific date/approach.)The product of the chemical-specific values and any remaining defaults is termed a data-derived safety factor or a data-derived uncertainty factor.

PBPK analysis. Used to replace the toxicokinetic aspect of defaults (requires information on mode of action). PBPK models are useful for describing interspecies, high-to-low dose, and temporal extrapolations but require key physiological and biochemical kinetic constants. These may be available from literature or experimentally derived. (Factors may be required to consider uncertainty in chemical-specific date/approach.)

BBDR (biologically based dose-response used to replace the pharmacokinetic aspects together with part or all of the pharmacodynamic aspects). Usually applied to interspecies comparisons; rarely used to model human variability. (Factors may be required to consider uncertainty in chemical-specific date/approach.)

ACKNOWLEDGMENTS

The steering committee and workshop participants would like to thank the following organizations for sponsoring the workshop: American Industrial Health Council, National Institute of Environmental Health, Society for Risk Analysis, Society of Toxicology, and United States Environmental Protection Agency. The following participants were members of the steering committee: Matthew S. Bogdanffy, Gail Charnley, George Daston, Elaine Faustman, Carole Kimmel, Gary Kimmel, Steve Lewis, George Lucier, Jennifer Seed, and Vanessa Vu. The Committee would also like to thank the following invited participants for their contributions during the meeting and for their review and comment on this manuscript: Sandra Baird, Hugh Barton, Michael Bolger, Richard Bull, James Bus, Byron Butterworth, Harvey Clewell, Jim Cogliano, Rory Conolly, Kerry Dearfield, Vicki Dellarco, Michael Dourson, Michael Gallo, David Gaylor, Jay Goodman, Jean Harry, Dale Hattis, Richard Hill, Annie Jarabek, Abbi Li, Mike Luster, Lynn McGrath, Elizabeth Meek, Suresh Moolgavkar, Jack Moore, Edward Ohanian, Chris Portier, Andrew Renwick, Rob Schnatter, Bernard Schwetz, Woodrow Setzer, William Slikker, Lewis Smith, Mark Stanton, Tom Starr, Lewis Smith, James Swenberg, Jane Teta, Mark Utell, and Lauren Zeise.

REFERENCES

Allen, B. C., Strong, P. L., Price, C. J., Hubbard, S. A., and Daston, G. P. (1996). Benchmark dose analysis of developmental toxicity in rats exposed to boric acid. *Fundam. Appl. Toxicol*. **32,** 194–204.

Andersen, M. E., Clewell, H., and Krishnan, K. (1995). Tissue dosimetry, pharmacokinetic modeling, and interspecies scaling factors. *Risk Anal.* **15,** 533–537.

Bushnell, P. J. (1997). Concentration-time relationships for the effects of inhaled trichloroethylene on signal detection behavior in rats. *Fundam. Appl. Toxicol.* **36,** 30–38.

Butterworth, B. E., and Bogdanffy, M. S. (1999). A comprehensive approach for integration of toxicity and cancer risk assessments. *Regul. Tox. Pharm.* **29,** 23–36.

Conolly, R. B. (1995). Cancer and noncancer risk assessment: Not so different if you consider mechanisms. *Toxicology* **102,** 179–188.

Crofton, K. M., and Zhao, X. (1997). The ototoxicity of trichloroethylene: Extrapolation and relevance of high-concentration, short-duration animal exposure data. *Fundam. Appl. Toxicol*. **38,** 101–106.

Dourson, M. L., and Stara, J. F. (1983). Regulatory history and experimental

support of uncertainty (safety) factors. *Regul. Toxicol. Pharmacol*. **3,** 224–238.

Faustman, E. M., Ponce, R. A., Seeley, M. R., and Whittaker, S. G. (1996). Experimental approaches to evaluate mechanisms of developmental toxicity. In *Handbook of Developmental Toxicology* (Hood, R., Ed.), pp. 13–41. Boca Raton, FL, CRC Press.

Gaylor, D. W., and Slikker, W. (1990). Risk assessment for neurotoxic effects. *Neurotoxicology* **11,** 211–218.

Haber, F. (1924). Zur Geschichte des Gaskrieges. In *Fuenf Vortraege aus den Jahren 1920–1923*, pp. 76–92. Julius Springer, Berlin.

Hattis, D., Banati, P., and Goble, R. (2001). Distributions of individual susceptibility among humans for toxic effects. How much protection does the traditional tenfold factor provide of what fraction of which kinds of chemicals and effects? *Ann. N.Y. Acad. Sci*. **895,** 286–316.

Health Canada (1994). Human Health Risk Assessment for Priority Substances. Environmental Health Directorate, Health Canada, Ottawa, Ontario.

Hertzberg, R. C. (1989). Extrapolation and scaling of animal data to humans: Fitting a model to categorical response data with application to species extrapolation of toxicity. *Health Phys.* **57**(Suppl. 1), 405–409.

Hertzberg, R. C., and Miller, M. (1985). A statistical model for species extrapolating, using categorical response data. *Toxicol. Ind. Health* **1,** 43–57.

Hill, A. B. (1965). The environment and disease: Association or causation. *Proc. R.. Soc. Med.* **58,** 295–300.

IPCS (1994). *Environmental Health Criteria No. 170: Assessing Human Health Risks of Chemicals: Derivation of Guidance Values for Health-based Exposure Limits.* International Programme on Chemical Safety, World Health Organization, Geneva.

IPCS (1999). Environmental Health Criteria 210. *Principles for the Assessment of Risks to Human Health from Exposure to Chemicals*. International Programme on Chemical Safety, World Health Organization, Geneva.

IPCS (1999). IPCS workshop on Developing a Conceptual Framework for Cancer Risk Assessment, 16–18 February 1999, Lyon, France, IPCS/99.6.

IRLG. Working Group on Risk Assessment, International Regulatory Liaison Group (1979). Scientific basis for identification of potential carcinogens and estimation of risks. *J. Natl. Cancer Inst.* **63,** 241–268.

Jarabek, A. M. (1995). Inhalation RfC methodology: Dosimetric adjustment and dose-response estimation of noncancer toxicity in the upper respiratory tract. *Inhal. Toxicol.* **6**(Suppl.), 301–325.

Kavlock, R. J., Allen, B. C, Faustman, E. M., and Kimmel, C. A. (1995). Dose response assessments for developmental toxicity: IV. Benchmark doses for fetal weight changes. *Fundam. Appl.. Toxicol.* **26,** 211–222.

Lehman, A. J., and Fitzhugh, O. G. (1954). 100-Fold margin of safety. *Assoc. Food Drug Off. U.S.Q. Bull* **18,** 33–55.

Meek, M. E., Newhook, R., Liteplo, R. G., and Armstrong, V. C. (1994). Approach to assessment of risk to human health for priority substances under the Canadian Environmental Protection Act. *Environ. Carcinog. Ecotoxicol. Rev.* **C12,** 105–134.

Moore, J. A., (1995). An assessment of lithium using the IEHR Evaluative Process for Assessing Human Developmental and Reproductive Toxicity of Agents. IEHR Expert Scientific Committee. *Reprod. Toxicol*. **9,** 175–210.

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

Renwick, A. G. (1991). Safety factors and establishment of acceptable daily intakes. *Food Addit. Contam.* **8,** 135–149.

Renwick, A. G. (1993). Data-derived safety factors for the evaluation of food additives and environmental contaminants. *Food Addit. Contam.* **10**, 275–305.

Renwick, A. G. (1999). Duration of intake above the ADI/TDI in relation to toxicodynamics and toxicokinetics. *Regul. Toxicol. Pharmacol*. **30,** S69–78.

Renwick, A. G., and Lazarus, N. R. (1998). Human variability in noncancer risk assessment—an analysis of the default uncertainty factor. *Regul. Toxicol. Pharmacol.* **27,** 3–20.

Schlosser, P. M., and Bogdanffy, M. S. (1999). Determining modes of action for biologically based risk assessments. *Regul. Toxicol. Pharmacol.* **30,** 75–79.

Sheehan, D., Young, J. F., Slikker, W., Gaylor, D., and Mattison, D. (1989). Workshop on risk assessment in reproductive and developmental toxicology: Addressing the assumptions and identifying the research needs. *Regul. Toxicol. Pharmacol.* **10,** 110–122.

Simpson, D. G., Carroll, R. J., Zhou, H., and Guth, D. (1996). Weighted logistic regression and robust analysis of diverse toxicology data. *Commun. Stat. Methods* **25**, 2615–2632.

U.S. EPA (1992). Draft report: A cross-species scaling factor for carcinogen risk assessment based on equivalence of $mg/kg^{3/4}$/day. *Fed. Reg.* **57,** 24152–24173.

U.S. EPA (1996). *Proposed Guidelines for Carcinogen Risk Assessment.* Fed. Reg. **61,** 17960–18011.

U.S. EPA (1998). *Assessment of Thyroid Follicular Cell Tumors*. U.S. Environmental Protection Agency, Washington, DC, EPA/630/R-97/002.

U.S. EPA (1999). Guidelines for Carcinogen Risk Assessment. Risk Assessment Forum, U.S. Environmental Protection Agency, NCEA-F-0644. July 1999, Review draft.

Weller, E., Long, N., Smith, A., Williams, P., Ravi, S., Gill, J., Henessey, R., Skornik, W., Brain, J., Kimmel, C., Kimmel, G., Holmes, L., and Ryan, L. (1999). Dose-rate effects of ethylene oxide exposure on developmental toxicity. *Toxicol. Sci*. **50,** 259–270.

Zhao, Q., Unrine, J., and Dourson, M. (1999). Replacing default values of 10 with data-derived values: A comparison of two different data-derived uncertainty factors for boron. *Hum. Ecol. Risk. Assess*. **5,** 973–983.