Special Issue

Of Mice Men and Models: Future Research for Improving Risk Assessment Methods

Leslie Stayner, Mark Toraason, and Dale Hattis, Guest Editors

Human and Ecological Risk Assessment

An International Journal

Volume 8, Number 6, October 2002
ISSN 1080-7039

CRC PRESS

Human and Ecological Risk Assessment

An International Journal

Volume 8, Number 6, October 2002
ISSN 1080-7039

CRC PRESS LLC

Table of Contents

Toxicological Issues

Statistical and Biological Models of Dose and Response

Workshop Consensus Reports

Contributed Paper

Dual Track

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1195-1202 (2002)

Risk Assessment at the Crossroads of the 21st Century: Opportunities and Challenges for Research

Leslie Stayner,[1]* Mark Toraason,[2] and Dale Hattis[3]
[1]Risk Evaluation Branch, Education and Information Division, NIOSH C15, 4676 Columbia Parkway, Cincinnati, Ohio 45226; [2]Division of Applied Research and Technology, NIOSH; [3]George Perkins Marsh Institute, Clark University, 950 Main Street, Worcester, MA 01610

INTRODUCTION

Although one could say that assessing risks is as old as man, formalized human health risk assessment is a relatively new discipline that has largely developed as a result of environmental (U.S. Environmental Protection Agency [USEPA]) and occupational regulations (Occupational Safety and Health Administration [OSHA]) that were adopted in the 1970s. Court decisions, such as the U.S. Supreme Court's ruling on the OSHA benzene standard (*Industrial Union Department v. American Petroleum Institute,* 448 U.S. 607, 655 [1980]), have reinforced the requirements that these agencies make their best efforts to quantify risks and benefits when setting standards for protecting the public health. For better or worse, risk assessment has become a *sine qua non* for regulatory decision making in the U.S.

CHALLENGES

One word that would best describe the last 20 years of experience with risk assessment in the U.S. is "controversy." Performing quantitative assessments of risk requires extensive toxicological dose-response information in animals and, to the extent possible, in humans. Controversy arises largely from the gaps in the scientific data available for risk assessments. There is often considerable debate regarding the practice of predicting human risks based on outcomes in experimental toxicological studies with their accompanying assumptions regarding similarities or differences in interspecies metabolism of xenobiotic compounds (*e.g.*, Ames and Gold 1990).

Risk assessments based on epidemiologic data are often no less contentious. For example, risk analyses of the effects of diesel exhaust on human health has been the subject of numerous analyses, reanalyses, and debates in the last decade (Crump

* Corresponding author.

1080-7039/02/$.50

2001; Dawson and Alexeeff 2001). There are many potential biases and other factors in epidemiologic investigations that are difficult to control for and that can distort the shape of the exposure-response relationship, such as the "healthy worker survivor" effect (Steenland and Stayner 1991; Steenland *et al.* 1996; Kolstad and Olsen 1999).

Methods used for performing risk assessments have also been a major source of uncertainty and controversy. USEPA and other agencies have used the linear multistage model (Crump *et al.* 1976, 1977) for cancer risk assessment. This model has been under considerable attack over recent years for its failure to consider possible effects of carcinogens on cell growth and differentiation, and for ignoring alternatives such as the "two-stage clonal expansion model" (Moolgavkar *et al.* 1980; Moolgavkar and Knudson 1981; Moolgavkar 1994). In response, the USEPA (1996) has developed draft guidelines for cancer risk assessment to address these issues. However, the fact that these guidelines have been under review for the last 5 years reflects the degree of debate over this issue. Similar debates exist over current methods for assessing noncancer risks, particularly over the continued use of the no observed adverse effect level (NOAEL) and uncertainty factors for determining "safe" levels of exposure and alternative methods have been proposed (Bailer *et al.* 1997; Hattis 1998; Hattis *et al.* 1999; Hattis *et al.* 2002).

Controversy also surrounds risk assessment because it provides the scientific basis for regulations that have major social and fiscal implications. Groups most affected by these regulations frequently raise questions about either the data and/or the methods used in risk assessments as a means of either strengthening or weakening the proposed regulation. The net effect of these debates has often been to delay the finalization of a risk assessment and associated regulatory actions. Diesel exhaust particulates (DEP) is a classic example (see Figure 1) of how difficult the risk assessment process has become for many regulatory agencies (Stayner *et al.* 1999). The USEPA initiated its efforts to assess the potential lung cancer risk associated with environmental exposures to DEP before 1980 (Albert *et al.* 1979, Albert, 1983). In 1987 the USEPA formally reinitiated its efforts, and after four drafts the risk assessment was just recently approved for finalization by their scientific advisory committee and it is anticipated that this assessment will be published in the next few months. Thus, it has taken the USEPA more than 20 years to complete its risk assessment for DEP. The USEPA risk assessment for dioxin has taken nearly as long and has also not yet been finalized.

The process of risk assessment and regulatory actions has been equally difficult in the occupational arena. To illustrate this, the number of occupational permissible exposure limits (PELs) set by OSHA since its existence is presented in Figure 2. OSHA set a relatively large number of RELs in the 1970s with a peak of 15 standards in 1974. There appears to have been a clear drop off in standard setting after the benzene Supreme Court case in 1980, which made the quantification of risks and benefits a requirement for setting standards. Only 10 standards have been set since 1980. In the last 4 years, OSHA only finalized one standard, the ergonomics rule (CFR, 29CFR 1910.900). However, this rule was just recently overturned by Congress under the Congressional Review Act of 1996. The development of NIOSH Recommended Exposure Levels (RELs) has been nearly as slow, as illustrated in Figure 3 with only 6 new RELs set in the 1990s.

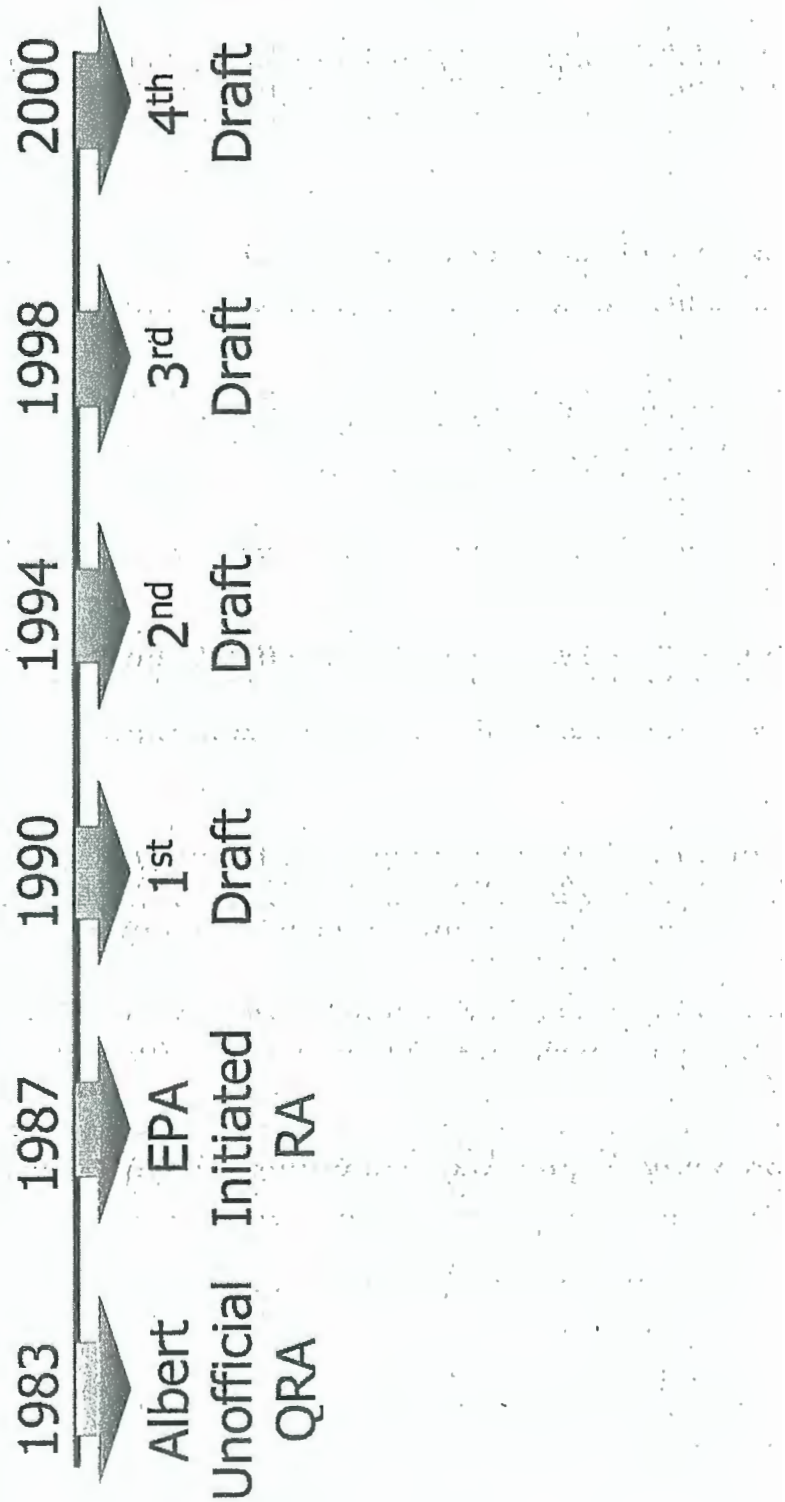


Figure 1: History of EPA Diesel Risk Assessment

Figure 2: Number of OSHA PELS set since passage of the Occupational Safety and Health Act in 1970

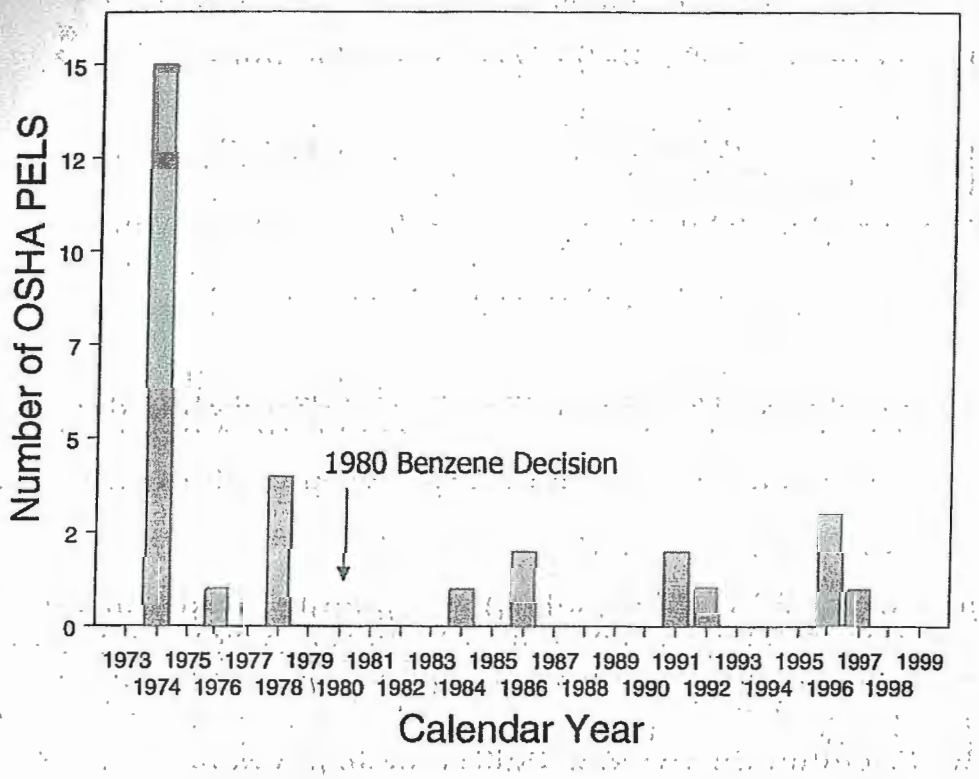
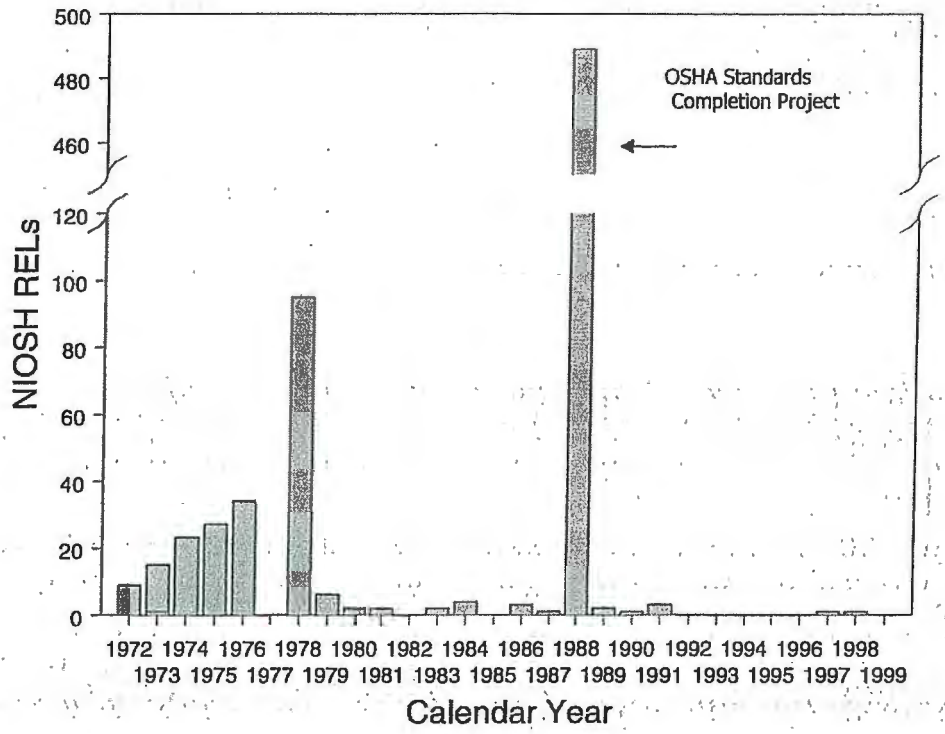


Figure 3: Number of NIOSH RELs set since passage of the Occupational Safety and Health Act in 1970



The extremely large amount of resources and time that agencies have had to invest in developing risk assessments has led some to question the utility of risk assessments for regulatory decision making (Silbergeld 1993). Recent environmental legislation in the European Union has emphasized the use of what has been referred to as the "precautionary principle" as an alternative to risk assessment for setting regulations (Epstein 2000). Simply put, the precautionary principle emphasizes that when there is insufficient evidence to characterize risk then one should set standards that err on the side of protecting the public health. This is really not a new concept, and in fact is essentially the philosophy that was the basis for much of the regulatory action in the U.S. prior to the development of the formal requirements for risk assessment in the 1980s. Some in the U.S. risk assessment community have reacted negatively to the precautionary principle as a substitute for risk assessment, in part because it does not provide either information needed to juxtapose expected costs and benefits of different policy options, or to judge the fairness (equity) of the distributions of benefits and risks to different parties that are expected to result from different public policy choices (Hattis and Anderson 1999; Graham 1999).

Finally because risk assessment often spans several different disciplines with different analytical paradigms and different traditions for what counts as "good" information, there have been important philosophy of science and even ethical disputes as scientists trained in different fields have misunderstood or misinterpreted work and the information standards used by others (Hattis and Smith 1987; Hattis 2000).

OPPORTUNITIES

Given what some might consider a crisis in the current state of affairs of risk assessment and risk management in our country, it seems an appropriate time to consider what as researchers we could do to resolve some of the issues discussed above. In order to identify the research opportunities that would address these issues, a workshop was convened on August 16 to 18th, 2000, in Aspen, Colorado, on "Future Research for Improving Risk Assessment Methods." The primary objective of this workshop was to bring together prominent scientists in the field of risk assessment and related sciences (*e.g.*, epidemiology, toxicology, industrial hygiene and statistics) to assist in the development of a national agenda for research that would enhance risk assessment methodology. The first few papers in this journal were from presentations intended to identify issues with current risk assessment methods, and to broadly identify opportunities for research solutions to these issues. The last set of papers in this issue are from three workgroups on mice (toxicology), men (epidemiology), and models (toxicokinetics and dose-response). These workgroups were charged with developing specific research ideas that could improve our ability to perform risk assessments in the future that better reflect our scientific understanding, and are more helpful and informative for decision making.

It would be misleading to suggest that further research is all that is needed to solve our problems with risk assessment. Nonetheless, there are significant research opportunities that may improve the current situation. The most significant new development is the burgeoning new area of genomics which was the subject of one of the

presentations in this special issue (Morgan 2002). The recent announcement of the successful mapping of the genome (Lander *et al.* 2001) is clearly going to usher in a whole new era in our understanding of the molecular basis for diseases including those induced by environmental agents. Future risk assessments will need to address how the risk associated with a particular agent are modified by genetic characteristics. Current mechanistic models for cancer (*e.g.*, multistage and 2-stage clonal expansion) will need to be modified to fit our increasingly complex knowledge of the carcinogenic process. Toxicologic bioassays may be improved so that they more accurately predict human risk, and require less time and resources to perform. Handling the vast amount of information generated from the high output DNA assays will present a challenge for risk assessors, and will require the development of new methods.

The explosion of information available for risk assessments in the future will also provide an even greater burden on risk assessors to develop methods that are not overly complex. Silbergeld (2002) has a clear warning to the risk assessment community that current risk assessments are already too complex, which contributes to distrust on the part of the general public. In developing new methods we should bear this warning in mind and remember the principle of Ockham's razor, which for risk assessment might state that the simplest model that adequately explains a phenomenon is probably the most useful. On the other hand, we might also consider that Einstein reportedly said that theories should be as simple as possible, but no simpler. The balance between faithfulness to our mechanistic understanding and simplicity in describing limited available data is addressed by Krewski *et al.* (2002) in these proceedings. Clearly, with the explosion of genetic and other mechanistic information that will be available to us, striking this delicate balance is probably the greatest challenge that risk assessors will face in the future.

REFERENCES

Albert RE. 1983. Comparative carcinogenic potencies of particulate from diesel engine exhausts, coke oven emissions, roofing tar aerosols and cigarette smoke. Environ Health Perspect 47:339-41

Albert RE, Pasternack BS, Shore RE, *et al.* 1979. Identification of occupational settings with very high risks of lung cancer. J Natl Cancer Inst 63:1289-90

Ames BN and Gold LS. 1990. Too many rodent carcinogens: Mitogenesis increases mutagenesis. Science 249:970-1

Bailer AJ, Stayner LT, Smith RJ, *et al.* 1997. Estimating benchmark concentrations and other noncancer endpoints in epidemiology studies. Risk Anal 17(6):771-80

CFR (Code of Federal Regulations). USA29CFR 1910.900. Office of the Federal Register. Government Printing Office. Washington, DC, USA

Crump K. 2001. Modeling lung cancer risk from diesel exhaust: suitability of the railroad worker cohort for quantitative risk assessment. Risk Anal 21:19-23

Crump K, Hoel D, and Peto R. 1976. Fundamental carcinogenic processes and their implications for low dose risk assessment. Cancer Research 36:2973-9

Crump KS, Guess HA, and Deal KL. 1977. Confidence intervals and test of hypotheses concerning dose response relations inferred from animal carcinogenicity data. Biometrics 33:437-51

Dawson SV, Alexeeff GV. 2001. Multi-stage model estimates of lung cancer risk from exposure to diesel exhaust, based on a U.S. railroad worker cohort. Risk Anal 21:1-18.

Epstein SS. 2000. Legislative proposals for reversing the cancer epidemic and controlling runaway industrial technologies. Int J Health Serv 30:353-71

Graham JD. 1999. Making sense of the precautionary principle. Risk Perspective 7:1-6

Hattis D. 1998. Strategies for assessing human variability in susceptibility, and using variability to infer human risks. In: Neumann DA and Kimmel CA (eds), Human Variability in Response to Chemical Exposure: Measures, Modeling, and Risk Assessment, pp 27-57. CRC Press, Boca Raton, FL, USA

Hattis D. 2000. Draft risk analysis ideals. Human Ecol Risk Assess 6:913-9

Hattis D and Anderson E. 1999. What should be the implications of uncertainty, variability, and inherent 'biases'/'conservatism' for risk management decision making? Risk Anal 19:95-107

Hattis D and Smith J. 1987. What's wrong with quantitative risk assessment? In: Humber JM and Almeder RF (eds), Quantitative Risk Assessment, Biomedical Ethics Reviews: 1986, pp 57-105. Humana Press, Clifton, NJ, USA

Hattis D, Banati P, and Goble R. 1999. Distributions of individual susceptibility among humans for toxic effects–for what fraction of which kinds of chemicals and effects does the traditional 10-fold factor provide how much protection? Annals NY Acad Sci 895:286-316

Hattis D, Baird S, and Goble R. 2002. A straw man proposal for a quantitative definition of the RfD. Drug Chem Toxicol (*in press*)

Kolstad H and Olsen J. 1999. Why do short term workers have high mortality? Am J Epidemiology 149(4):347-52

Kreski D, Brand KP, Burnett RT, *et al.* 2002. Simplicity *vs.* complexity in the development of risk models for dose-response assessment. Human Ecol Risk Assess (*this issue*)

Lander ES, Linton LM, Birren B, *et al.* 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921

Moolgavkar SH. 1993. Cell proliferation and carcinogenesis models: general principles with illustrations from the rodent liver system. Environ Health Perspect 101(suppl 5):91-4

Moolgavkar SH. 1994. Biological models of carcinogenesis and quantitative cancer risk assessment. Risk Anal 14:879-82

Moolgavkar SH and Knudson AG Jr. 1981. Mutation and cancer: a model for human carcinogenesis. J NCI 66:1037-52

Moolgavkar SH, Day NE, and Stevens RG. 1980. Two-stage model for carcinogenesis: Epidemiology of breast cancer in females. J Natl Cancer Inst 65:559-69

Morgan KT, Brown HR, Benavides G, *et al.* 2002. Toxicogenomics and human disease risk assessment. Human Ecol Risk Assess (*this issue*)

Silbergeld EK. 1993. Risk assessment: the perspective and experience of U.S. environmentalists. Environ Health Perspect 101:100-4

Silbergeld EK. 2002. An NGO perspective on risk assessment and scientific research. Human Ecol Risk Assessment (*this issue*)

Stayner L. 1999. Protecting public health in the face of uncertain risks: The example of diesel exhaust. Editorial. Am J Public Health 98(7):991-3

Steenland K and Stayner L. 1991. The importance of employment status in occupational cohort mortality studies. Epidemiology 2(6):418-23

Steenland K, Deddens J, Salvan A, *et al.* 1996. Negative bias in exposure-response trends in occupational studies: Modeling the healthy workers survivor effect. Am J Epidemiology 143(2):202-10

USEPA (U.S. Environmental Protection Agency) 1996. Proposed guidelines for carcinogen risk assessment. Fed Reg, April 23, 1996:1790-01

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1203-1221 (2002)

The Joy Before Cooking: Preparing Ourselves to Write a Risk Research Recipe

Adam M. Finkel
Directorate of Health Standards,[1] U.S. Occupational Safety and Health Administration, Room N- 3718, 200 Constitution Avenue, NW, Washington, DC 20210; Tel(voice):202-693-2256, Tel(fax):202-693-1658; adam.finkel@osha.gov

ABSTRACT

Rather than the conventional practice of compiling a list of interesting research projects and then attempting to make the case that each represents a high priority, I will attempt an approach rooted in decision analysis. The information of greatest value, according to decision theory, is that which most enables us to make more reliable, transparent, and cost-effective decisions. Therefore, I begin with a brief discussion of how and why typical decisions relying on cancer dose-response information can fall short, in an attempt to assess where and how this aspect of risk assessment is "broken" before generating a list of research projects to "fix" it. I discuss the problem of model uncertainty in dose-response assessment, and conclude it is impossible to gauge how valuable it might be to know the correct model until we agree on guidelines for how to make decisions given imperfect information in this regard. After discussing four broad research areas that arguably represent particularly high priorities given this framework, I conclude by identifying three overarching areas of risk assessment and management that, if not given commensurate attention, threaten to render even perfect dose-response information of dubious value.

Key Words: uncertainty, value of information, risk assessment guidelines, research priorities, holistic approach.

INTRODUCTION

Tradition dictates that speakers at conferences about research needs for risk assessment present a list of knowledge gaps that, if filled, would make the control

1 Currently Regional Administrator (Region VIII), U.S. Occupational Safety and Health Administration, 1999 Broadway, Suite 1690, P.O. Box 46550, Denver, CO 80201-6550. The views expressed here are those of the author and not necessarily those of OSHA or the U.S. Department of Labor. This manuscript is considered to be a work of the U.S. Government and therefore is not copyrighted.

1080-7039/02/$.50

of environmental and occupational hazards tidy and uncontroversial. I will present a short list at the end of this talk, so as not to disappoint (unless the substance of my suggested priorities does that for me). But first, I want to try to do something I hope will prove more useful: to step back and ask what we are trying to fix here—and why—before unrolling the laundry list of desirable research projects. Inasmuch as we are all going off to breakout groups to vote on which research questions are most meritorious, it might help to have a common vision (or at the very least an appreciation of the diversity of possible visions) of what kind of party we're planning before we vote on the menu.

To tackle the "why" question first, I start from the premise that for risk assessment to be more useful and beneficial, it needs to perform well along various dimensions:

- Producing estimates of both individual risk (probability of harm) and population risk (aggregate harm) that are reliable enough to lead us toward sound decisions—both with regard to what hazards to address (priority-setting) and how to address them;

- Allowing individual decision-makers (citizens, consumers, workers, parents) and social decision-makers (regulators, judges, legislators) to choose the best decision *according to preferences they specify;*[2]

- Allowing consumers of risk assessment information, who are willing to put in some effort of their own, to understand the strengths and limitations of the analysis itself, and to assess what information might be worth deferring a decision in order to obtain.

These attributes—to oversimplify, "accuracy," "honesty" and "transparency"—are indispensable and make all the analytic work worthwhile, even though they sometimes conflict with each other and can each be taken to undesirable extremes (Hattis 2000).

A SENSE OF WHERE WE ARE

To debate how risk assessment research can improve the *status quo*, and to narrow down which research initiatives deserve highest priority, we also need a snapshot of what 30 years of building a risk assessment system has brought us. Let me begin sketching this out with five generalizations, designed as all generalizations are to capture something fundamental while admittedly doing some violence to the details.

1. *Current QRA methods often predict population risk with surprising accuracy.* I refer here to the few data sources we have to compare predictions of excess human risk derived from animal experimentation to actual enumeration of excess morbidity or mortality. Such comparisons, as the U.S. Environmental Protec-

2 I contrast this to the decision that is only best given a hidden, and perhaps manipulative set of preferences (such as extreme risk-aversion or (more commonly) the strange belief that errors resulting in needless expenditures to control risks are exactly equivalent to errors causing needless human suffering or ecological damage, "suitably monetized").

tion Agency (USEPA) notes in its current draft Guidelines for Carcinogen Risk Assessment, generally show that rather than a one-sided parade of false positives and gross exaggerations of the magnitude of risk, epidemiology (and exposure measurement) often corroborate the expectations derived from toxicology (and exposure modeling) (Allen *et al.* 1988; Dedrick and Morrison 1992). A similar (though more qualitative than quantitative at present) pattern can be discerned from data comparing some non-cancer effects between rodents and humans (Takihi *et al.* 1994; Olson *et al.* 2000). That the deviations from perfect predictive power tend to balance out, with roughly as many errors of "conservatism" as its opposite, attests to a set of rules and assumptions that, while certainly not of uncanny accuracy, are not systematically "rigged" either to exaggerate or to gloss over the risks to humans from exposures we study. Given the many physiologic similarities among mammalian species (and the long track record of using these test animals to gauge the efficacy and side-effects of pharmaceuticals), and the firm scientific foundation of exposure modeling in various environmental media, the usefulness of current risk assessment methods should not be surprising. I only use that adjective here to emphasize how surprising these data might be to someone acquainted only with the thousands of articles, editorials, and other pronouncements of the many observers determined to expose and rectify the purported "hyper-conservatism" of risk assessment.

2. *Human beings vary substantially in their susceptibility to carcinogenesis and other toxic processes.* The combination of genetic differences (in metabolism, DNA repair, immunocompetence, *etc.*), concurrent diseases, and lifestyle factors make for a tremendous diversity in the individual risks each of us would face upon exposure to an identical toxic stimulus. A few very preliminary estimates, based either on models of the influence of these various factors or on analysis of cancer age-incidence data, tend to agree with each other and to suggest that a substantial minority of persons (above and beyond those afflicted with visible predisposing conditions such as Down Syndrome) may be 50 to 100 times more susceptible than the average person (with a similar fraction equally less susceptible) (Finkel 1995a; Hattis and Barlow 1996). Related studies of interindividual variability in susceptibility to non-cancer effects suggest a similar degree of heterogeneity in the human population (Hattis *et al.* 1999).

3. *Therefore, current QRA procedures understate risk to individuals of above-average susceptibility.* I offer this as a truism—the only logical conclusion if statements 1 and 2 above are true. If we have a black box that spits out a reasonable prediction of population risk (which is merely a scalar multiple of the individual risk to the average person), then it does so by simultaneously overestimating risks to persons of low susceptibility and underestimating risks to those of higher-than-average susceptibility. The USEPA, unfortunately, has completely garbled this truism in its proposed Carcinogen Risk Assessment Guidelines, proclaiming that one plus one equals zero instead of two. The guidelines (p. B-5 of the 1999 version, the most recent version available on the USEPA Web site, at http://www.epa.gov/ncea/raf/crasab.htm) assert that the Agency

does not need to consider adjusting its cancer risk assessments to account for individuals of above-average susceptibility, because "linear extrapolation is sufficiently conservative to protect public health," but then cites as *support* for this proposition evidence that "linear approaches... from animal data are consistent with linear extrapolation on the same agents from human data (Goodman and Wilson 1991; Hoel and Portier 1994)." Of course, if linear extrapolation indeed yields reasonable predictions of risk to the individual of average susceptibility, it cannot also be sufficiently conservative to protect persons of above-average susceptibility!

4. *When QRA does miss the mark, it misses badly, due to faulty model assumptions.* The "half full" optimism of the first two generalizations above needs to be tempered with the sober observation than on occasion, we have learned how error-prone risk assessment can be. Certain substances which would appear to be likely human carcinogens turn out to cause animal tumors via processes unlikely or impossible to occur in people, while in other cases (notably cigarette smoke), relying on exculpatory evidence from bioassays would have been folly. In exposure assessment, models are rarely qualitatively wrong, but still we sometimes fail to apply the right models in the right situation—which is really what happens whenever we end up saying "we found quantities of substance X when we didn't expect to find them" or "we measured the environment or human beings and didn't find any substance X even though our predictions said we would."

5. *When such "model uncertainty" looms large, it is "bad science" to conceal the controversy, either by refusing to consider credible alternatives to a default model or by rushing to declare the default model "dead."* I hope this is an uncontroversial statement, and that anyone particularly aggrieved by one of these two types of abuses will be able to acknowledge that the opposite problem occurs no less frequently in public discourse about chemical risks. We tend to either give short shrift to promising new ways to look at the science or else switch—too soon for some critics, too late according to others—to a new model with no allowance for the possibility that the old approach may still be quite credible. It stands to reason that except in the rare cases where one definitive publication "rocks the world," data accrue to support a theory (and in so doing cast doubt on currently-used assumptions) gradually rather than instantaneously. This suggests that instantaneous shifts from not mentioning a new approach to not mentioning the old one are failures of risk communication.

I will later provide some recommendations for how we might accommodate new model information when our knowledge paints a picture in shades of gray rather than black or white. However, I strongly believe that one way **not** to manage model uncertainty is so tempting and so wrong-headed it deserves mention as a corollary. Taking two or more fundamentally incompatible scientific theories and "averaging them together"—that is, concocting a risk prediction based on a "probability"-weighted sum of the estimates each model produces—is worse than arbitrarily using

one theory only, because the former practice can be said to yield a "best estimate" of risk, a siren's song if ever there was one.

Even if you believe in expected-value decision-making (and volumes have been written on the limitations of "the greatest good for the greatest number" as a precept), surely you should at least know how to choose correctly under this framework. Unfortunately, proponents of "decision-making by expected-value" have thoroughly confused two basic concepts and are thus unable to choose (or advise) correctly. The decision that yields the greatest expected value over an uncertain spectrum of possible outcomes is generally *not* the decision that follows if the outcome was known with certainty to be equal to its expected value (which is the trap "best estimate" decision-making leads one into). To belabor a parable I've used to excess elsewhere (Finkel 1996), if competing scientific theories predict either that a hurricane brewing in the Gulf of Mexico will turn west and hit New Orleans or turn east and hit Tampa, it makes no sense to "average" the predictions together and evacuate (or warn) the residents of Mobile.

The only way to determine "the greatest good for the greatest number" in such a case is to compare the costs of incorrectly evacuating Tampa given the probability that the hurricane will hit New Orleans, versus the costs under the converse situation. The interplay of probabilities and costs (in turn related to the number of lives that might be lost in either city, the difficulty of evacuating residents of either location, *etc.*) determines which of the two sensible decisions is superior "on average." For this reason, the 1994 National Academy of Sciences committee charged with evaluating USEPA's risk assessment methods warned against indiscriminate use of central-tendency estimates, especially when the main thing we don't know is which of two models is correct. We recommended in *Science and Judgment in Risk Assessment* that "EPA should try to quantify the parameter and other uncertainty that exists for each plausible choice of scientific model," but should not force all these uncertainties "under one roof," and certainly not present only a single point estimate concocted out of this misapplication of decision theory. Note that we thought it sufficient to stress the logical fallacy inherent in "model averaging," despite the more obvious and well-studied concern about the processes that would have to be used to derive the subjective weights assigned to each model.[3]

STRUCTURED RESEARCH

If this snapshot is even roughly correct, then I would argue that before we can identify worthwhile research endeavors, we need to develop guidelines for how to identify, communicate, and (most importantly) react to uncertainties in risk—both of the quantitative variety (the sum of parameter uncertainty and inter-individual

3 When, as is often the case, the scientific controversy reduces to a choice between two models, one that predicts substantial risk at current exposure levels and one that predicts zero risk, the "best estimate" if the models are to be averaged becomes entirely a function of the fraction of "experts" surveyed who believe the former model is correct. Thus, the choice of a single expert over another in a group of 10 could cause the "average risk" to be cut by half.

variability) and qualitative variety (the fundamental model choices referred to above). Such "ground rules" would enable decision-makers to:

- Find the strength to avoid "paralysis by analysis"—either the familiar inability to choose if and how to control a hazard because there will always be another study to commission, or the utter catatonia of not even being able to start reducing uncertainty out of despair over how intractable it seems;

- Identify the critical contributors to uncertainty, in order to estimate how much more robust the apparently best available control decision would be if each component uncertainty was resolved. This estimation relies on the "quantitative value-of-information" approach developed several decades ago (Raiffa and Schlaifer 1961; Morgan and Henrion 1990) and more recently applied to the environmental risk assessment paradigm (Finkel and Evans 1987; Dakins *et al.* 1996);

- Provide sensible incentives for needed research to occur, grounded in a sober assessment of who controls resources and what motivates them to allocate resources to research rather than to any other use; and

- Empower risk managers to favor "erring on the side of safety" (or to favor a strategy biased toward protecting financial interests, or any other calculus) in the full light of day.

For me, the dilemmas surrounding model uncertainty are particularly more interesting and deserving of elaboration here—but I would be remiss not to mention that several years ago, the Presidential/Congressional Commission on Risk Assessment and Management threw a wrench into what seemed to be a growing consensus about the need to better quantify parameter uncertainty. The Commission recommended that analysts **not** routinely compute and report uncertainty in risk, on the grounds that this would needlessly complicate matters, and confuse decision-makers and citizens who would mistakenly tend to assume that all values were equally likely within an interval given as the plausible endpoints of the magnitude of a risk. This latter objection, if valid, could be easily remedied by a bit of explanatory material, but I believe the Commission fell into a more fundamental misunderstanding. Quantitative uncertainty analysis (QUA) does not have to burden decision-makers and the public with ranges and probability density functions rather than point estimates of risk; done properly, it is instead a tool for replacing erroneous point estimates with valid ones.[4]

4 Curiously, the Commission was bullish on the need to rigorously quantify interindividual variability in exposure at the same time it threw cold water on quantifying uncertainty in potency or inter-individual variability in susceptibility ("The Commission strongly supports using mathematical descriptions of variability, particularly distributions of a population's possible contaminant exposure concentrations," p.89). Since the additional risk communication complexity is no different, I wonder whether this inconsistency can be explained by the hope that full distributions of exposure would reveal how conservative current methods of point estimation are in this part of the process (coupled perhaps with the concern that full distributions of potency, if brought to light, might reveal something quite different).

Assuming for the sake of argument that the recipient of risk information cannot cope with the distribution or range emerging from a QUA, all the analyst has to do is provide a point estimate from out of this distribution, preferably after discussing what type of estimator ("typical" or median value, expected value, rather conservative, extremely conservative, *etc.*) the recipient desires. Without having performed a QUA, s/he can still gin up a desired point estimate, of course—agencies do this all the time, as when they multiply and divide a bunch of "typical" estimates together and somehow declare the result to be the "best estimate" (or worse, when they combine a string of purportedly upper-bound estimates together, then "back away" by some arbitrary factor and declare the result to be either an average value or a "reasonably high-end estimate.") Contrary to the spirit of the Commission's recommendation, a decision-maker who can't be bothered with more than a single point estimate does not have to "dumb down" the entire process—all she/he needs to do is ask the analyst whether the point estimate proffered has consciously been chosen from a distribution acknowledging uncertainty, or has merely been labeled a central or other type of estimate without any assurance this is appropriate.

An extreme view at the opposite extreme (that *any* use of a point estimate ruins a risk assessment) came recently from consultant Warner North, who testified before Congress in 1997 that OSHA's methylene chloride regulation should be invalidated because the Agency used a point estimate of individual risk (and a corresponding estimate of the expected number of cancer deaths averted) at various places within several hundred pages of analysis. North offered this opinion even though he acknowledged that OSHA had made it quite clear (through narrative, tabular, and graphical presentations) that it had chosen these estimates only after having undertaken an exhaustive and transparent QUA whose validity North did not question. Assuming this was a genuine expression of a general view that displaying any one point estimate is always an error that no amount of analysis can overcome, I can see why the Commission might have felt there was some zealous worship of complicated QUA out there that they had to combat.

Returning to model uncertainty, the first crossroads in developing ground rules for coping with it is to decide whether new scientific information should be fodder for a process that combines all credible models together, or should instead determine which models shall have primacy over others. We are all familiar with the latter system, wherein we start with a set of "default" models (one for each inference point in the risk assessment) and switch to one or more alternative models when it seems the right thing to do. In addition to the misuse of expected-value decision-making and the precarious nature of eliciting the subjective weights (see above), the other approach to model uncertainty also degrades research—it may simply be easier to influence the weights by warping the process rather than by truly advancing knowledge. The binary approach has its problems, but they are self-made, as we've never had a clear formula for implementing it—when improved, this approach is a clear winner over the construction of hybrid estimates. A better binary system needs, above all, two things it has never really had: (1) a commitment to clearly explaining (either in a generic document or *de novo* each time a risk assessment is published) the scientific and policy rationale for each of the default models used; and (2) a set of clear criteria—preferably published prior to the risk assessment in which they are used—that establish the nature and quality of research needed to abandon or

modify each one of the default models and adopt some new alternative. Without the forthright description of how we came to follow a "default pathway" through the branching maze of inference points, we play into the hands of those who denigrate the defaults as bureaucratic rather than scientific rules, or as "purely policy-driven." Without an explanation of the various sound theoretical bases for (and strong empirical support for) linear extrapolation from relatively high exposures to a carcinogen (exposures causing roughly 5 to 50 percent tumor incidence) to exposures 10- to 100-fold lower, for example, the naïve reader of a risk assessment might be excused for believing the propaganda that this practice amounts to "ruler toxicology."

More importantly, without the up-front criteria for telling the scientific community what constitutes sufficient evidence to overturn a default in favor of an alternative, agencies will deservedly face severe criticism that each such decision is *ad hoc*—succumbing to persuasive evidence can be derided as "giving away the store," just as holding on to a time-tested default in the face of some preliminary (and dubious) new hypothesis can be blasted as being immune to common sense. *We in the agencies owe it to those who want to understand our decisions, and especially to those who want to conduct fruitful research to influence our decisions, to spell out why we will act a certain way in the absence of evidence to the contrary, and what we will need to be persuaded to change our minds.*

To be sure, such criteria are inextricably bound up in policy choices. The height of the hurdle required to abandon a default model and replace it with an alternative model can, in simplest terms, be very low—the general premise being something like "move off the default whenever there is any evidence (of any quality) that the alternative is reasonable"—or very high, ranging up to a rule such as "abandon the default only when airtight evidence has proven beyond a shadow of a doubt that the alternative is correct and the default is incorrect." Within the infinite shades of gray between these two extremes, we could adopt philosophies ranging from "a non-trivial chance the alternative is more reasonable than the default," through various gradations of wanting "persuasive," "compelling," or "clear and convincing" (but not airtight) evidence before switching approaches.

Most of us believe that by and large, the current set of risk assessment defaults tend to be somewhat "conservative," although I believe it quite noteworthy that the 1994 National Academy of Sciences committee (which encompassed experts who worked for the chemical manufacturing industry as well as others who worked for or with environmental advocacy groups) expressed the consensus view that because cancer risk assessment currently treats all humans as having identical susceptibility, a significant "missing default" needs to be addressed for us to be confident that the outputs of traditional cancer risk assessment do in fact tend to err on the side of overestimation. The NAS panel went even further, and affirmed that a system based on "conservative" defaults is the right way to begin—that the more we know, the less protective we need to be. Our committee could have questioned this logic or conceivably even endorsed the mirror-image process, wherein we would start with an "anti-precautionary" bias and allow or encourage research that would substitute models that "erred against the side of safety" with conservative ones—more on this later. But as night follows day, the acceptance of a system that *starts* with a conservative stance means that the decision about how high an evidentiary hurdle we

should erect to evaluate new science *becomes a referendum on how reluctant or how eager we should be to relax (what we hope is) this built-in conservatism.*

Here the *Science and Judgment* committee reflected the deep division within the expert community and offered two sharply contrasting sets of views. I worked with some committee members to draft a chapter recommending "plausible conservatism" as an organizing principle for deciding when to switch to a more sophisticated and less conservative model, while Roger McClellan and Warner North worked with others to elaborate on the "maximum use of scientific information" as a desired guiding principle. The chapter I wrote basically argued for a less stringent version of a "clear and convincing" standard—that we need not be convinced that an alternative model is correct in order to adopt it instead of the default, only convinced that it is sufficiently more reasonable than the default to render the default no longer plausible by comparison. The other minority report advocated switching to an alternative model whenever the scientific community agreed it was more reasonable—greatly so or slightly so, no matter—than the default. The USEPA Cancer Guidelines seem to be in a constant state of flux, but in the 1996 draft, the Agency has essentially signaled its acceptance of the latter viewpoint—"if data support an alternative to the default as the more reasonable judgment, the data are [to be] used" (p. 22, emphasis added).[5]

We ought to be greatly concerned, I assert, about the practical consequences for health and environmental protection of adopting a "more reasonable judgement" standard rather than a somewhat more stringent one that embraces an alternative model only once a better case has been made than "it's slightly better than the default." The difference between "a 51/49 preponderance of the evidence" and "a 90/10 preponderance" may have profound consequences, and *I personally believe it is reckless to allow minor (and highly subjective) differences in relative plausibility to completely trump major differences in the degree to which our analysis-driven environmental control decisions tend to provide adequate protection in the face of uncertainty.* As with the confusion over "best estimates," it may be comforting to allege that using the "most reasonable" model is simply "being agnostic about the extent of conservatism and letting science drive the outcome," but this is again an affirmative (though concealed) statement that potential errors of overestimation are *exactly* as unwanted as the converse. To throw out the "49 percent possibility" (and thus deprive the decision-maker of the chance to take precautionary action) on the grounds that "science made us do it" is what I would have called a "cop-out" had I been born a few years earlier.

We should also be concerned about the implications for scientific research of one type of "departure from defaults" regime or another. Obviously, a draconian standard requiring absolute proof would quickly stifle research, as both human nature

5 The previous sentence of the Guidelines document ("If data support a plausible alternative to the default, but no more strongly than they support the default, both the default and its alternative are carried through the assessment and characterized for the risk manager") (emphasis added) seems clearly to address the special case where two equally plausible models exist. In context, therefore, "more reasonable" clearly refers to any situation where the alternative is deemed even infinitesimally more plausible than the default.

and financial realities weigh against conducting or funding research that has little or no chance of ever being accepted. But an overly permissive standard threatens research quality almost as severely. After all, why build a boat painstakingly, filling in all the cracks, shoring up weak areas that might not fail rapidly but will eventually, when the buyer has already told you that it only has to look slightly better than the old boat he already owns and no more questions will be asked? Consider, for example, the default assumption that tumors produced in experimental animals are considered relevant to humans. In a last-ditch attempt to block OSHA's pending regulation of methylene chloride, an industry trade association petitioned us to declare that observed tumor responses in mice were irrelevant to humans, on the grounds that the metabolizing enzyme had been found in the nuclei of some mouse cells but purpotortedly could be shown predominantly in the cytoplasm of analogous human cells. With several years' hindsight, I remain confident that OSHA would have correctly rejected this hypothesis (for a wealth of reasons detailed in the preamble to the 1997 final rule, 62FR 1494-1619, esp. pp. 1517-1529) even if we had used an explicit "the more reasonable model shall prevail" standard. However, there would have been no impetus for the research sponsors to even try to patch the numerous holes in their arguments had we accepted this alternative without pointing out what additional questions needed to be answered.

I also surmise that a permissive standard applied to new research would stifle exploration in another, more subtle way. The risk-and-reward calculus that underlies decisions to direct research toward improving chemical-specific risk assessment must be quite sensitive to the expected "shelf life" of changes in analyses and decisions prompted by successful acceptance of the completed studies. A system that promises to switch gears when one model is marginally more reasonable than another is, after all, by definition a system that will just as easily abandon new Model B in favor of new Model C (or go back to old Model A) when the next smidgen of information is brought to light. In other words, "most reasonable" is a fickle standard that could well discourage "suitors" considering investing in research.

One formidable obstacle remains in the path of developing any explicit standard, however high or low the evidentiary bar—the claim that it is impossible for risk assessment agencies to develop scientific criteria that would flesh out in detail how they would determine an alternative model to be "more reasonable," "compelling," or whatever. This is USEPA's rationalization in the Cancer Guidelines for refusing to choose between the two pathways the National Academy of Sciences offered (or to craft a third approach of its own). First, the Agency put its own "spin" on what the Committee recommended, referring to it as a call to "adopt a list of formal decision criteria in the sense of a checklist" (it puzzles me how a request that the USEPA should state whether it advocates a "clear and convincing" standard or some other hurdle for accepting new models could morph into a "checklist"). The USEPA then

6 To me, this makes about as much sense as saying "because every suspect in a murder case will have to be judged guilty or innocent based on a unique set of facts, we can't decide whether the standard of proof ought to be 'beyond a reasonable doubt'." Not only has our system of laws made that decision, but centuries of case law has clarified how juries might think about what makes a doubt "reasonable."

denigrates this straw man, claiming that because "risk assessments are highly variable in content and purpose... no uniform checklist will fit all cases."[6]

Such a claim does not stand up even to cursory scrutiny. The judgments made using criteria developed *a priori* will, of course, create new "case law" with each specific application, but the criteria themselves can certainly be developed—indeed, they already have been. OSHA confronted this question in 1995 when I suggested that the time had come to set an OSHA Permissible Exposure Limit based on a pharmacokinetic adjustment of the dose-exposure relationship across species rather than the traditional allometric adjustment (again, methylene chloride was the case in point). Before we made our decision, I initiated a discussion of what sorts of evidence, of what quality, we would need in the general case to rule that the PBPK approach was clearly more credible than the generic default. Out of this discussion we developed 11 criteria (see Appendix) that amply specify how OSHA would weigh this question in subsequent rulemakings. In the methylene chloride case, we ended up reasonably confident that researchers had already met the standard of proof along all 11 dimensions, and we calculated the cancer potency using a PBPK model.

Now at least for this one paradigm choice (allometry versus PBPK), researchers and their sponsors will know as they build a case for a "departure from default" what questions OSHA deems important, what data are needed to complete the case, and what must be done to sufficiently rule out alternative interpretations of the new data collected. OSHA recently began working on developing similar sets of criteria for several other recurring choices we will be called upon to make between traditional and novel scientific models. For example, we are discussing the recurring question of what constitutes sufficient reason to dismiss animal tumor responses as irrelevant to people, to adopt a non-linear dose-response extrapolation, or to use an exposure duration parameter other than the standard 45-year working lifetime.[7]

A FEW IDIOSYNCRATIC RESEARCH PRIORITIES

Having sketched out the prerequisites for determining which risk assessment research efforts are likely to produce the most valuable results, I cannot take the next step and identify specific priorities without information—quantitative estimates of uncertainties and their effect on how precarious our "best guess" environmental protection decisions are—that the decision-analytic framework demands but that our collective wisdom cannot yet supply. I will, however, offer four general avenues of research that appear to me to be likely candidates for a "most valuable research in cancer potency" list.

1. **How broad is the distribution of human interindividual variability in susceptibility to carcinogenesis?** I hope we all understand that "risk" is not a group concept, but applies to individuals and is determined by the combination of exposure and susceptibility unique to each individual. The more we know about the reasons for such differences, the better we will be able to fulfill two absolutely crucial functions of risk assessment and management: (1) determining what exposure

7 As I recently ceased directing the health regulatory arm of OSHA to become a Regional Administrator, I do not know what will become of this endeavor.

levels are actually necessary to protect (*i.e.*, reduce cancer risk to an "acceptably low" probability) "most or nearly all" of the exposed population—in contrast to current risk-based levels, which I assert are only designed to protect the person of average susceptibility; and (2) communicating to specific individuals (or giving them the raw materials to determine for themselves) what we think their individual risk actually is, based on our knowledge of their exposure plus knowledge of particular genetic or other factors that render them more or less susceptible than the "reference man." I suggest we greatly step up our research efforts both into determining the distribution (and pinpointing an individual's place within that distribution) of predisposing factors that would likely affect risk to carcinogenic stimuli in general (*e.g.*, inborn or acquired variations in the competence of the physiologic processes that detect and combat incipient tumors) as well as factors specific to particular stimuli (*e.g.*, inborn or acquired variations in the concentrations of particular enzymes that activate xenobiotics to proximate carcinogens or detoxify them). I also suggest that we not exclusively focus on quantifying the influence of specific factors we already suspect are prominent in affecting susceptibility (albeit a very important task, since an X-fold difference between two individuals in some factor may imply any degree—including zero—of difference in the persons' overall susceptibilities). Instead, we might well supplement this work with some thoughtful "fishing expeditions" that worked backward from individuals who appear to be highly susceptible or highly resistant to cancer, and catalog the phenotypic and other differences between them in an effort to discern the important determinants of susceptibility. For example, I have long wondered if and in what way(s) lifelong heavy smokers who do not develop lung cancer even by an advanced age differ from young adults whose lung cancers (non-adenocarcinomatous lesions) developed seemingly after very few pack-years of exposure.

2. **How much error is introduced by testing substances in the standard rodent bioassay?** "Mice are not little men," the inane saying goes, but do they predict cancer risk as if they were? One difference that tends to make bioassay results underestimate human risk is that we care about human exposures that begin at birth (or *in utero*) and can extend for 80 or more years, whereas we begin exposing rodents after weaning (the equivalent of missing the first several years of a human life) and sacrifice them at 24 months (10 or more human-equivalent years before the end of the natural lifespan). Research that enabled us to lengthen the duration of the bioassay at both extremes, or that allowed us to estimate how many more tumors would be produced if we did, would be of great value in making mice act more like little men. (Peto *et al.* 1991). On the other hand, a systematic catalog of physiologic differences between rodents and humans—not just what proteins or enzymes one species has that the other doesn't (or has much more or much less of), but how the species differ functionally—could greatly demystify the contentious process of deciding that a positive rodent response should not be relied upon as indicative of equal (or any) human risk.

3. **When (if ever) does a series of exposures to different carcinogenic stimuli over a lifetime or working lifetime not yield similar cumulative risk to an**

equivalent exposure to a single substance? This is not an exposure-assessment question, but a knowledge gap in toxicology; it has taken on great practical importance with the emphasis on survey data quantifying what we already know—that people move from home to home and job to job during their lifetimes. Some have argued (Hamilton, *et al.* 1997) that if an average person changes residences seven times in seventy years, then a risk assessment assuming seventy years of being a "porch potato" exaggerates risk sevenfold. But relaxing all exposure standards by a factor of seven to account for this would simply and perversely increase individual risk by the same factor, unless one of two conditions held: either only one 10-year period of exposure conferred any risk at all, or each period "restarted the clock" such that the variety of risks broadened but the magnitude did not increase. The former condition is clearly absurd—there would be "no escape" if all standards are relaxed in the name of adjusting for duration, and more to the point, adjusting them in one arena (say the Superfund program, as Hamilton *et al.* advocate) ignores the grim reality that the people most exposed to those hazards in the first place are those least likely to be able to move at all or (what counts) move to pristine (*i.e.*, expensive) locations.[8] If, however, exposures in location B did not operate by related mechanisms to the exposures already accrued earlier in location A, then adjusting for transience might have some merit. Bioassays that compared two consecutive 12-month exposures to substances (with different degrees of structural similarity) to a 2-year exposure (historical data?) to each substance singly might shed valuable light on this question.

4. **Where on the continuum between "high" and "low" (non-zero) doses might significant non-linearities in dose-response occur?** The field of risk assessment has been very concerned about "the" shape of particular dose-response functions, intent on discovering which are linear, which are steeper at high doses than at lower ones, and which exhibit threshold behavior. Some analysts have focused on the behavior of these functions at vanishingly low doses, with critics of traditional multistage extrapolation arguing it is implausible to posit that a single molecule can present any non-zero risk, and defenders of this default pointing out sound theoretical reasons why risk should often rise in proportion to dose even in the limit of low dose. This stark controversy has frustrated regulators enough that the current version of the draft USEPA Cancer Guidelines abandons risk quantification altogether in favor of a "margin of expo-

[8] The analogy to the occupational setting would be failing to appreciate that a worker exposed to high levels of a given chemical is likely to move (if at all) to a new employer and situation where exposures are identical or similar. Sandblasters who report a change of employer very often remain sandblasters, and are unlikely to become stockbrokers. Hence the conclusion of Burmaster (2000) that "the assumption often dictated by federal and state environmental agencies that all people work in all jobs for 30 years is false and misleading" misses the mark badly. A worker who merely changes employer several times during a working lifetime (which is all that the data Burmaster analyzed can reveal) may well be exposed to 30, 45, or more years of exposure to the same (or functionally equivalent) chemical exposures.

sure" (MOE) approach whenever "the curve is thought to be non-linear, based on [assumptions about] mode of action."[9]

But we hardly need to build a road to infinity in order to be fairly confident that we are on a useful path down the proverbial block. In cases where lifetime risk can reasonably be assumed to be the sum of a vast number of molecular encounters with DNA, each one of which carries with it a very small probability of yielding a particular unrepaired mutation that pushes a cell irreversibly along the pathway to tumorigenesis, surely the difference between the (say) 10^{28} molecules a test animal encounters during a bioassay and the (say) 10^{26} molecules an exposed human could encounter during a lifetime has very little to do with "extrapolation to zero." We don't need a referendum on whether "one molecule can cause cancer" to be concerned about being this close (relative to zero) to exposures that cause "mini-epidemics" of cancer in the laboratory. The interesting questions, which do lend themselves to risk quantification rather than the MOE approach, involve cases where something important happens as exposure is reduced this relatively tiny increment toward the vanishing zero, by virtue of discontinuities in fundamental biological processes. If, for example, X ppm saturates a detoxifying metabolic pathway, leading to some production of a dangerous metabolite, but X/2 ppm results in 100 percent of the substance being detoxified, then linear extrapolation even over a factor of 10 could be misleading. Research—difficult as it may well be—to shed light on when "10^{26}" molecules per lifetime is qualitatively different from "10^{28}" may be our only avenue for answering questions that matter. "Does 'ruler toxicology' beginning at a TD_{10} and heading toward zero miss important non-linearities or discontinuities in dose-response, in the range of exposures relevant to humans?"—this is a research question worthy of our attention.

PARTING SHOTS

At meetings such as these, I realize that it is easy to grouse about how the organizers have defined the topic too narrowly and stifled efforts to step back and discuss "the forest rather than the trees." At the time the Aspen conference was held, the phrase "the weakest link" had not yet become grating, so I would note that if dose-response assessment was in fact the weakest link in the chain of events leading to the control of environmental hazards, it would be efficient to meet and discuss research needs in this area alone to the exclusion of others. To the contrary, I believe that if we do draw some concentric circles around the larger problem, dose-response may be at the core but is far from the area needing the most attention. As we expand our view, let me offer three broader areas where "weaker links" can be found:

1. *exposure assessment.* It may seem odd to caution that the tidy world of exposure assessment, where analytical devices have errors in the percent range or less

9 As of this writing, the USEPA has not begun to craft any interpretive guidance that might help decision-makers and the public cope with its having abdicated the responsibility to produce estimates of risk and uncertainty for this important class of situations. How should the consumer react to the news that "we can't estimate your risk—all we can tell you is that an exposure of X units produces an apocalyptically high tumor incidence in animals, and that your exposure is less than X by an 'MOE'?"

and physical/chemical models are well-specified, may be a shakier ground than the "black box" of dose-response assessment. I would suggest, however, that we understand exposures more in theory than in practice, and certainly more as single-substance snapshots than as integrated profiles of encounters with multiple substances over time. Even if you draw the semantic boundary so that variations among humans in uptake and metabolism count as uncertainties in effective potency rather than variabilities in true exposure, the remaining variations in ambient concentrations are often substantial and not well studied. Nowhere is this more evident than as in the workplace. For less than a handful of substances (really only lead and crystalline silica) is the probability significantly different from zero that a given U.S. worker will ever have had even one eight-hour exposure measurement taken during her career (unless she works for one of the few companies that systematically collects such information on its own and discloses it to those sampled). For example, in the 12 months ending October 1999, OSHA took 17 samples nationwide for carbon monoxide exposure, 607 samples from among the 250,000 workers exposed to methylene chloride, 204 samples for perchloroethylene exposure from among the nation's 30,000-plus dry-cleaning establishments, and zero samples for such substances of concern as vinyl chloride, methyl-*tert*-butyl ether, and pesticides such as aldrin, chlordane, heptachlor, *etc.* Having attended numerous interagency meetings where the paucity of exposure data has been recognized, I'm hopeful that help may be on the way. However, based on the current proposals I've seen, it is no exaggeration to predict that within several years, we will probably know more about the body burdens of toxic chemicals in domestic cats and dogs that we will about the same burdens in U.S. industrial workers.

2. *regulatory economics.* It is the rare environmental control decision where costs are not considered, either explicitly or implicitly. And yet we sit here, polishing the proverbial chrome on our system for quantifying risks, while the parallel system for estimating costs is in disarray, with little impetus for improvement or signs thereof. No one can say for sure whether the typical error in cost estimation is larger or smaller than that in risk estimation, but several valid generalizations do bear on this issue: (1) the potential errors in cost estimation dwarf those in risk estimation—except for rare and speculative cases where hormesis may occur, risk analysts may misidentify a situation posing zero risk as risky, but not mistake a net minus for a net plus. But because economists almost invariably consider compliance costs (the amount paid out by regulated entities as a result of a government intervention) rather than the relevant measure of social cost (the sum of all changes in economic welfare, negative and positive, set in motion by the intervention), they can easily report as a cost something that actually benefits the economy as well as the environment (Finkel 1995b); (2) whereas risk assessors have been accused (often falsely, in my opinion) of a systematic bias toward overestimation, regulatory economists have a clear track record of producing exaggerated estimates—the evidence for this claim is persuasive and growing more voluminous each year (OTA 1995: Hodges 1997; Goodstein and Hodges 1997), and the natural causes of

this bias (*e.g.*, incentives not to underestimate cost on the part of regulated firms and agency economists and lawyers concerned with court challenges for underestimating cost, failure to account for economies of scale and technological improvements), are ingrained and recalcitrant to change; and (3) uncertainty analysis—the raw material for revealing the possible extent of error and pointing the way to data or research to reduce it—has barely begun to catch on among regulatory economists. Economists have begun surrounding their benefits estimates with ranges (generally to account only for controversies in how to assign monetary values to effects), but only the very best cost analyses ever include ranges (and never to my knowledge have included probability density functions that indicate expected values and other estimators) to be discerned. At OSHA, I've seen cost analyses presented as "correct" to the nearest penny! This glaring disparity between the worlds of risk and cost in willingness to acknowledge uncertainty seems to be a "vicious circle" of overconfidence and lack of guidance. Why else would the last dozen "regulatory reform" bills drafted by Congress be chock-full of rules to micro-manage the process and substance of risk estimation, without a single admonition that economic analysis demands at least equal attention to transparency, "sound science," peer review, disclosure of assumptions, *etc.*? This "free pass," I suspect, has emboldened economists to the point that recently several workshops have been convened to "bridge the gap" between risk assessors and economists by considering only ways in which the former group could improve its procedures to make them more useful to the latter. Obviously, I think something is seriously askew here, and that risk assessors should resist these ambushes and instead seek to expose and correct the "garbage in the numerator" problem that stymies efforts to make sensible social decisions via cost-per-lives-saved and similar metrics.

3. *risk management.* Bernie Goldstein has often made the astute observation that (to paraphrase) we obsess over "fixing" risk assessment to the detriment of the weaker link of risk management. Of the many ways in which risk management arguably is "broken," my recent experiences as a government official have alerted me to a set of problems stemming from a reductionist orientation to our mission. By undertaking risk assessment and risk management on substances rather than processes, I believe we are both creating new problems and missing opportunities for more fundamental progress. Examples abound of controls on one substance impelling regulated entities to substitute a more dangerous, unregulated substance in response. Currently, for instance, I am concerned that OSHA's 1997 methylene chloride regulation will, if further interventions are not forthcoming, encourage firms in foam fabrication and many other sectors to switch to *n*-propyl-bromide, unregulated despite clear evidence from animal studies (and clear human evidence on an isomer contained in the commercial material) of potent reproductive toxicity. The pitfalls at the USEPA of focusing on a single substance, wherein different agency programs acting at different times have the net effect of moving exposures from one environmental medium to another and back again, have been well documented (Davies and Mazurek 1988). More recently, USEPA and OSHA

jointly convened a workshop to discuss a growing set of cases where environmental controls applied to single substances have the effect of reducing emissions from measured point sources at the expense of increased concentrations inside workplaces. These three problems—adverse substitution, intermedia pollutant transfers, and environment-workplace transfers—are untoward outcomes that a less reductionist approach could anticipate and avert. The more far-reaching benefits of coordinated action aimed at improving industrial processes rather than simply reducing particular exposures could include cheaper control strategies (as companies are permitted to engineer compliance with air, water, workplace, and other standards simultaneously rather than in fits and starts), resolutions that satisfy both the dictates of cost-benefit analysis and the preferences of industries and local citizens, and multiple risk reductions that surpass our limited expectations.

Ultimately, dose-response assessment is an input to risk assessment, which in turn is raw material for risk-based decision-making. Research to refine dose-response assessment has value of its own, but realizing its full value depends on creating a system so that we can use risk assessment to discriminate among a rich set of possible solutions to the web of health, safety, welfare, and economic problems that continue to plague modern society.

REFERENCES

Allen BC, Crump KS, and Shipp AM. 1988. Correlation between carcinogenic potency of chemicals in animals and humans. Risk Anal 8:531-44

Burmaster DE. 2000. Distributions of total job tenure for men and women in selected industries and occupations in the United States, February 1996. Risk Anal 20(2):205-24

Dakins ME, Toll JE, Small MJ, *et al.* 1966. Risk-Based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information. Risk Anal 16(1):67-79

Davies JC and Mazurek J. 1988. Pollution Control in the United States: Evaluating the System. (ISBN 0-915707-87-X). Resources for the Future, Washington, DC, USA

Dedrick RL and Morrison PF. 1992. Carcinogenic potency of alkylating agents in rodents and humans. Cancer Res 52(9):2464-7

Finkel AM. 1995a. A quantitative estimate of the variations in human susceptibility to cancer and its implications for risk management. In Olin S, *et al.* (eds), Low-Dose Extrapolation of Cancer Risks: Issues and Perspectives, pp 297-328. International Life Sciences Institute, Washington, DC, USA

Finkel AM. 1995b. A second opinion on an environmental misdiagnosis: The risky prescriptions of *Breaking the Vicious Circle.* New York University Environ Law J 3:295-381

Finkel AM. 1996. Who's exaggerating? Discover :48-54, May

Finkel AM and Evans JS. 1987. Evaluating the benefits of uncertainty reduction in environmental health risk management. J Air Pollution Control Assoc 37(10):1164-71

Goodstein E and Hodges H. 1997. Polluted data: Overestimating environmental costs. The American Prospect (Nov-Dec):64-9

Goodman G and Wilson R. 1991. Quantitative prediction of human cancer risk from rodent carcinogenic potencies: a closer look at the epidemiological evidence for some chemicals not definitively carcinogenic in humans. Reg Toxicol Pharmacol 14(2):118-46

Hamilton JT, Viscusi WK, and Dockins PC. 1997. Conservative versus mean risk assessments: implications for superfund policies. J Environ Economics Management 34:187-206

Hattis D. 2000. Draft risk analysis ideals. Human Ecol Risk Assess 6:913-9

Hattis D and Barlow K. 1996. Human interindividual variability in cancer risks–Technical and management challenges Human Ecol Risk Assess 2:194-220

Hattis D, Banati P, and Goble R. December 1999. Distributions of individual susceptibility among humans for toxic effects–For what fraction of which kinds of chemicals and effects does the traditional 10-fold factor provide how much protection?" Annals NY Acad Sciences 895:286-316

Hodges H. 1997. Falling Prices: Cost of Complying with Environmental Regulations Almost Always Less Than Advertised. Briefing paper. #199711. Economic Policy Institute, Washington, DC, USA

Hoel DG and Portier CJ. 1994. Nonlinearity of dose-response functions for carcinogenicity. Environ Health Perspect 102(supp 1):109-13

Morgan MG and Henrion M. 1990. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, NY, NY, USA

Olson H, Betton G, Robinson D, *et al.* 2000. Concordance of the toxicity of pharmaceuticals in humans and animals. Reg Toxicol Pharmacol 32(1):56-67

OTA (Office of Technology Assessment, U.S. Congress). 1995. Gauging Control Technology and Regulatory Impacts in Occupational Safety and Health–An Appraisal of OSHA's Analytic Approach. OTA-ENV-635. U.S. Government Printing Office, Washington, DC, USA

Peto R, Gray R, Brantom P, *et al.* 1991. Dose and time relationships for tumor induction in the liver and esophagus of 4080 inbred rats by chronic ingestion of N-nitrosodiethylamine or N-nitrosodimethylamine. Cancer Res 51:6452-69

Raiffa H and Schlaifer R. 1961. Applied Statistical Decision Theory. Harvard University Press, Cambridge, MA, USA

Takihi N, Rosenkranz HS, Klopman G, *et al.* 1994. Structural determinants of developmental toxicity. Risk Anal 14(4):649-57

APPENDIX
OSHA'S 11 CRITERIA FOR ACCEPTING A PHARMACOKINETIC MODEL (FROM THE 1997 METHYLENE CHLORIDE FINAL RULE).

(1) The predominant and all relevant minor metabolic pathways must be well described in several species, including humans.

(2) The routes of metabolism must be adequately modeled.

(3) There must be strong empirical support for the putative mechanism of carcinogenesis (*e.g.*, genotoxicity) and the proposed mechanism must be plausible.

(4) The kinetics for the putative carcinogenic metabolic pathway must have been measured in test animals *in vivo* and *in vitro* and in corresponding human tissues at least *in vitro*, although *in vivo* human data would be the most definitive.

(5) The putative carcinogenic metabolic pathway must contain metabolites which are plausible proximate carcinogens (for example, reactive compounds such as formaldehyde or S- chloromethylglutathione).

(6) The contribution to carcinogenesis via other pathways must be adequately modeled or ruled out as a factor. For example, there must be a reasonable analysis of why reactive metabolites formed in a second pathway would not contribute to carcinogenesis (*e.g.*, formyl chloride produced via the MFO pathway is likely to be too short-lived to be important in MC carcinogenesis).

(7) The dose surrogate in target tissues (lung and liver in the case of MC) used in PBPK modeling must correlate with tumor responses experienced by test animals.

(8) All biochemical parameters specific to the compound, such as blood:air partition coefficients, must have been experimentally and reproducibly measured. This must be true especially for those parameters to which the PBPK model is most sensitive.

(9) The model must adequately describe experimentally measured physiological and biochemical phenomena.

(10) The PBPK models must have been validated with data (including human data) which were not used to construct the models.

(11) There must be sufficient data, especially data from a broadly representative sample of humans, to assess uncertainty and variability in the PBPK modeling.

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1223-1228 (2002)

Topics in Hazard Identification: Oxygenated Fuels, Safety Assessment, Hematological Neoplasms, and the Precautionary Principle

Bernard D. Goldstein[1]
Director, Environmental and Occupational Health Sciences Institute (EOHSI), 170 Frelinghuysen Road, Piscataway, New Jersey 08854; Tel(voice):732-445-00205, Tel(fax):732-445-0131; bgold@eohsi.ruters.edu

ABSTRACT

Hazard identification is based upon the second "law" of toxicology, the specificity of toxic effects caused by a chemical agent. Specificity reflects the differential reactivity inherent in chemical structure and in the biological niches in which chemicals interact. Just as Paracelsus is identified with the first "law" of toxicology, the dose makes the poison, Paré, a century French surgeon, should be credited with an early formulation of the second "law" of toxicology, the specificity of chemical effects. I discuss a number of aspects of hazard identification, including issues related to oxygenated fuels, to routine safety assessment, to the interpretation of hematological neoplasms and to the Precautionary Principle.

Key Words: butadiene, methyl tert-butyl ether, multiple myeloma, non-Hodgkins lymphoma.

INTRODUCTION

I have chosen to discuss aspects of the hazard identification step of risk assessment in large part because the distinguished presenters and panelists at this session seem to cover every other available topic. This choice also reflects my concern that we are losing sight of the important scientific principle that underlies hazard identification — the specificity inherent in the chemistry of an agent and in the biology of its receptor.

We are all aware that very slight changes in structure can greatly alter the reactivity of a chemical, and we know of numerous instances in which important biological niches are exquisitely sensitive to seemingly minimal changes in the structure of chemical compounds with otherwise similar reactivity. This specificity of chemical effects is as much a

[1] Current address: Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, 15261; goldstein@gsphdean.gsph.pitt.edu

1080-7039/02/$.50

"law" of toxicology as is the dose makes the poison. My colleague Michael Gallo and I have suggested that the name Paré be associated with this second "law" of toxicology, just as Paracelsus is associated with the first "law," the dose makes the poison (Goldstein and Gallo 2001). Paré is the father of experimental surgery. He lived in 16th century France, where his willingness to question accepted theories of medical care put him at odds with the academic medical authorities of his time, but his ability to test and prove his hypotheses soon made him famous. This fame led him to be asked about the anti-poisoning efficacy of a bezoar recently bought at great expense by the King of France. A bezoar is a concretion from the stomach of goats, which was believed in the Middle Ages to be an antidote against all poisons. Paré told the king this was not possible.

> I answered, that nature could not admit it; for neither have all poysons the like effects, neither doe they arise from one cause; for some worke from an occult and specifick property of their whole nature, others from some elementary quality which is predominant (Paré 1984).

As a test, a criminal condemned to be hanged accepted the option of being poisoned and receiving the bezoar antidote, which, unfortunately for the prisoner, did not work.

OXYFUELS

Hazard identification is under attack from different directions. We are hearing about the need to treat every chemical as if it is a carcinogen, or assuming that a chemical that can cause cancer in one organ system can do so in all organ systems. We are also told, as in the example of oxygenated automotive fuels, that because one compound is potentially without an effect then a chemically related compound is surely harmless and does not need thorough testing.

Since controversy began in the winter of 1992 about possible adverse health and environmental consequences of methyl tert-butyl ethers (MTBE), it has been clear that more information was needed on other ether oxyfuels, such as tert-amyl methyl ether (TAME) and ethyl tert-butyl ether (ETBE), both because they were already being used as oxyfuels and because any phase out of MTBE would result in an increase in their use. Yet rather than beginning in 1993 to obtain the necessary toxicological information, this information is only now being required. Instead, there was a dependence on structure-activity relationships (SAR) which, even if one could argue that MTBE was harmless, flew in the face of the many examples in which just one methyl group could make an enormous difference. Just consider benzene vs. toluene, n–hexane vs. n-heptane or ethanol vs. methanol. I do not question that SAR can be a highly valuable technique, particularly with newer computer-based approaches (Zhang *et al.* 1997). But it is foolhardy to expose perhaps 100 million Americans to an agent as ubiquitous as a gasoline component with an incomplete consideration of its toxicity (Franklin *et al.* 2000, Erdal and Goldstein 2000).

SAFETY ASSESSMENT

We clearly need better tools for hazard assessment. There is insufficient recognition that our most useful approaches to determining the potential of a chemical to

produce a specific adverse effect come from improved basic science understanding of the mechanisms of chemical toxicity. The Ames test, for example, could not have been developed without an understanding of the mechanisms of chemical mutagenesis and of the relationships between mutagenesis and carcinogenesis. It was thus somewhat disappointing that the otherwise positive agreement between the U.S. Environmental Protection Agency (USEPA), environmental groups, and the chemical industry to perform testing on the thousands of untested chemicals in commerce depends so heavily on routine safety assessment.

I suggest a simple paper experiment. Choose a dozen or so chemicals that are of major public health concern, *e.g.*, lead, arsenic, benzene, DDT, mercury, DBCP. Based on the literature, analyze whether the toxicity of these human problem chemicals would be picked up the battery of safety assessments studies being employed to assess existing untested chemicals. I may be wrong, but I suspect that at least a few of these known problem chemicals would slip past our current standard safety assessment procedures. This outcome would argue for investment of resources in the development of the mechanistic understanding needed for better safety assessment so that we can be more protective than with our current routine safety assessment methodology.

MOLECULAR BIOLOGY OF HEMATOLOGICAL NEOPLASMS

The organizers of this conference have asked for specific research ideas. In keeping with my theme of the importance of hazard identification, I suggest that it is time to use advances in molecular biology to sort out issues presented by hematological neoplasms.

The central issue exemplifies the old medical argument between the lumpers and the splitters. The splitters have carried the day in two recent controversial analyses of potential chemical carcinogenicity. In the case of butadiene, the alleged lack of concordance between the two major positive epidemiological studies, one reporting lymphoma and the other reporting leukemia, played a major role in a recent review by the International Agency for Research on Cancer (IARC), which narrowly voted butadiene to be a probable rather than a known human carcinogen (IARC 1999). In the case of MTBE, the finding of a statistically significant increase in all hematological neoplasms in female rats exposed lifetime was not considered of regulatory significant because the increase for none of the individual hematological tumor types was statistically significant (Belpoggi *et al.* 1998; USEPA 1999). These decisions would be understandable, and in my view supportable, if in both cases the tumors were of completely different organ systems, *e.g.*, leukemia and kidney cancer instead of leukemia and lymphoma.

Benzene presents a somewhat different issue in hazard identification. There is no question about it being a cause of human acute myelogenous leukemia (AML). But there is controversy concerning whether benzene can cause other hematological cancers (Goldstein 1990; Goldstein and Witz 1999). In my view, it is highly probable that benzene causes acute lymphatic leukemia, non- Hodgkin's lymphoma and multiple myeloma and various myeloproliferative cancers. This view is based on epidemiological as well as toxicological findings supportive of this relationship. For the lymphatic tumors evidence includes the sensitivity of lymphocytes to benzene,

the readily detectable presence of chromosomal abnormalities in lymphocytes of exposed individuals, and the fact that a carcinogenic metabolite of benzene reaches the bone marrow resulting in AML. Yet the findings are not quite at the level of scientific proof due in large part to the inherent weakness of our methodology. A recent blatant example of a misleading approach are two meta-analyses of petroleum workers looking at the incidence of multiple myeloma (Bergsagel *et al.* 1999; Goldstein and Shalat 2000) and non-Hodgkin's lymphoma (Wong and Raabe 2000; Goldstein and Shalat 2001). Both were said to be negative but the level of benzene exposure was sufficiently trivial that the same cohorts would not have had an increase in their incidence of AML. A study purporting to look at the relationship between cigarette smoking and AML would be considered irrelevant if the level of cigarette smoking was too low to cause a measurable increased incidence of lung cancer.

Central to resolution of the controversies concerning all three of these pollutants is an improved understanding of the relationship between myelopoietic and lymphopoietic cells. There has been a longstanding uncertainty about the extent to which these two hematological cell types are related to each other through a pluripotential stem cell. Whether they are or not, and whether this is relevant to hazard identification for agents that produce hematological tumors, can and should be addressed.

PRECAUTIONARY PRINCIPLE

My last point about hazard identification concerns the Precautionary Principle (Goldstein 1999). This principle is increasingly in use as a means to regulate environmental threats. A reasonably standard definition derived from the 1989 Rio Declaration is:

> Nations shall use the precautionary approach to protect the environment. Where there are threats of serious or irreversible damage, scientific uncertainty shall not be used to postpone cost- effective measures to prevent environmental degradation (United Nations 1992).

Past and proposed actions that could fit under the Precautionary Principle suggest that the value of hazard identification may be lost to the detriment of effective environmental control. I give two examples.

In the United States the 1990 Clean Air Act Amendments (CAAA) radically shifted regulatory control strategy for so-called hazardous air pollutants (HAP). Until then, the Clean Air Act required the USEPA to make an active determination that there was sufficient evidence that a chemical could produce adverse health effects. Once this was determined, regulatory activities were risk based, targeted only at those sources whose elimination or control would be effective in reducing risk. Congressional frustration at the slow rate of regulatory control led to the 1990 CAAA specifically listing more than 180 pollutants to be controlled. In keeping with the Precautionary Principle the burden of proof was shifted from the need for evidence to prove adverse effects to be listed, to the need for evidence of no adverse effects to warrant removal from the list. Secondly, regulatory control is based on

maximal available control technology (MACT) with risk being relegated to a secondary consideration. Ten years later it is still uncertain whether this new approach to HAPs has been more effective. The process has not been as simple as proponents believed.

We need to study more throughly whether ignoring hazard in the control of HAPs is an effective control approach. The key issue for the present discussion is whether this shotgun approach is effective; whether treating known carcinogens such as benzene no differently from reasonably well-studied compounds for which there is no evidence of effects at the levels likely to be present in worst-case community air pollution situations is of value. Also of note is the issue of whether writing specific regulations for current MACT inhibits further development of even better control technology. By maintaining a secondary risk-based approach in which additional measures are required if MACT does not sufficiently reduce risk, the 1990 CAA amendments in essence recognize and attempt to cope with a limitation inherent in an approach solely based on the Precautionary Principle.

A more egregious example of ignoring the specificity of effects of chemical agents is the campaign, in essence, to ban any organic chemical with a chlorine atom, irrespective of hazard. Using the Precautionary Principle as the basis for such a sweeping ban suggests that hazard identification is being threatened by at least some proponents of the Precautionary Principle.

REFERENCES

Belpoggi F, Soffriti M, and Maltoni C. 1998. Pathological characterization of testicular tumours and lymphomas-leukemias, and their precursors observed in Sprague-Dawley rats exposed to methyl tertiary-butyl ether (MTBE). Eur J Oncol 3(3):201-6

Bergsagel DE, Wong O, Bergsagel PL, *et al.* 1999. Benzene and multiple myeloma: appraisal of the scientific evidence. Blood 94:1174-82

Erdal S and Goldstein BD. 2000. Methyl tert-butyl ether as a gasoline oxygenate: Lessons for environmental public policy. Ann Rev Energy Environ 25:765-802

Franklin PM, Koshland CP, Lucas D, *et al.* 2000. Clearing the air: Using scientific information to regulate reformulated fuels. Environ Science Technol 34:3857-63

Goldstein BD. 1990. Is exposure to benzene a cause of human multiple myeloma? Annals NY Acad Science 609:225-334

Goldstein BD. 1999. The precautionary principle and scientific research are not antithetical. Editorial. Environ Health Perspect 107:12-594-95

Goldstein BD and Gallo MA. 2001. Paré's law: The second law of toxicology. Toxicological Sciences 60:194-5

Goldstein BD and Shalat SL. 2000. (Letter) The causal relation between benzene exposure and multiple myeloma. Blood 95:4

Goldstein BD and Shalat S. 2001. (Letter) To the Editor of JOEM. J Occup Environ Med 42-12:1133-6

Goldstein BD and Witz G. 1999. Benzene. In: Lippman M (ed), Environmental Toxicants: Human Exposures and Their Health Effects, chapter 4, 2nd ed, pp 121-50. John Wiley & Sons, NY, NY, USA

IARC (International Agency for Research on Cancer). 1999. Re-Evaluation of Some Organic Chemicals, Hydrazine and Hydrogen Peroxide (Part One). IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, vol 71. IARC Press, World Health Organization, Geneva, Switzerland

Paré A. 1984. Bezoars. In: Bogdonoff MD, *et al.* (eds), The Apologie and Treatise Containing the Voyages Made into Divers Places with Many of His Writings upon Surgery, pp 199. The Classics of Medicine Library, Division of Gryphon Editions, Ltd., Birmingham, AL, USA

United Nations. 1992. Conference on Environment and Development. Earth Summit. Rio Declaration on Environment and Development, Rio de Janeiro, Brazil. United Nations Environment Programme, NY, NY, USA

USEPA (U.S. Environmental Protection Agency). 1999. SB521 Research Project Review: Executive Summary, Health and Environmental Assessment of MTBE. Office of Mobile Sources, Office of Air and Radiation, Review of Volumes I-IV, Washington, DC, USA

Wong O and Raabe GK. 2000. Non-Hodgkin's lymphoma and exposure to benzene in a multinational cohort of more than 308,000 petroleum workers, 1937 to 1996. J Occupat Environ Med 42:554-68

Zhang YP, Macina OT, Rosenkrantz HS, *et al.* 1997. Prediction of the Metabolism and toxicological profiles of gasoline oxygenates. Inhalation Toxicol 9:237-54

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1229-1242 (2002)

Does the Emperor Have Any Clothes: Using Mechanistic Information or Doing Houdini Risk Assessments?

Franklin E. Mirer
Health and Safety Department, International Union, UAW, 8000 East Jefferson Avenue, Detroit, MI 48214; Tel(voice):313-926-5563, Tel(fax):313-824-4473; FMIRER@uaw.net

ABSTRACT

Risk assessment research rarely quells controversy. Mega-mouse, and mega-rat, experiments contradicted a threshold for carcinogenesis, yet thresholds are still argued. High to low dose continuity of response from cigarette smoking to environmental tobacco smoke, and from occupational asbestos exposure to take-home asbestos, contradict thresholds in people. Nevertheless, mechanistic hypotheses allege "Houdini Risk Assessments", which make risks disappear or allow industries to escape from protecting workers. Despite concerns for animal-to-human extrapolations, priority occupational exposures with sufficient or substantial evidence of carcinogenicity in people not addressed by new exposure limits include silica, sulfuric acid mist, chromates, diesel particulate matter, particulate matter generally, metalworking fluids, welding fume, and formaldehyde. "Houdini Risk Assessments" are exercises in "anti-hypothesis generation": ignore selected tumor sites and types; ignore data from people (as with formaldehyde and diesel); choose the most resistant species in laboratory tests; select biochemical parameters in which the most resistant species resembles people; assume a mechanism that gives threshold or steep exposure response for carcinogenic effect; and reduce estimated people risk by the parameter ratio to the most resistant species. NORA research should focus on quantitative reconciliation of laboratory and epidemiology studies, and develop a counter "anti-hypothesis" generation research agenda for key exposure circumstances.

Key Words: carcinogenesis, Houdini, risk assessment, silica, diesel, chromate, particulate.

1080-7039/02/$.50

Risk assessment research is distinct from research into risks. Risk assessment research includes methods development and analyzes the body of available data on specific substances to support decision rules to assess the risks of additional substances or exposures, but does not usually include investigations of risks of specific substances. Certain studies of model compounds, discussed below, are examples of risk assessment research.

Our national research agenda, and particularly our public sector research agenda, must redress a distinct imbalance that inhibits public health progress in the occupational environment. These views are informed by the United Auto Workers' (UAW's) experience, including participation in a dozen major Occupational Safety and Health Administration (OSHA) standards. The UAW settled with industry lawsuits over OSHA's formaldehyde (OSHA 1991) and methylene chloride (OSHA 1998) standards through negotiations with industry by refusing to address industry-advanced, unproven and likely incorrect mechanistic hypotheses that supported Houdini (Figure 1) risk assessments. Had the UAW and OSHA reengaged on these issues, we would still be litigating instead of having negotiated protections workers needed, and management could live with.

I also come here bruised by being on the losing end of votes on the carcinogenicity of phthalates at the February 2000, International Agency for Research on Cancer (IARC) Working Group.

Risk assessment research rarely quells controversy. The most famous, but nearly forgotten example is the mega mouse experiment, the mother of all bioassays, which would settle the controversy over whether there was a threshold[1] for carcinogenesis. (Littlefield *et al.* 1980)

The study found no apparent threshold for liver tumors, the possibility of a threshold for bladder tumors, and was ignored in the subsequent debate.[2] Subsequently dueling statistical analyses evaluating the threshold have obscured any conclusions about the significance of the study.[3]

---

[1] A threshold is a dose below which there is no dose response relationship, where increasing dose has no increasing risk.

[2] The conclusion of the authors was: "Although bladder neoplasms exhibited a minimum effect level (or a nonlinear response) for specific conditions, the total results were consistent with a "no threshold concept. The late appearing liver neoplasms displayed a nearly linear type response that extrapolated directly to zero dose." Others may have interpreted the results differently. This study also demonstrated the importance of time and mortality adjusted analysis: "The liver neoplasms appeared very late in the study but were shown to be induced at a very early point in the exposures and did not require the continuous presence of the carcinogen in order to develop. A standard 18-month bioassay study, if conducted under the same conditions, would have classified this chemical as a weak acting carcinogen. These studies demonstrate the importance of the time factor in safety evaluation or risk assessment in carcinogenesis." We note that National Toxicology Program (NTP), as a result, pioneered such analysis, while these methods are rarely used elsewhere, weakening the sensitivity and specificity of chronic bioassays. Recently, investigators conducting bioassays of methyl t-butyl ether, and butadiene metabolites, have acquired a taste for 18-month rather than two-year bioassays.

[3] The study used a model compound, 2-acetylamino fluorene, of no economic significance. Thus, it didn't impact any exposures.

Figure 1. Houdini.

Similarly, a mega rat experiment found complete linearity to the lowest doses (Peto *et al.* 1991), and was also ignored.[4]

Two paradigm examples of high-dose to low-dose potency of carcinogens in people are continuation of response from direct cigarette smoke to environmental tobacco smoke, and from occupational asbestos exposure to take-home asbestos.[5]

---

[4] The authors' conclusion was: "At these low dose rates, the number of liver (but not of esophageal) neoplasms induced by treatment was simply proportional to the dose rate. This finding is not surprising, since the background incidence of liver (but not of esophageal) neoplasms was appreciable. The linear relationship observed at low dose rates (below 1 ppm) suggests that under these experimental conditions, among rats allowed to live their natural life span, a dose of 1 ppm of NDEA or NDMA in the drinking water will cause about 25% to develop a liver neoplasm, a dose of 0.1 ppm will cause about 2.5% to do so, and a dose of 0.01 ppm will cause about 0.25% to do so, etc., with no indication of any 'threshold.'" The study found a higher order slope at higher dose rates, so simply extrapolating the exposure response from the high dose range would have underestimated the low dose risks actually observed.

[5] Asbestos is not reactive to DNA in a conventional matter. While tobacco smoke contains DNA reactive material, it also might be carcinogenic in the lung via an inert particle effect such as that hypothesized for washed carbon black. Thus, the two paradigm examples in people appear to contradict the notion that only DNA-reactive materials are carcinogenic to low doses or over a wide range.

Thus, in the four examples testing the threshold or low dose extrapolation, low dose toxicity, if not linearity was observed.

Risk assessment applies equally to the dangers of musculoskeletal disorders, the largest cause of occupational disability, or the need to implement NFPA Standard 97, on guarding of machinery, to prevent fatalities, or to the prevention of asthma among workers exposed to metalworking fluids. It does not just apply to cancer.

However, occupational cancer accounts for about 90% of currently identified work-related mortality. And, the longest running and most developed scientific and political controversies around risk assessment for cancer.

Harry Houdini, the magician, gave his last performance in Detroit, where he collapsed on stage and died (Silverman 1996). Like their namesake, Houdini risk assessments either make a risk disappear, or allow a public health agency or industry escape from action to prevent a health problem. Houdini also enjoyed exposing other illusionists who claimed their performances were reality.

Most risk assessment commentaries focus on interventions based on laboratory studies of carcinogenicity in the face of absent or negative epidemiology. Public health advocates argue for "intervention in the face of uncertainty" and the "precautionary principle."

The opposite reality prevails in the occupational environment. Large amounts of epidemiology identify hazardous exposures. Workers endure the refusal to intervene in the face of certain evidence of significant risk at prevailing exposure levels.

Examples of occupational exposures, with sufficient or substantial evidence of carcinogenicity in people not addressed by new exposure limits include silica (IARC 1997), sulfuric acid mist (IARC 1992), diesel particulate matter (IARC 1989), particulate matter generally (Mauderly 1997), metalworking fluids (NIOSH 1998), welding fume (IARC 1990), and, formaldehyde.

The steps in risk assessment which define areas for improvement were codified by the NAS's Committee on Institutional Means for Risk Assessment in the Federal Government (NAS 1983):

- Hazard Identification
- Dose-Response Assessment
- Exposure Assessment
- Risk Characterization

Least contentious of these opportunities is improvement of exposure assessment methodology, which would immediately narrow uncertainty. In addition, better understanding of limitations of exposure assessment in the occupational environment would improve evaluation of studies.

The most important need is to understand variability and uncertainty of exposure. The presence or absence of an exposure-response relationship is important in hazard identification from human studies. Therefore, methods for testing the power of detecting an exposure-response relationship, if it were there, would improve interpretation of such studies.

Reconstructing exposures in studies of known human carcinogens, or laboratory carcinogens with negative human studies, would permit contrast and comparison of potency in human and laboratory studies.

Much past risk assessment debate questioned whether dose response assessment and quantitative risk characterization was valid, feasible, or appropriate. Such assessment leads inevitably to defining an acceptable risk, reasonable risk, significant risk, or a body count for cost benefit analysis at a given exposure level.

Quantitative risk assessment was initially introduced to rationalize not banning certain chemicals by defining the Virtually Safe Dose (VSD) at 1 in a million risk. Later, the method was forced on OSHA by the Petroleum Institute in the Benzene Standard controversy, with the Supreme Court acknowledging a 1 in a thousand lifetime risk for cancer as "significant." This debate has been overrun by the large number of agents for which a carcinogenic hazard has been identified, especially in people. The large number of unbannable agents identified as carcinogens, including silica and diesel particulate matter, compel quantitative analysis at least to set priorities.

Now, many former users of straightforward quantitative extrapolation have turned against the method, instead advocating complicated black-box modeling approaches.

Risk identification remains a significant area of controversy. Identification of carcinogenicity from laboratory studies had moderately well-defined decision rules until the optional exceptions for mechanistic hypotheses was introduced. We lack clear decision rules for use of mechanistic information. For epidemiology, we lack consistent decision criteria, and also suffer from a lack attention to dose in reconciling human and laboratory data. We also lack clear decision rules for genetic toxicology data.

Many conflicts over quantitative risk assessment actually arise from dispute over the hazard identification stage, and what data is relevant and should be used in the risk assessment.

Laboratory and human studies researchers, and risk assessors, must use a common metric for magnitude and limits of detection for each others' endpoints. Toxicologists deal mostly in unit risks, while epidemiologists deal in relative risk without much attention to quantitative dose measures. The metrics used by public health agencies in assessing risks further diverge.

For toxicologists, the following "thought epidemiology" study (Figure 2) illustrates the relationship of dose rate and limit of detection in mortality studies.

Consider a population exposed to an agent *at a level* where the exposure-related risk of lung cancer is 1/100. The background risk of lung cancer among American white males is 6/100. A common outcome measure of a mortality study is the standardized mortality ratio (SMR), the ratio of observed risk (background + attributed) to expected risk (background). In this example, the background risk is 6/100, the hypothetical attributable risk is 1/100 on top of 6/100 adding to 7%. Therefore, the SMR is 117, or a 17% increase.

Such a risk ratio might be statistically significant in a large study. Yet, it would be very suspect for hazard identification For example a study of textile workers exposed to formaldehyde at sub-part per million levels found such a risk ratio for lung cancer to be statistically significant. (Stayner *et al.* 1988) A larger study found a 30% excess risk for lung cancer in a somewhat higher exposed population (Blair *et al.* 1986).

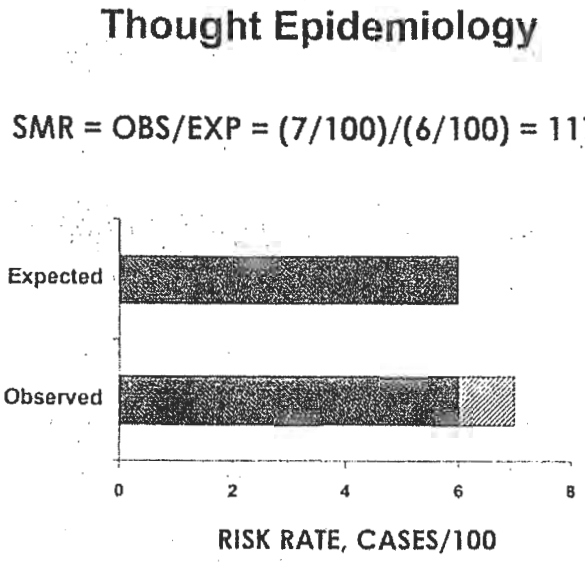


Figure 2. Thought epidemiology.

Neither study has been used for human risk assessment.[6] This detectable elevated risk ratio is a 1/100 compared to 1/1000 level of concern for OSHA or 1/1,000,000 for the U.S. Environmental Protection Agency.

Note that the risk ratio at a dose is the "real" biological phenomenon, while statistical significance is an artifact of sample size, variability, and background rate.

For epidemiologists, another thought experiment chart illustrates a dose-response relationship over several orders of magnitude in risk rate.

The range of direct observation through the typical animal bioassay is limited from 10% tumors, at the lower limit of statistical significance (against a zero background), to 50% tumors because of mortality. This leaves little room for a slope. High background tumors can be observed only at higher absolute risks, and therefore higher doses. Low background tumors can be observed at lower dose levels, but only in specialized or larger studies. The sensitivity of bioassays for hazard identification arises from the ability to administer and measure high doses.

By contrast, epidemiology can detect somewhat smaller relative risks. For lung cancer, we have illustrated a 2% attributable risk as marginally detectable. Most tumors have a lower background than lung cancer, so a lower limit of detection against background can be found. However, detection of more rare tumors requires large groups or specialized studies. The limit of quantitation for epidemiology is perhaps 1 in 1000.

But, mortality studies have many additional limitations, which both compromise their sensitivity, and challenge their use in hazard identification. The most frequent limitation is low dose, compared to the effect levels in laboratory studies, and failure to take human dose into account for risk assessment.

[6] Notably, the investigators in the first study concluded it provided evidence for carcinogenicity, while the second study team concluded it provided "little evidence."

The previous charts illustrate that a no observed effect level (below an observed effect level] still represents a significant risk. A "No Observed Effect Level" in laboratory study (below an established effect level) corresponds to a risk of about 5% (above a zero background). The benchmark dose is a statistically derived equivalent to the NOEL, and corresponds to a 5% risk.

As demonstrated above, the "No Observed Effect Level" in epidemiology depends on the background risk of the cancer site, quality and size of the study.

Thus, extrapolation from laboratory studies to population exposure levels, or high exposure epidemiology to lower exposed populations requires extrapolation beyond the dose range where direct observation is feasible (Figure 3).

In long-range extrapolation, the SLOPE of extrapolated dose response curve determines risk estimates at lower exposures. The observed range has little to do with the end result. The choice of the extrapolation curve to these lower doses, except for the four examples cited above, is based on ASSUMED mechanism, because direct observation is not feasible (Figure 4). A shallower slope projects higher low dose risk. A steeper slope projects lower low dose risk and may be the equivalent of a threshold. Laboratory studies yield an exposure response relation for in-bred and homogeneous animals living under controlled conditions. This suggests that the population exposure response would likely be steeper in laboratory studies — and predict lower low-dose risks — than that observed for free living people.

Multistage mechanisms predict a break point in the exposure response relationship, where a linear process takes over from higher order processes. The break point happens in a dose range that can not be directly observed. The break point dominates the slope of the first-order process, which determines the low dose risk (Figure 5).

Thus, risk assessment models are based on choices of the modeler. Slopes greater than linear in the low dose range predict lower risk. Outcomes of the choices are known in advance.

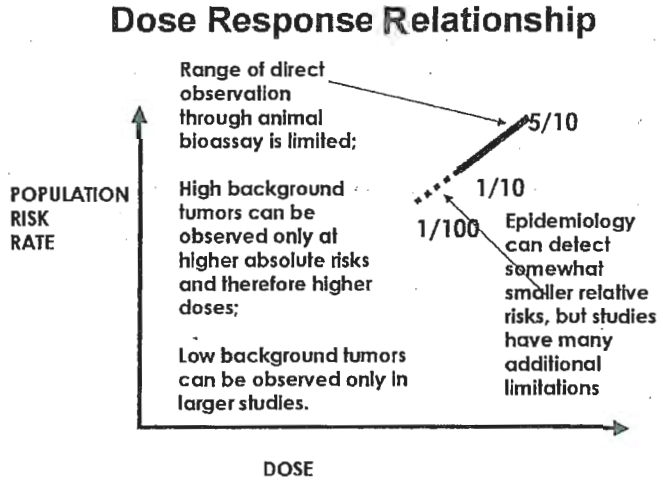


Figure 3. Dose response relationship direct observation.

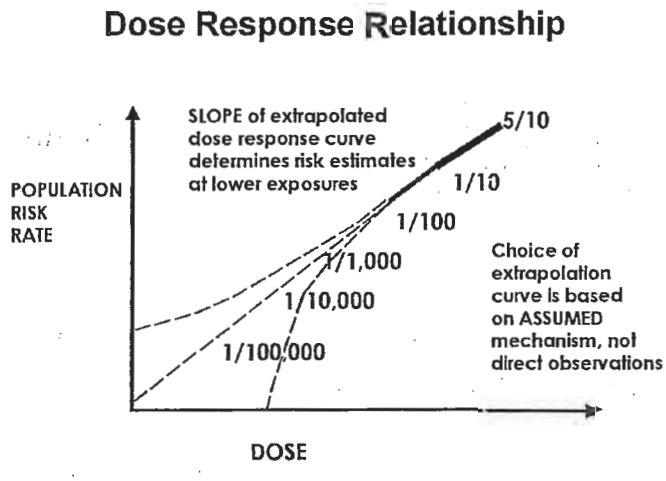


Figure 4. Dose response relationship extrapolation.

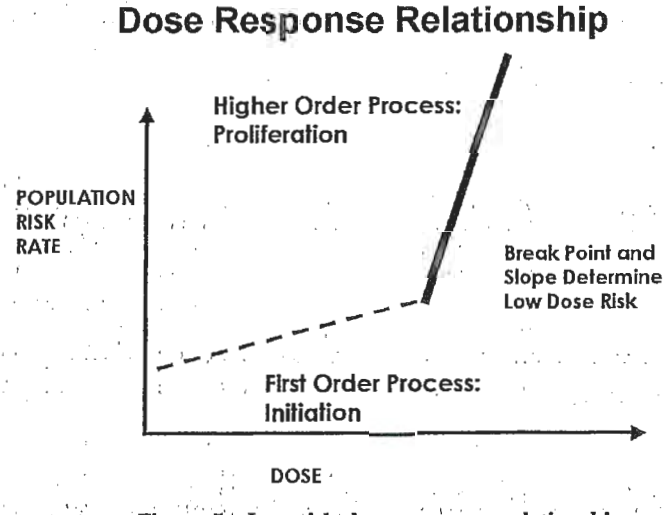


Figure 5. Low risk dose response relationship.

The Houdini risk assessment approach is now routine in regulatory controversies. It has been applied to important chemicals like formaldehyde, methylene chloride, and butadiene.

Houdini risk assessment steps are

- Ignore selected tumor sites and types, and all data from people (as with formaldehyde and diesel);
- Chose most resistant species in laboratory test;
- Select a biochemical parameter in which most resistant species is more like people;
- Assume a mechanism which gives threshold or steep exposure response for carcinogenic effect;
- Reduce estimated people risk by the parameter ratio to most resistant species;
- For gasoline, people are mice, not rats (if you ignore clear evidence of liver tumors in mice);
- For methylene chloride, people are rats, not mice (if you ignore fatal breast tumors in rats).

Included in the Houdini approach are a set of suspect tumors and sites to be discounted or ignored:

- thyroid follicular-cell tumors in rodents;
- renal tubule cell tumors in male rats;[7]
- calculi- and microcrystalluria-associated urinary bladder neoplasms in mice and rats;[8]
- liver tumors in mice (sometimes rats);
- lung tumors in mice exposed to solvents by inhalation (clara cell);
- lung tumors in rats exposed to particulate matter.

And also included are a set of anti-hypotheses of varying age and sophistication:

- toxicity in the target organ;

---

[7] Renal cell tumors are discounted based on the alpha-2u-globulin hypotheses, especially to escape from new controls on gasoline, aliphatic petroleum hydrocarbon materials such as Stoddard Solvent, and a variety of other materials.

[8] Saccharine is the target of this mechanism. The NTP Board of Scientific Counselors group to review the NTP Report on Carcinogens declined to support this hypothesis by a narrow vote, but was overruled by the agencies when saccharine was delisted.

- High dose changes metabolic pathway
- not genotoxic[9]
- Rat-monkey hybrid DNA cross-link model for formaldehyde;
- alpha-2μ-globulin;
- peroxisome proliferation;
- overwhelming lung clearance.

The method attracts laboratory toxicologists and modelers because it appears to be more "scientific," "mechanistic," and "biologically based." But is it real? Some of this attraction is simply a cultural bias, an attraction for a professional role. A model is not evidence, it is a hypothesis to be tested against all the data, including human risk levels. To policy makers and the general public, it is a black box machine that only scientists can operate and understand.

You get what you model for. Models with thresholds and steep slopes inevitably project lower risk. The parameters and precursor toxic effects, which direct the models, are rarely validated against *chronic exposure outcomes* for the target chemical or positive and negative controls.

After a positive bioassay is reported for any economically important chemical, our friends on the industry side get together and devise "mechanistic" studies to systematically explore hypotheses for that specific chemical which projects lower human risk.

By contrast, public sector funding follows the typical RO1 course, not especially informed by the importance of an agent.

Devising models helps understand systems. This simple compartment model in Figure 6 was devised to help explain solvent and lead toxicity in training courses for union representatives. Even this simple diagram has 11 arrows, which means 11 rate constants for uptake and distribution before taking metabolism into account.

Adding competing metabolic pathways generates "William Tell" diagrams with a blizzard of arrows, and usually more than one apple.

But every one of those arrows is another parameter, many of which can't be measured directly, and the choice of the important apple is often debatable.

Chemical kineticists are taught (Figure 7):

- With 5 parameters you can draw an elephant.
- With 7 you can make it wag its tail
- With 30 you can draw the Mona Lisa.[10]

[9] The mutagenicity bioassay was intended to quickly identify carcinogens so that animal tests weren't needed. Instead, no agents are identified as possible carcinogens on genotoxicity alone, while clear evidence of carcinogenicity in animals is discounted based on negative gene-tox.

[10] Some of the models for methylene chloride debated in regard to the OSHA standard involved upwards of 30 parameters.

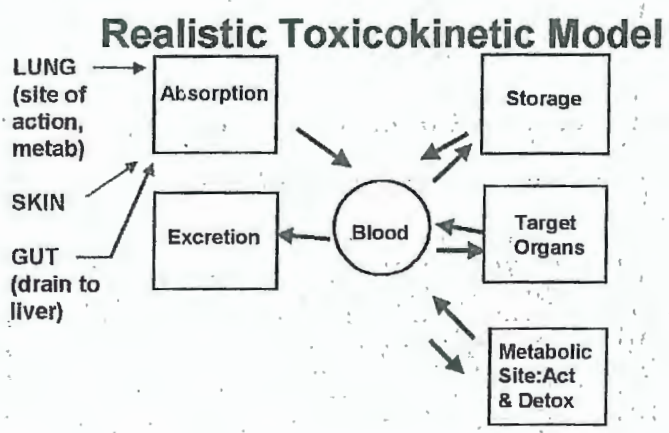


Figure 6. Realistic toxicokinetic model.

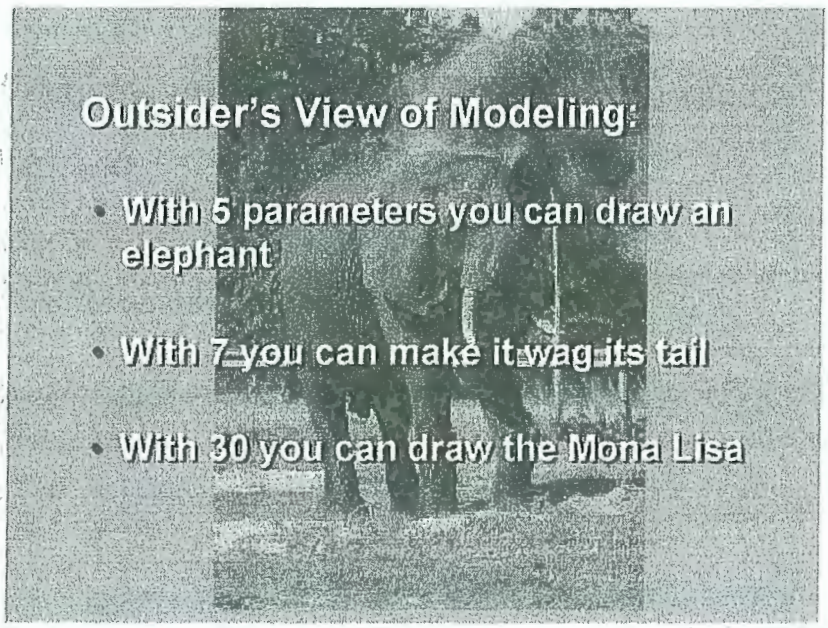


Figure 7. Outsider's view of modeling.

Models are only hypotheses: they must be tested against chronic endpoints, or the behavior of known carcinogens in either the animals or people. These models include many parameters but are fitted to the very few data points of an exposure-response relationship.

Modelers also advance a black box lingo to dominate policy debate. Consider the following quote from a paper on benzene, which predicts a low dose protective effect on leukemia in people:

"A Monte-Carlo uncertainty analysis based on maximum-entropy probabilities and Bayesian conditioning is used to develop an entire probability distribution for the true but unknown dose-response function." (Cox 1996)

Quantitative risk assessments should use all the data, including human data. All positive studies should be used to estimate the unit risk observed in the cohort as a whole or exposure groups. Where the exposure levels are not known, or determined concurrently with the effects, those levels should be estimated by expert knowledge or determined by follow up modeling or demonstration studies. Negative studies may define an upper limit to risk at the exposure level of the studied cohort. Absent studies, the possibility of detection of predicted risks from various animal models can be estimated.

Uncertainty in risk assessment from laboratory study includes statistical uncertainty in effect rates and model uncertainty in species to species and high to low dose extrapolation (Figure 8). However, exposure and administered dose are precisely known. Studies in people are also subject to statistical uncertainty in effect, and to the uncertainty of causality. Model uncertainty in species extrapolation is absent, and high dose to low dose extrapolation unusually involves a much smaller range. Additional uncertainty arises from statistical uncertainty in dose.

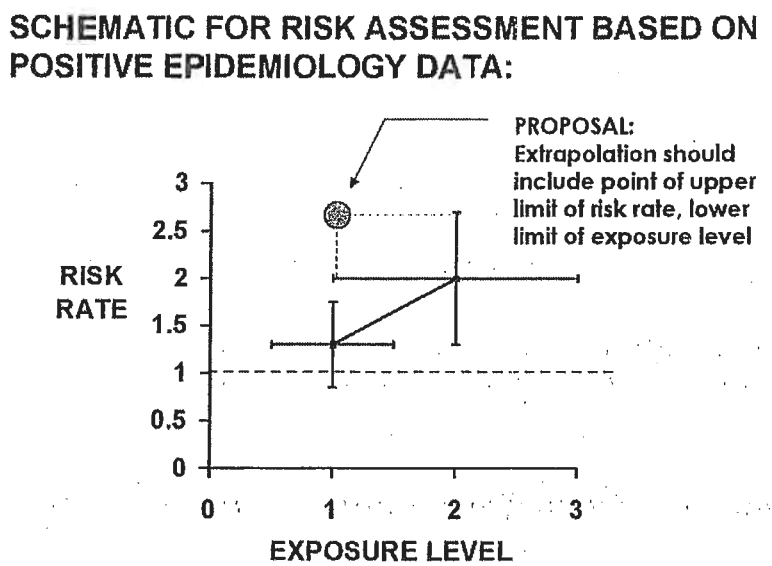


Figure 8. Schematic for risk assessment.

An appropriate approach for quantitative risk assessment from a population study includes:

- Convert relative risk in the epidemiologic study to absolute risk (lifetime risk estimate) based on background population rates;
- Correct for Healthy Worker Effect (relative risk for cancer in study is typically 10% lower than true risk; HWE is an important obstacle to hazard identification, but has modest impact on absolute risk);
- Calculate unit risk from estimated exposure level;
- Upper limit of risk is calculated from upper confidence level for the risk rate and the lower confidence interval for exposure;
- Where the studies reveal no association, the upper confidence interval for the risk rate should still be used.

Present practice avoids risk extrapolation from human data. Multiple epidemiological studies showing excess rate ratios are frequently disregarded. The Health Effects Institute convened an expert panel to justify not using human data for diesel particulate matter (two of the twenty studies finding excess lung cancer in truck drivers and railroad personnel with expected higher diesel particulate matter exposure than the general population.)

Devising and applying a common metric between laboratory toxicology, human epidemiology and clinical studies, and modeling will permit the three disciplines to converse and combine their currently disparate contributions. Our NORA risk assessment initiatives should support that goal.

REFERENCES

Blair A, Stewart P, O'Berg M, *et al.* 1986 Jun. Mortality among industrial workers exposed to formaldehyde. J Natl Cancer Inst 76(6):1071-84

Cox LA. Reassessing benzene risks using internal doses and Monte-Carlo uncertainty analysis. 1996. Environ Health Perspt 104(suppl 6):1413-29

IARC (International Agency for Research on Cancer). 1989. Monographs 46:41. Lyon, France

IARC (International Agency for Research on Cancer). 1990. Monographs 49:447. Lyon, France

IARC (International Agency for Research on Cancer). 1992. Monographs 54:41. Lyon, France

IARC (International Agency for Research on Cancer). 1997. Monographs 68:41. Lyon, France

Littlefield NA, Farmer JH, Gaylor DW, *et al.* 1980. Effects of dose and time in a long-term, low-dose carcinogenic study. J Environ Pathol Toxicol 3(3 Spec No):17-34

Mauderly JL. 1997. Relevance of particle-induced rat lung tumors for assessing lung carcinogenic hazard and human lung cancer risk. Environ Health Perspect 105(suppl 5):1337-46

NAS (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. Committee on the Institutional Means for Assessment of Risks to Public Health, National Academy Press, Washington, DC, USA

NIOSH (National Institute for Occupational Safety and Health). 1998. Criteria for a Recommended Standard: Occupational Exposure to Metalworking Fluids. NIOSH (DHHS). Publication No. 98-102. Cincinnati, OH, USA

OSHA (Occupational Safety and Health Administration). 1991. Response to court remand for proposed rule on occupational exposure to formaldehyde. Federal Reg 56:32302

OSHA (Occupational Safety and Health Administration). 1998. Methylene chloride: Final rule. Federal Reg 63:50711-32

Peto R, Gray R, Brantom P, *et al.* 1991 Dec 1. Effects on 4080 rats of chronic ingestion of N-nitrosodiethylamine or N-nitrosodimethylamine: A detailed dose-response study. *Cancer Res* 51(23 Pt 2):6415-51

Silverman K. 1996. Harry Houdini!!! The Career of Ehrich Weiss: American Self Liberator, Europe's Eclipsing Sensation, World's Handcuff King and Prison Breaker. Harper Collins, NY, NY, USA

Stayner LT, Elliott L, Blade L, *et al.* 1988. A retrospective cohort mortality study of workers exposed to formaldehyde in the garment industry. Am J Ind Med 13(6):667-81

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1243-1247 (2002)

An NGO Perspective on Risk Assessment and Scientific Research

Ellen K. Silbergeld[1]
University of Maryland School of Medicine, Department of Epidemiology and Preventive Medicine, Baltimore, MD

ABSTRACT

Concerns over risk assessment have been raised by Non-Government Organizations (NGO) and the environmental community for decades. In considering proposals for research in this area, it is important for both scientists and policymakers to consider the following points: (1) risk assessment as a method of policymaking is increasingly inaccessible to meaningful public participation, (2) the lack of fundamental toxicological data constrains the application of risk assessment methods more than any other factor, and (3) the importance of individual susceptibility in risk assessments must be tempered by the lack of control over individual exposures.

Key Words: risk assessment, environmentalists, susceptibility, toxicology.

INTRODUCTION

Speaking as a representative of the community of environmentalist/Non-Government Organizations (NGOs), I will draw attention in this paper to the range of opinions that exist within that community—not necessarily by endorsing all of them, but in the spirit of this conference, by raising the issues that may define a relevant and important domain for considering where research can help move us forward. We are, of course, not politicians, at least not mainly. We're not economists. But we may have the opportunity to add some bits of information, methodology, and perspective that can help resolve outstanding issues.

Speaking on behalf of Environmental Defense (ED), I make only two introductory notes: this organization has a long history of involvement in the development and use of risk assessment from its birth in this country, in a formalized sense, and in science-based regulatory policy more generally. ED was founded by scientists and

[1] Current address: The Johns Hopkins University, Bloomberg School of Public Health, Department of Environmental Health Sciences, 615 N. Wolfe Street, Baltimore, MD 21205; Tel(voice):410-955-8578, Tel(fax):410-955-9334; esilberg@jhsph.edu

1080-7039/02/$.50

continues to value the importance of scientific research and analytic methods in developing policy recommendations. We can date to 1979 when the Interagency Regulatory Liaison Group (ILRG) proposed methodologies for cancer risk assessment (OTA 1987); ED scientists have been involved in many advisory committees, at the U.S. Environmental Protection Agency (USEPA), the National Research Council (NRC), and elsewhere, in the continuing development of risk assessment as a policy tool. Second, despite that long engagement and involvement, ED shares the disquiet of many (*e.g.*, O'Brien 2000; Raffensperger and Tichner 1999) with the actual practice and history of risk assessment as a methodology of policymaking.

From the NGO perspective, a major problem with risk assessment has been the increasing complication and inaccessibility of the process. Now the notion of complexity in a context such as this conference is attractive. The bread and butter of research is increasing complexity so that we can then get money to solve it. In my experience, however, truly elegant research simplifies, or at the least helps us find paths through extremely complicated concepts. So, a concern about complexity does not necessarily translate into a disquiet over research, although sometimes it does when research seems poorly focused to meet the challenge of complexity.

From the policy perspective, unnecessary complexity can create a climate of distrust, and Frank Mirer's Houdini analogue (Mirer 2002) is a beautiful way of expressing that. But in addition to Houdinism, the complexity of risk assessment—as is commonly practiced now—denies access to the public. It was Bill Ruckelshaus, in an extremely important essay called "Risk Assessment, Science and Democracy," who raised fundamental and real concerns about an increasingly complex system of decision-making within the context of a participatory democratic society (Ruckelshaus 1985). No one has really acknowledged or resolved the Jeffersonian issues he identified.

There is in addition another cost of complexity, as raised by many speakers at this conference, that the complexity of risk assessment necessarily devours time and resources. It prolongs decision-making indefinitely and it consumes a great deal of scientific expertise, the resources of people, of money, of agency process, and animals as well. Complexity is over encouraged by an extreme bias against Type I errors, which springs from a very strange philosophy within the U.S. regulatory system, regarding the burden of proof. That is, the burden of information production and definition of risk is placed, almost always, on those who have a concern for the adverse effects of a technology or other kinds of action, rather than the reverse. In practice, risk assessors seem to operate on the "hasn't killed anyone yet" principle, demanding proof of harm rather than substantiation of safety. At present, the tension between Type 1 and Type 2 errors is very unbalanced (Weiss 2001; Needleman 1995). We seem enormously more concerned at overestimating a risk, rather than underestimating it, despite the wisdom of Bradford Hill to temper statistics with judgment (Schwartz *et al.* 1999; Weed 2000).

There have been many attempts to deal with this problem of balancing the demands and burdens of knowledge. The most successful, in terms of process, in this country was drafted in California by ED. California's Proposition 65 has been extraordinarily successful; it was simply an attempt to shift the burden in order to promote action but not to discourage the proffering of relevant and more definitive information. So, precaution does have not to be a shutting of the door upon science

and research, but rather an encouragement of research from all sides, not merely from those, or mainly from those, who seek to do something about a perceived risk. The primary impetus for the Precautionary Principle is, at heart, an attempt to shift the burden of process and discourse (Kriebel *et al.* 2001; O'Brien 2000). Seen in that light, the Precautionary Principle can engage a more equal tension between the kinds of uncertainties and information gaps that always bedevil us as scientists and policymakers.

Finally, the complexity of risk assessment in practice frequently supports a risk assessment process that is irrelevant. Much of what we do in risk assessment really doesn't relate to the kinds of policy interventions that we can actually bring to bear upon reducing risks. In most cases, for instance, technology based standards are not as fine-tuned as the demands we place on risk assessment. So we have to ask whether it is a useful expenditure of private and public resources of all types. Sometimes, in fact, it can be antiprotective. In this context, I want to raise the great research horizon opening up before us in using molecular and other techniques to understand individual variations in individual susceptibility. We have to ask ourselves right now (and I think it's almost too late): What are we going to do with that information? We have a relatively clumsy hand of intervention when it comes to making decisions. How will we use the fine structure of information about individual susceptibility? I participated in the Johnson Controls Case about lead and discrimination against women in occupations (Sever 2000). We have not gone very far in this society to understanding how to use susceptibility information in a socially acceptable way.

A second issue associated with new technologies in epidemiology, toxicology and risk assessment is the issue of data demands. We've heard a lot of reasons why the data demands are so extreme and continue to bedevil us. Part of that, I think, is our unequal attention to Type I and Type II errors. But more generally, I believe that the risk assessment methodologies we have developed over the last 30 years have created a hungry god. This is a god that doesn't know when it's full. We don't know what constitutes a minimally acceptable risk assessment. What is the minimal amount of information we need to go forward? We can all talk about the ever-receding horizon of information that we would like to have, but that can play into the social obstacles against reaching decisions. I believe that an urgent research goal is to define the nature of information we require in order to support the kinds of analytic methodology that are feasible and efficient. How relevant is it to define the cancer risk of dioxin when we argue about differences in the femtogram range? RFDs of that order generally require preventing all possible releases of dioxins. We need to solve this problem in order to use risk assessment and to generate the kinds of supporting ancillary research, both basic and applied— toxicology, epidemiology— in order to make risk assessment a more useful endeavor in public policy.

The High Production Volume (HPV) Chemical Testing Initiative, which Goldstein (2002) alluded to, is an attempt to do this (Environmental Defense, Toxic Ignorance, available on www.environmentaldefense.org/documents/243_toxicignorance.htm). I'm very proud of the partnership among government, industry, and ED, among others, that has continued to generate momentum for this process now on a worldwide basis. The goal of the HPV initiative is not to provide definitive data for elaborate risk assessments. It is rather the beginnings of light in a mostly darkened world. So to

suggest that these tests have flaws in terms of definitive hazard identification, when we have little to no information for most high production volume chemicals, is not arguable.

The HPV program will generate a real research challenge: Within some 4 or 5 years, we will have an enormous database of information on a very large number of chemicals. How are we going to use that information? How do we, for example, use it to select those chemicals for which further research—on hazard or risk—might be undertaken? How will we incorporate that information in a rational program of exposure assessment, going out into the real world to define priorities for action? I believe that these questions offer some exciting opportunities for applied and basic research in risk assessment, using a dataset much larger than we have ever had before. I would warn that there are, in fact, two paths when it comes to innovations, if you will, in hazard identification. One is the HPV approach, a very rough-and-ready but easily understood approach of opening doors into closed rooms. If you make a lot of it, you need to know at least a little about it. The other approach is exemplified by the endocrine disrupter testing process. I think that is a potentially very dangerous harbinger of what I might call "boutique toxicology" in which we invest millions of dollars on a very limited set of endpoints with no real sense as to how the data are going to be incorporated into a rational database for risk assessment.

Finally, I would like to end with a comment that we can take forward into our groups. I hope some day in my life to come to a risk assessment conference in which the word "risk assessment" is not merely a code for cancer. We have got to start developing both the basic knowledge and epidemiologic and biostatistical approaches for confronting non-cancer risks (Weiss 2001). We still have no science that underlies our current approach. It's some "one size fit all." Whether we use a bench mark dose, a No Observed Effect Level (NOEL) approach, a slope, an implicit threshold; it's all the same for developmental toxicology, reproductive toxicology, immunotoxicology, neurotoxicology. Speaking as a scientist involved in some basic research in these physiologic systems, I consider this is absolute nonsense. There is no way that one model can capture all the sets of mechanisms that define the differences among these systems and their responses to chemicals. We have got to go beyond this.

The most egregious example of how risk assessment really means cancer has come up in the most recent evaluations by the USEPA of the risks of dioxin (not yet finalized). I carry the burden of being part of what has been called "the December Group," but I hope we don't emulate the Decembrists of Tsarist Russia, who were all executed, when we try to institute reform. But nonetheless as a Decembrist, I was dismayed by our work product, despite an extraordinary amount of work, an extraordinary body of contributions from the scientific community worldwide, extraordinary innovations in risk assessment — much done by Chris Portier's group at the National Institute of Environmental Health Sciences (NIEHS)—and certainly a growing and compelling epidemiologic database on the non-cancer health effects of dioxins; specifically, their developmental effects (Birnbaum and Tuomisto 2000). What did we put forward for regulators and the American public? Yet another cancer risk assessment for dioxins. We need to learn from other recent evaluations, particularly that done on methylmercury, undertaken first by the USEPA, then

reviewed by the National Institutes of Health and finally by the National Academy of Sciences (NRC 2001). These risk assessments may finally convince us, first, that non-cancer health effects are likely to be more significant for the public's health; and second, that there are ways of using basic and applied insights from toxicology and epidemiology to craft new pathways in risk assessment that will allow us expeditiously, efficiently, and scientifically to move to confront these categories of risk.

REFERENCES

Birnbaum LS and Tuomisto J. 2000. Noncarcinogenic effects of TCDD in animals. Food Addit Contam 17(4):275-88

Goldstein BD. 2002. Topics in hazard identification: Oxygenated fuels, safety assessment, hematological neoplasms, and the precautionary principle. Human Ecol Risk Assess (*this issue*)

Kriebel D, Tickner J, Epstein P, *et al.* 2001. The precautionary principle in environmental science. Environ Health Perspect 109:871-6

Mirer FE. 2002. Does the emperor have any clothes?: Using mechanistic information or doing Houdini risk assessments? Human Ecol Risk Assess (*this issue*)

Needleman HL. 1995. Making models of real world events: The use and abuse of inference. Neurotoxicol Teratol 17:241-2

NRC (National Research Council). 2001. Toxicology of Methylmercury. National Academy Press, Washington, DC, USA

O'Brien M. 2000. Making Better Environmental Decisions. MIT Press, Cambridge, MA, USA

OTA (Office of Technology Assessment). 1987. Identifying and Regulating Carcinogens. U.S. Government Printing Office, Washington, DC, USA

Raffensperger CA and Tickner J. 1999. Protecting Public Health and the Environment. Island Press, Washington, DC, USA

Ruckelshaus W. 1985. Risk, science, and democracy. Issues Science Technol 1:19-38

Schwartz S, Susser E, and Susser M. 1999. A future for epidemiology? Ann Rev Pub Health 20:15-33

Sever LE. 2000. Ethics, society and occupational reproductive hazards: fetal consequences? Women's Health 30:25-37

Silbergeld EK. 1993. Risk assessment: The perspective and experience of US environmentalists. Environ Health Perspect 101:100-4.

Weed DL. 2000. Epidemiological evidence and causal inference. Hematol Oncol Clin North Am 4:797-807

Weiss B. 2001. Ethics assessment as an adjunct to risk assessment in the evaluation of developmental neurotoxicants. Environ Health Perspect 109:905-8

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1249-1251 (2002)

Current Perspectives on Issues in Risk Assessment Methods

Herman J. Gibb[1]
U.S. Environmental Protection Agency, National Center for Environmental Assessment, Ariel Rios Building, 1200 Pennsylvania Avenue, NW, Washington, DC, 20460; Tel(voice):202-564-3334, Tel(fax):202-565-0059; gibb.herman@epa.gov

ABSTRACT

The Office of Research and Development (ORD) of the U.S. Environmental Protection Agency was reorganized in 1995 to follow the risk assessment paradigm developed by the National Research Council. With the reorganization, a number of different research strategies and plans were developed on problem-driven topics (*e.g.*, arsenic, particulate matter, and microbial pathogens/disinfectant byproducts) as well as on core research (*e.g.*, ecological research, human health risk assessment research, and pollution prevention research). The human health risk assessments research strategy, which is currently under development, addresses a variety of issues which affect all human health risk assessments. These include mechanism of action, variation in response, aggregate risk, and effect on public health. While all of the issues that have been identified are important, the challenge to ORD is to identify which will have the greatest impact on improving risk assessment so that resources can be most strategically applied.

Key Words: mechanism, aggregate risk, public health, variation in susceptibility.

The U.S. Environmental Protection Agency (USEPA) has maintained a research program in human health risk assessment since its inception in 1970. In 1995, the Office of Research and Development at the USEPA was reorganized to focus on the risk assessment paradigm outlined by the National Research Council (NRC 1983). The reorganization created National Laboratories of Effects, Exposure, and Risk Management and Centers of Risk Assessment and Exploratory Research. With the reorganization, research strategies and plans on different topics, including human health risk assessment, were developed. Final research strategies or plans include the Waste Research Strategy, Pollution Prevention Research Strategy, Research Plan

1 The discussion in this paper does not necessarily reflect the view of the U.S. Environmental Protection Agency. This manuscript is considered to be a work of the U.S. Government and is therefore not copyrighted.

1080-7039/02/$.50

for Endocrine Disruptors, Ecological Research Strategy, Research Plan for Arsenic in Drinking Water, Research Plan for Microbial Pathogens and Disinfectant By-Products in Drinking Water, Action Plan for Beaches and Recreational Water, Mercury Research Strategy, and the Strategy for Research on Environmental Risks to Children (http://www.epa.gov/ORD/WebPubs/final). Strategies or plans currently in draft include the Particulate Matter Research Program Strategy and the Global Change Research Strategy. Other plans or strategies currently under development include the Human Health Risk Assessment Research Strategy, Air Toxics Research Strategy, Environmental Monitoring and Assessment Program Research Strategy, and the Drinking Water Contaminants Candidate List Research Plan (http://www.epa.gov/ORD/resplans).

Some of these plans or strategies are what the Agency has called problem-driven issues, such as arsenic, particulate matter, or microbial pathogens/disinfectant by-products and focus on health risk assessment issues regarding those issues. Other strategies are considered core research such as ecological research, human health risk assessment research, and pollution prevention research. The core research program develops information needed across multiple Agency programs (*e.g.*, air, water, pesticides) and address pollution problems that are multimedia (*e.g.*, air, water, pesticides on food, *etc.*). One of the plans longest under development has been the Human Health Risk Assessment Research Strategy. Human health risk assessment research is of course the subject of this meeting. The Agency has drawn on multiple references in its preparation of the human health risk assessment strategy including the NRC's (1996) *Science and Judgment in Risk Assessment* report and various reports of its Science Advisory Board such as *Human Exposure Assessment: A Guide to Risk Ranking, Risk Reduction, and Research Planning* (SAB 1995). A number of areas have been identified as important to the strategy:

1. Mechanism of action. What is the dose response at exposures below the range where effects were seen in animals or humans? Is the mechanism of action in animals different from that in humans? How does the mechanism of action affect route-to-route extrapolation?

2. Variation in response. What is the variation in response to a chemical substance across the human population as a result of genetic polymorhphisms, preexisting disease, variation in exposure, diet, life stage, and other factors?

3. Aggregate exposure. What is the total exposure (and risk) to a single agent from different routes of exposure (*e.g.*, inhalation, ingestion, dermal, *etc.*)?

4. Cumulative exposure. What is the combined risk from aggregate exposures to multiple agents or stressors?

5. Effect on Public Health. What is the effect that USEPA's risk management actions have on the health of the country?

While we have been good at identifying the issues of concern that face risk assessment, we have been less successful in determining the strategy for dealing with these issues (*i.e.*, which of these issues will have the greatest effect on a risk assess-

ment and how do use our resources accordingly?). Thus, a good strategy, in addition to identifying the issues, has to pick the priorities for research. That means if we identify new areas for research we also need to concurrently identify what areas of research we will be deemphasizing, difficult to accept since the tendency is to want to do everything and hard for the individual scientist who may have invested years of research on a particular topic. Of course, the inclination is to include areas of research in the strategy that one's laboratory or center is currently doing. That helps to justify the current research program of the laboratory, but may do little for improving risk assessments. One of the most positive features of the USEPA Grants Program, which was markedly increased with the Office of Research and Development reorganization in 1995, is that it allows the Agency to go outside of the expertise currently existing in the laboratories. Still there is the issue of which research will have the biggest impact on our assessments, and identifying those areas is a challenge for the Federal agencies and others engaged in improving risk assessment.

While we have heard and seen estimates of how much this or that factor could affect a risk assessment, which of these factors has the *greatest* effect on a given risk assessment? Which is the short term and which is the long-term research, and what should be the balance? Have we tried to think out of the box as much as we can? For example, would it be better to first identify what people are actually exposed to and where those exposures come from before we begin to evaluate the effects of particular chemicals? Would long-term cohort studies such as the Framingham study be the best (and perhaps only) approach to assessing effects of environmental and occupational insults for different ages, different susceptibilities, different lengths of exposure, *etc.*, many of the questions for which we have little or no answers now? It is my hope that this meeting will not simply identify the issues that face risk assessment, but will identify the one or two issues or approaches which are of greatest priority for the limited resources available.

REFERENCES

NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the Process. National Academy Press, Washington, DC, USA

NRC (National Research Council). 1996. Science and Judgment in Risk Assessment. National Academy Press, Washington, DC, USA

SAB (Science Advisory Board). 1995. A SAB Report: Human Exposure Assessment: A Guide to Risk Ranking, Risk Reduction and Research Planning. EPA-SAB-IAQC-95-005. Prepared by the Indoor Air Quality/Total Human Exposure Committee, U.S. Environmental Protection Agency. Washington, DC, USA

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1253-1265 (2002)

The Use of Epidemiology in Environmental Risk Assessment

Joel Schwartz[1,2]
[1]Department of Environmental Health, Environmental Epidemiology Program, and [2]Department of Epidemiology, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA, 02115, USA; Tel(voice): 617-432-1245, Tel(fax): 617-277-2382

ABSTRACT

Epidemiology provides estimates of the concentration–response relation for environmental and occupational toxicants in the species of interest, in or close to the dose range of interest. As such, when available, they provide the primary source for risk assessments. Further information can be acquired by using modern biostatistical techniques to assess the shape of the dose response relation, examine effect modification, and assure control for confounding. These approaches are particularly effective if they are done in the context of a meta-analysis or hierarchical model. This is illustrated with examples from the air pollution literature.

Key Words: epidemiology, risk assessment, hierarchical models, air pollution.

INTRODUCTION

The field of Risk Assessment is strangely bifurcated. On the one side is the subfield of Cancer Risk Assessment, on the other Noncarcinogenic Risk Assessment. Most of the cancer risk assessments are based on studies linking dose to cancer risk in animals. These involve carefully controlled studies with known doses. This information is then extrapolated to risk assessments for humans by modeling. These modeling assumptions involve models that relate exposure to dose and models that relate animal dose–response to human dose–response. The dose-response extrapolation includes species differences, which require corrections for, *e.g.*, differences in metabolism and metabolic pathways that affect breakdown products and rates, as well as size corrections, which are usually based on surface area adjustments whose appropriate form is still under debate. In general, the exposures are at high dose, which raises questions about the linearity of the relationships. Feeding concern

Supported by EPA grant R827353 and NIEHS grant ES 00002

1080-7039/02/$.50

about the assumed shape of the relationship are biological models such as the multiple hit model, as well as concerns over cell killing.

While progress is being made on biologically based models, substantial uncertainty remains. Epidemiology is seldom used in this area, ostensibly because it is not available, or because the possibility of confounding introduces uncertainty into the estimates. Often, questions are raised as to whether it is even possible to use epidemiology for risk assessment. The use of dose rather than exposure is touted as an advantage, despite the necessity to then introduce modeling uncertainty about activity pattern, micro-environmental concentrations, *etc.* when extrapolating to a risk assessment where typically the only generally available information is exposure. Research is focussed on obtaining better micro-exposure data to improve these models, not on obtain exposure–response relations.

Of course, the limited availability of epidemiologic concentration–response relations derives, in no small part, to a bias against using them. And while confounding and bias are ever-present threats to the validity of epidemiology studies, the four order of magnitude difference in the cancer potency of some carcinogens among different species of rodent indicates that there are uncertainties in that approach as well. One of the advantages claimed for the animal studies are that the exposures are controlled, and known. However, in the real world, people are exposed to multiple other substances that may act as promoters, initiators, and modifiers of the cancer potency of a single substance. Epidemiology by its nature incorporates that preexisting exposure into the subjects under study, and obtains dose-response relations in the presence of more or less typical exposures to these other substances. This is actually an advantage for the epidemiologic approach.

Risk assessment for noncarcinogenic environmental pollutants represents almost the mirror image. All of the major risk assessments have been based on epidemiologic data. The risk assessments and cost benefit analyses for lead in gasoline (Schwartz *et al.* 1985), lead in drinking water (Levin 1986), as well as unpublished analyses of lead abatement in housing and lead screening all relied entirely on epidemiologic data. Similarly, the risk assessments for the recent revisions of the ambient air quality standards for NO_2, O_3, and $PM_{2.5}$ were all based on epidemiology. Estimates of drinking water born gastrointestinal illness (Levin and Kleiman 1999) likewise rely on human data. Here risk assessments are generally based on exposure, not dose. Research has focused on better exposure measures, not on obtaining dose information, which is dismissed as too costly or impractical for an epidemiology study. Animal data is often ignored. For example, a recent paper (Pocock *et al.* 1994) evaluated the evidence for the effects of lead on cognitive function solely by examining the epidemiological literature, ignoring experimental studies showing lead-induced cognitive effects in primates (Rice 1985), lead–induced impairment in long–term potentiation in the hippocampal area of the brain of rodents (Lasley *et al.* 1993), lead–induced impairment of dopaminergic neurotransmission (Corey-Slecta and Widzowski 1991; Cory-Slecta *et al.* 1993), *etc.* When animal data is assessed, "splitters" rather than "lumpers" predominate, and positive findings are often dismissed as not directly relevant. We are, with reason, reminded that the nasal passages of the rat are far more effective at removing particles than the passages of humans, raising questions of equivalency of exposure, that the shape of and number of branches of the bronchial tree means that ozone deposition patterns in the lung are different, and that the rats

ability to make vitamin C may make the animals response to exposures that produce oxidative stress in the lung of less relevance to humans. Little tolerance for animal based risk assessment exists.

Clearly, better communication between these two ships passing in the night would be useful. However, it has become increasingly clear that epidemiology is extremely useful for risk assessment, and can be made more useful by advancements in technique. It is possible to extract more information than is usual from the epidemiological studies by applying better analytical methods. In this paper I address some ways to improve epidemiologically based risk assessments by the use of nonparametric smoothing and hierarchical models. In particular I will focus on three questions. They involve better assessment of the shape of the exposure–response curve (what is happening?), better assessment of heterogeneity and predictors of heterogeneity in that response (who is it happening to?), and better assessment of which subspecies of exposure is primarily responsible for the effect (what is doing it?).

While better concentration–response relations can be obtained in single studies, I will address these issues primarily in the context of hierarchical models. A hierarchical model is a model that has more than one level of analysis. Meta–analysis is a simple example of such a model. A separate regression coefficient for exposure is estimated in each study, and in a second stage model those coefficients are combined together. A more sophisticated approach would be to regress the slope in each study against study characteristics, to see if any of those characteristics explain differences in slopes. Meta–analysis has formed the basis for most of the risk assessments mentioned above. In addition to providing more stable estimates it implicitly recognizes the need to have multiple confirmatory studies to generate enough scientific consensus to justify a risk assessment. Confidence is also gained when the hierarchical modeling is planned in advance—that is, when all of the studies are analyzed in a similar manner in a prearranged study where concerns about file draw problems are eliminated.

WHAT IS THE SHAPE OF THE EXPOSURE-RESPONSE CURVE?

The shape of the exposure–response curve is a critical issue in risk assessment, and most recent arguments about carcinogenic risk assessment have fundamentally been arguments about this issue. It is no less critical for noncarcinogens. To answer it we need a flexible approach (so we don't predetermine the answer), as well as enough power to get an answer. Traditional approaches to modeling continuous exposure data have either assumed linearity, or converted the exposure into categories, such as quintiles. Linear models make full use of the available data, and hence have greater power than quintiles. On the other hand, by assuming linearity, they fail to address the issue at hand. Quintiles give some indication of the shape of the exposure–response curve, but such crude categorization, in addition to losing power, produces shapes that are sensitive to the choice of cutoffs between the categories. Using quintiles removes subjectiveness from the choice of cut points for categories, but not the sensitivity. Moreover, categorizing a continuous variable is a form of measurement error that can produce a distortion of the shape of the exposure–response curve.

Fortunately, the last 2 decades have seen an explosion in the statistical literature on nonparametric smoothing. Nonparametric smoothing represents a flexible alternative

to these approaches. It uses the data to determine the shape of the curve with minimal assumptions. This approach has begun to be more common in environmental epidemiology. For example, smooth functions were used to estimate the functional dependence of lung function on age (Schwartz *et al.* 1988; Wypij 1996) and air pollution (Schwartz 1989), of children's IQ on tooth lead (Schwartz 1993) or blood lead (Schwartz 1994), and of daily deaths on airborne particle concentrations (Schwartz 1994; Daniels *et al.* 2000). The basic idea of nonparametric smoothing is similar to that of categorization, with one key difference. Instead of estimating the average response in each of fixed categories of exposure, we can estimate the average response in moving windows of exposure. For example, instead of estimating blood pressure for each decade of life (20 to 29, 30 to 39, *etc.*), which assumes 39 year olds have more in common with 30 year olds than with 40 year olds, we can estimate blood pressure in moving 10 year windows about each year of age. This will trace a smooth curve without making any assumptions about its shape. More sophisticated smoothing algorithms are usually used, but they are essentially generalizations of this approach. The most important generalization involves generalized additive models, which implement this approach for multiple variables and for non-gaussian data. Figure 1 illustrates this approach. It shows the covariate adjusted exposure–response curve between particulate air pollution and daily deaths in Boston (Schwartz 1998). In this case, the association is quite linear. This is not always the case, however. Schwartz (1993) showed the relationship between blood lead and erythrocyte protoporphyrin levels in children in the Second National Health and Nutrition Examination Survey. It was clearly nonlinear, with a very shallow slope at blood lead concentrations below 17 μg/dl. This is shown in Figure 2.

These are examples of single studies, however, and I have argued that multiple studies are usually required to reduce uncertainty. Quintiles cannot readily be combined across studies because the boundaries of the quintiles will differ from study to study. A recent study (Schwartz and Zanobetti 2000) shows how nonparametric smoothed curves can be combined. It involves using pointwise meta-analyses to combine the curves at many exposure levels across the studies. The details have been published elsewhere. A key point was that simulation studies have shown that this approach was unbiased, and was adequate to detect thresholds when they exist. Table 1 shows the results of 500 simulations for each of three scenarios—a true linear relationship, a threshold at 20 $\mu g/m^3$, and a true logarithmic relationship. In each scenario, the true relationship was simulated for each of 10 studies. The table first shows the true relationship that generated the data. Then, a hierarchical model was used, where smoothed functions were fit to the data in each of the 10 studies in the scenario. Those 10 smoothed curves were combined, using the piecewise metasmoothing technique in the reference. To test how well the combined smoothed curve did in capturing the linear, threshold, or logarithmic fit that was the true relationship across the 500 simulations of each scenario, a regression model was fit in a third stage to the smoothed data points. For example, the smoothed points from the threshold simulations were fit to a piecewise linear model, to see if on average, the smoothing captured the fact that the relationship was zero below the cutpoint, and had the correct slope above it. These results are shown in the table for each of the three simulated relationships. For comparison, it also shows the results of fitting the true relationship to the simulated data, to indicate the degree of uncertainty in the estimates even when the true relationship is known. Further details have been

Table 1. Results of simulation study of meta-smoothing.

Dose Response Curve	True β	Meta-Smoothing	Model Based
Linear Model	.00150	.00150 [.00134, .00166]	.00150 [.00138, .00160]
Threshold			
Below	0	.00005 -.00021, .00032]	.00003 [-.00049, .00057]
Above	.00150	.00147 [.00130, .00164]	.00149 [.00136, .00163]
Logarithmic	.0400	.0367 [.0318, .0410]	.0402 [.0355, .0444]

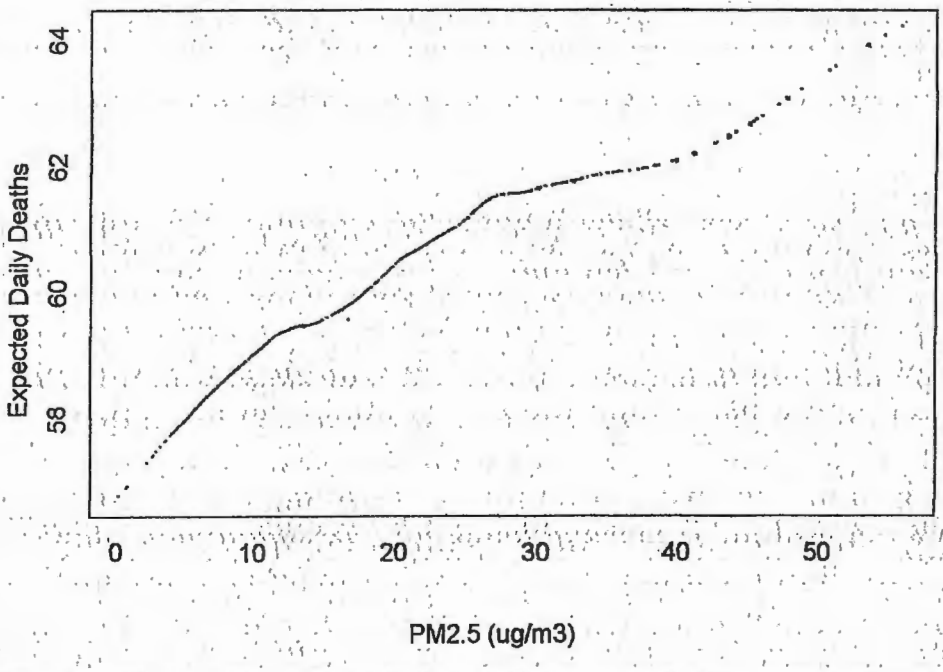


Figure 1. **The covariate adjusted daily deaths in Boston versus the $PM_{2.5}$ concentrations in the city on the current and preceding days.** No threshold is evident for the relationship, which exists entirely below the current ambient air quality standards of 65 μg/m³ for $PM_{2.5}$.

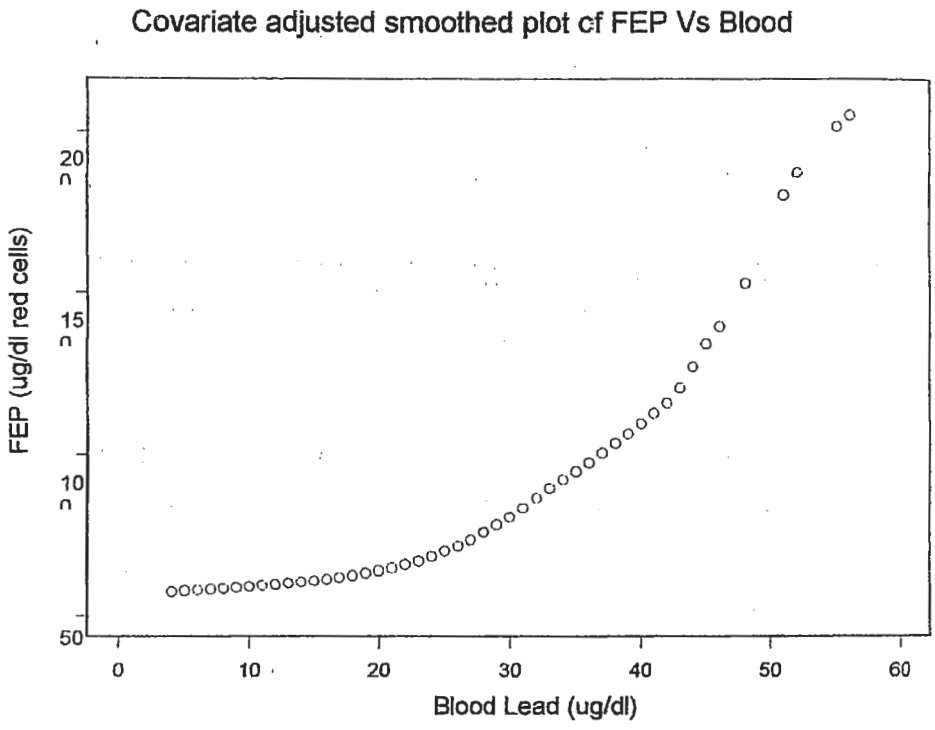


Figure 2. **The covariate adjusted dose-response between concentrations of free erythrocyte protoporphyrin levels in children aged 0 to 5 years, from the Second National Health and Nutrition Examination Survey and lead.** The adjustment was for sex, race, and a smoothed function of age.

published. Hence, by using more flexible methods to assess exposure–response relations and combining them across studies in hierarchical models, it is possible to learn a great deal about what those relations look like. Figure 3 shows the results of applying this technique to the relation between daily hospital admissions for heart disease and PM_{10} in 10 US cities (Canton, OH; Birmingham, AL; Chicago, IL; Colorado Springs, CO; Detroit, MI; Minneapolis/St. Paul, MN; New Haven, CT; Pittsburgh, PA; Seattle, WA; and Spokane, WA.) This analysis allows for a random effect that reflects heterogeneity in the true relationship across cities. In this case, the evidence for a linear association at low exposures is compelling. It also provides some indication of a higher slope at lower concentrations.

For existing data, this approach may involve rerunning regression models to obtain smoothed dose-response curves in each study. Modern computing makes this a relatively small burden, and only the estimated covariate adjusted curves need to be sent to the coordinating center for the second stage.

WHO IS SUSCEPTIBLE TO AIR POLLUTION?

Zanobetti and co-workers (2000a) analyzed daily counts of hospital admissions for cardiovascular disease (International Classification of Disease 9-th revision, 390-429), chronic obstructive pulmonary disease (ICD-9: 490-496, except 493) and

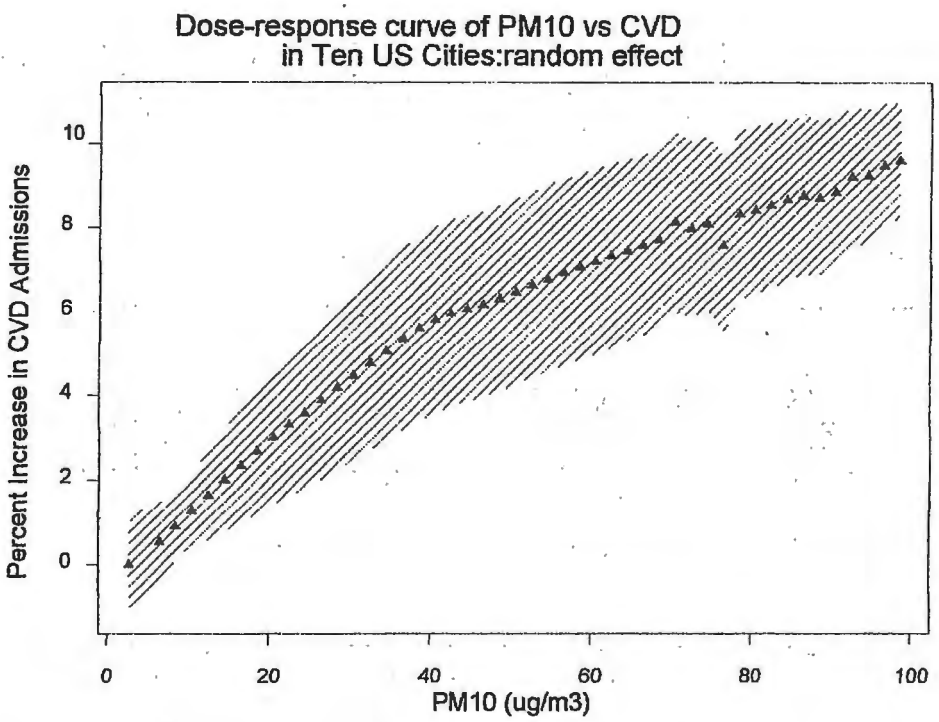


Figure 3. The covariate adjusted relative risk of hospital admissions for heart disease versus PM_{10} concentration in 10 US cities. The results average the concentration–response curves across the 10 cities allowing for heterogeneity. The risk is defined as one for days with PM_{10} concentrations below 4.5 μg/m³.

pneumonia (ICD-9: 480-487), in persons aged 65 years and older in the same ten cities. They built city-specific models including season, weather variables (temperature, relative humidity and barometric pressure) and day of the week. They then fit a second stage regression to examine effect modification by socioeconomic status:

$$\hat{\beta}_i = \beta^* + \delta P_i + \varepsilon_i$$

where $\hat{\beta}_i$ is the estimated PM_{10} effect in city i, P_i is the socio-economic index in that city, and inverse variance weighting is used to estimate the coefficients. δ then tells us how much the effect of PM_{10} changes for a unit increase in the social index. They considered the percentage of the population living below the federal poverty level, the percentage with college degrees, and the percentage of the population that was non-white as potential modifiers of the effect of PM_{10} on hospital admissions of the elderly. As seen in Figure 4, none of them was even suggestive as an effect modifier. This was not the case for medical conditions. For example, Figure 5, taken from Zanobetti *et al.* (2000b), shows the increased risk of an admission for COPD associated with a 10 μg/m³ increase in PM_{10} in all elderly subjects in Chicago, and in subsets defined by pre-existing medical conditions. Clearly, the presence of heart

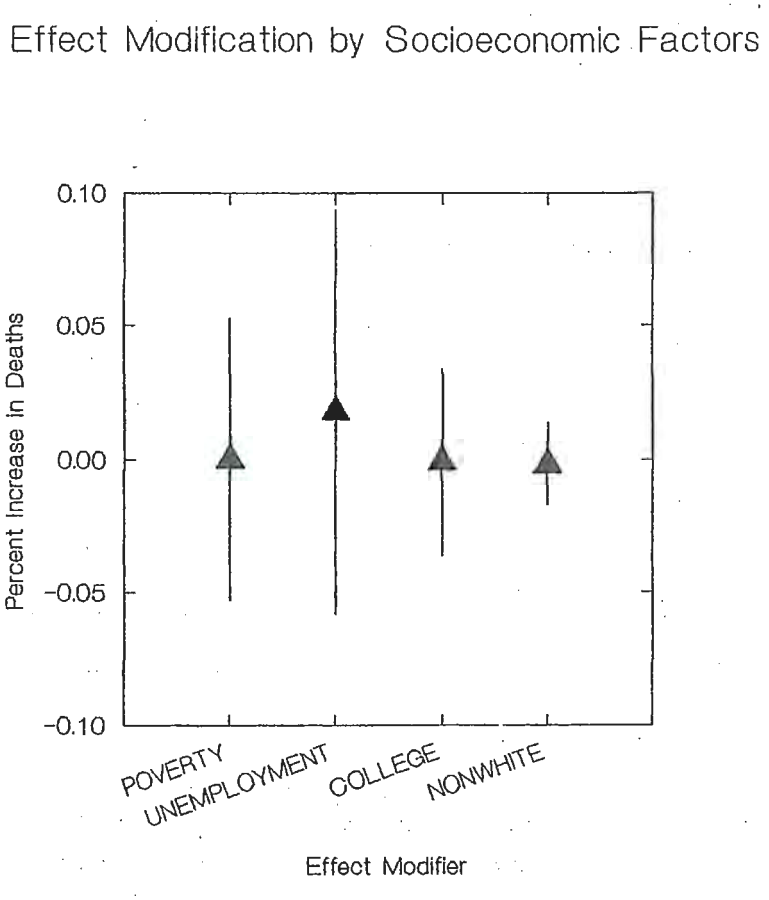


Figure 4. **The increase in the effect of PM_{10} on daily deaths for a 5 percentage point increase in the population living below the poverty level in a city, a 5 percentage point increase in the population with college degrees, and a 5 percentage point increase in the unemployment rate.** None of these factors modifies the effect of PM_{10} .

Effect Modification by Current and Prior Conditions: Chicago Hospital Admissions

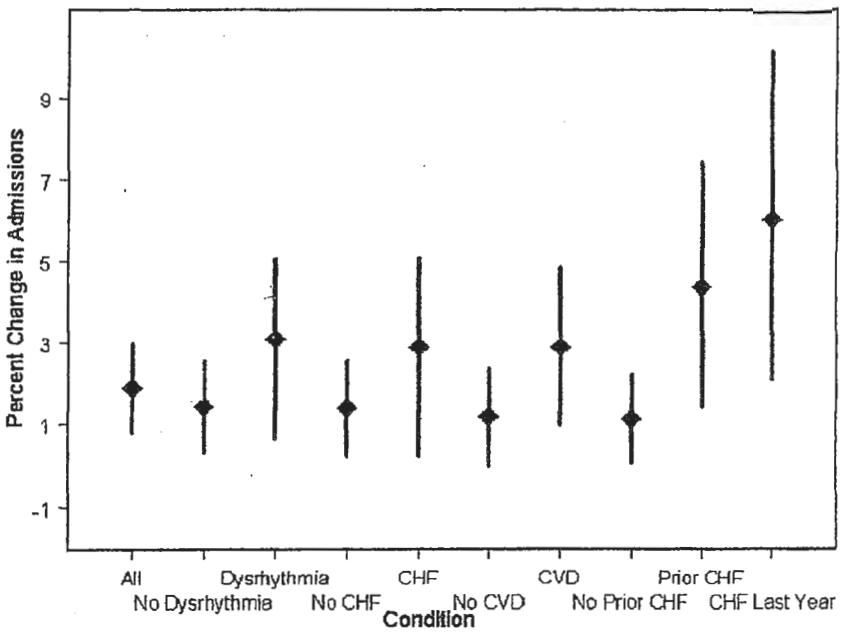


Figure 5. A covariate adjusted risk of hospital admissions for COPD for a 10 $\mu g/m^3$ increase in PM_{10} for subjects with and without specific co–morbidities. Heart failure appears to modify the risk of COPD admissions.

failure, and particularly an admission in the last year for heart failure, modifiers the risk of PM_{10} associated admissions for COPD.

WHAT POLLUTANT IS DOING IT?

A second use of multiple studies is to gain power, not just to assess the mean effect of an exposure, but to assess the mean effect of multiple correlated exposures. This has been an important issue for particulate air pollution. Many have argued that regulation is inappropriate not because there is no evidence of mortality risk from exposure, but because we may end up regulating the wrong source. What if particles from some sources were more or less toxic than average? Since the issue is inappropriate regulation of a source, one approach would be to identify exposure due to individual sources, rather than total exposure to fine particles, or even exposure to subspecies defined by chemical composition or size. A recent paper (Laden *et al.* 2000) took that approach using data from six US cities. They used elemental analysis of the chemical composition of the fine particle filters on each day in each city. A factor analysis in each city was used to identify factors that represent contributions from identifiable sources. These included motor vehicle exhaust, long range transport particles from coal burning power plants, particles from residual oil combus-

tion (only present in two cities), and particles from fine dust and soil. The mass contribution of each source on each day, in each city, was estimated by regressing daily total fine particle mass in that city on each factor score, and computing the mass contribution of each factor using those regression coefficients. These separate mass concentrations were then simultaneously included in regression analyses relating their mass concentration to daily deaths in each city. Estimating them in six locations, and combining the effects across the six cities produced greater stability of the resulting coefficients. Figure 6 shows the results, which clearly demonstrate that both traffic particles and particles from coal burning power plants are associated independently of daily deaths. Each effect estimate is adjusted for the effects of the other three types of particles.

What if the health effects are associated not with particles but with other air pollutants? Here again, multistage models provide additional power to answer this question. Using the same 10–city analysis described before, Zanobetti and coworkers (2000b) analyzed the potential for confounding by other pollutants. As Figure 7 indicates, there is no evidence of any confounding. This is an important result, particularly because this can rarely be assessed in individual city analyses. The correlation among the pollutants is too high, and stochastic variability can drive

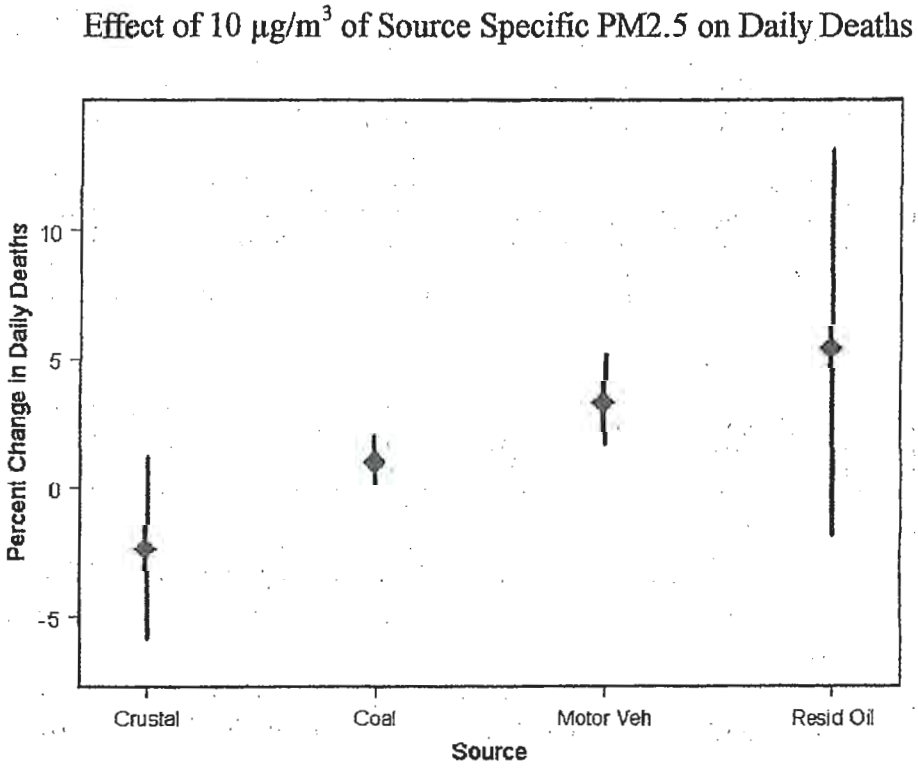


Figure 6. **The effects of particles from traffic, coal, residual oil, and windblown dust on daily deaths in six US cities.** The effect of each particle source is after control for the other sources.

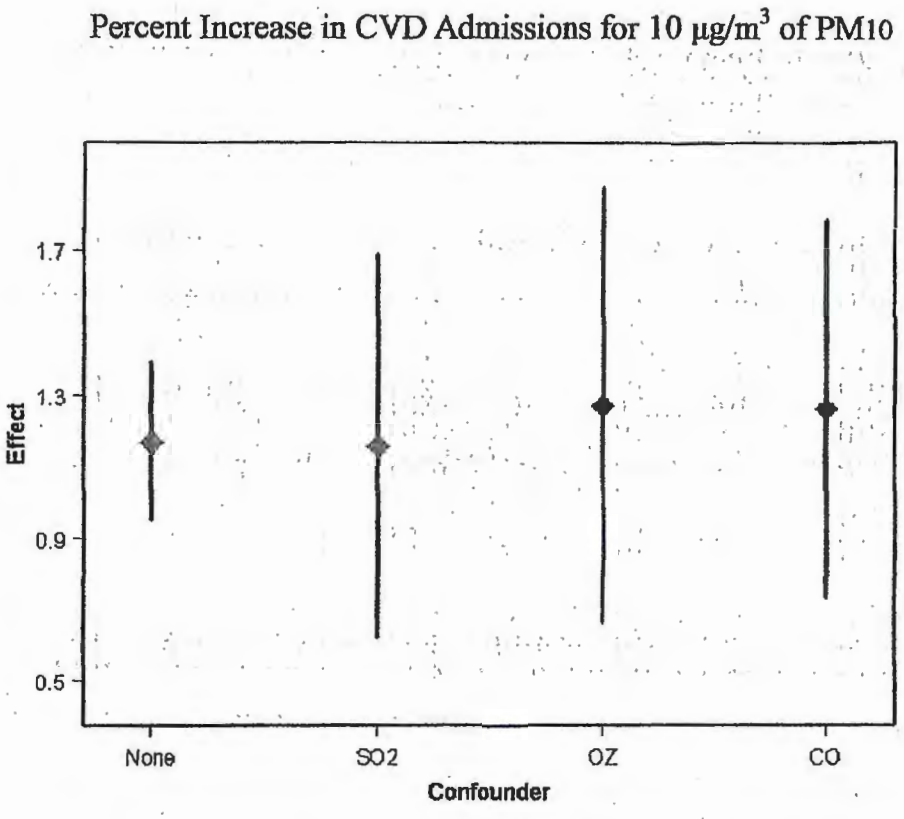


Figure 7. **The effect of PM_{10} on hospital admissions for cardiovascular disease before and after adjustment for potential confounding by CO, SO_2, NO_2, and O_3. No evidence of confounding is seen.**

results in individual cities. Even negative associations are possible. It is only when these results are combined across multiple cities that stable answers are obtained. This was demonstrated much earlier in the analyses of Schwartz and Marcus (1990). They analyzed the association between black smoke and SO_2 and daily deaths in London using 14 years of data. In two pollutant models for each year, wildly variable results were seen. Some years SO_2 would have a large effect and black smoke would appear protective. Other years the opposite was seen. Only when the results were combined over all years in an empirical Bayes analysis was a clear result obtained. In that case, there was no association with SO_2, only with black smoke.

CONCLUSIONS

Epidemiology is conducted in the species of interest, at or close to the exposure range of interest, and with the co–exposures of interest. It directly relates exposure to response without the need for modeling from dose–response relations. This should make it the preferred source of concentration–response for risk assessments.

While confounding is possible in epidemiology, the creative use of multiple studies provides a way to address this issue, as demonstrated above. Moreover, recent history suggests that this concern is overblown. When early studies of lead and children's IQ indicated an adverse effect, it was argued that this was due to poorly controlled socio–economic confounding. When Bellinger and co–workers (1987) conducted a study where the higher lead levels were in the higher socio–economic group, they found an even larger effect. Concerns that the association between airborne particles and daily deaths were due to inadequate control for weather and season were likewise put to rest in sensitivity analyses that either excluded extreme weather days (Schwartz 1998) or considered sensitivity to alternative control measures (Samet *et al.* 2000; Samet *et al.* 1998)).

The larger sample sizes than are available for animal studies allow the use of non-parametric smoothing to gain insight into the shape of the concentration–response relation at or near the exposure ranges of interest. These large sample sizes also allow the investigation of sensitive subgroups of individuals within the context of a single study, whereas animal toxicology would require multiple expensive studies to address as many potential effect modifiers.

It is often argued that epidemiology is insensitive at the low exposure levels that are mostly evaluated in risk assessment. Recent epidemiology has not borne this out. Studies have identified low relative risks of adverse pregnancy outcomes associated with disinfection byproducts (Swan and Waller 1998), of low relative risks of lung cancer associated with air pollution exposure (Cohen and Pope 1995), and of modest changes in cardiovascular risk factors associated with acute air pollution exposure (Gold *et al.* 2000; Schwartz 2001). The principal limitation is currently cultural. Old perceptions are impeding the funding of epidemiological studies for carcinogens, resulting in few risk assessments based on them. For non-carcinogens, this has not been the case, and the cancer community needs to revise its old perceptions to reflect new realities.

REFERENCES

Bellinger D, Leviton A, Waternaux C, *et al.* 1987. Longitudinal analyses of prenatal and postnatal lead exposure and early cognitive development. N Engl J Med 316:1037-43

Cohen AJ and Pope CA. 1995. Lung cancer and air pollution. *Environ Health Perspect* 103(suppl 8):219-24

Cory-Slecta DA and Widzowski DV. 1991. Low level lead exposure increases sensitivity to the stimulus properties of d1 and d2 agonists. Brain Res 553:65-74

Cory-Slechta DA, Wodzpwski DV, and Pokora MJ. 1993. Functional alterations in dopamine systems assessed using drug discrimination procedures. Neurotoxicology 14:104-14

Daniels MJ, Dominici F, Samet JM, *et al.* 2000. Estimating particulate matter-mortality dose-response curves and threshold levels: an analysis of daily time-series for the 20 largest US cities. Am J Epidemiol 152(5):397-406

Gold DR, Litonjua A, Schwartz J, *et al.* 2000. Ambient pollution and heart rate variability. Circulation 101:1267-73

Laden F, Neas LM, Dockery DW, *et al.* 2000. Association of fine particulate matter from different sources with daily mortality in six U.S. cities. Environ Health Perspect 108(10):941-7

Lasley SM, Polan-Curtain J, and Armstrong DL. 1993. Chronic exposure to environmental levels of lead impairs in vivo induction of long-term potentiation in rat hippocampal dentate. *Brain Res* 614(1-2):347-51

Levin R. 1986. Reducing Lead in Drinking Water: A Benefit Analysis. EPA-230-09-86-019. U.S. Environmental Protection Agency, Washington, DC, USA.

Levin R and Kleiman MAR. 1999. Drinking water treatment: balancing infectious disease, cancer, and cost. In: SS Nagel SS (ed), The Substance of Public Policy, pp 117-37. Nova Science Publishers, NY, NY, USA

Pocock SJ, Smith M, and Baghurst P. 1994. Environmental lead and children's intelligence: a systematic review of the epidemiological evidence. *Brit Med J* 309(6963):1189-97

Rice DC. 1985. Chronic low lead exposure from birth produces deficits in discrimination reversal in monkeys. Toxicol Appl Pharmacol 77:201

Samet J, Zeger S, Kelsall J, *et al.* 1998. Does weather confound or modify the association of particulate air pollution with mortality? An analysis of the Philadelphia data, 1973-1980. Environ Res 77:9-19

Samet JM, Zegar SL, Dominici F, *et al.* 2000. The national morbidity, mortality, and air pollution study part II: Morbidity, mortality, and air pollution in the United States. Health Effects Institute 94:1-84

Schwartz J. 1989. Lung function and chronic exposure to air pollution: A cross-sectional analysis of NHANES II. Environ Res 50:309-21

Schwartz J. 1993. Beyond LOEL's, p-values, and vote counting: Methods for looking at the strengths and shapes of associations. Neurotoxicol 14:237-46

Schwartz J. 1994. Particulate air pollution and daily mortality in Cincinnati, Ohio. Environ Health Perspect 102:186-9

Schwartz J. 1998. Health effects of particulate air pollution: Is there a threshold? In: Mohr U (ed), Relationship between Respiratory Disease and Exposure to Air Pollution. ILSI Press, Washington, DC, USA

Schwartz J. 2001. Air pollution and blood markers of cardiovascular risk. Environ Health Perspect 109(suppl 3):405-9

Schwartz J and Marcus A. 1990. Mortality and air pollution in London: A time series analysis. Am J Epidemiol 131:185-94

Schwartz J and Zanobetti A. 2000. Using meta-smoothing to estimate dose-response trends across multiple studies, with application to air pollution and daily death. Epidemiology 11(6):666-72

Schwartz J, Katz S, Fegley R, *et al.* 1988. Analysis of spirometric data from a national sample of healthy 6-24 year olds. Am Rev Respir Dis 138:1405-14

Schwartz J, Pitcher H, Levin R, *et al.* 1985. The Costs and Benefits of Reducing Lead in Gasoline. EPA 230-05-85-006. U.S. Environmental Protection Agency, Washington, DC, USA

Swan SH and Waller K. 1998. Disinfection by-products and adverse pregnancy outcomes: what is the agent and how should it be measured? *Epidemiology* 9(5):479-81

Wypij D. 1996. Spline and smoothing approaches to fitting flexible models for the analysis of pulmonary function data.. Am J Respir Crit Care Med 154(6 Pt 2):S223-8

Zanobetti A, Schwartz J, and Dockery DW. 2000. Airborne particles are a risk factor for hospital admissions for heart and lung disease. Environ Health Perspect 108:1071-7

Zanobetti A, Schwartz J, and Gold DR. 2000. Are the sensitive subgroups for the health effects of airborne particles? Environ Health Perspect 108:841-5

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1267-1293 (2002)

Issues in Exposure and Dose Assessment for Epidemiology and Risk Assessment

Thomas J. Smith
Environmental Science and Engineering Program, Harvard School of Public Health, Landmark Center, 404F, P.O. Box 15677, Boston, MA 02215, USA; Tel(voice):617-384-8804, Tel(fax): 617-384-8859; tsmith@hohp.harvard.edu

ABSTRACT

Epidemiologic studies have been effective in identifying human environmental and occupational hazards. However, most epidemiologic data has been difficult to use in quantitative risk assessments because of the vague specification of exposure and dose. Toxicologic animal studies have used applied doses (quantities administered, or exposures with fixed duration) and well characterized end points to determine effects. However, direct use of animal data in human risk assessment has been limited by uncertainties in the extrapolation. The applied dose paradigm of toxicology is not suited for cross species extrapolation, nor for use in epidemiology as a dose metric because of the complexity of human exposures. Physiologically based pharmacokinetic (PBPK) modeling can estimate the time course of tissue concentrations in humans, given an exposure-time profile, and it has been used for extrapolating findings from animals to humans. It is proposed that human PBPK modeling can be used in appropriately designed epidemiologic studies to estimate tissue concentrations. Secondly, tissue time courses can be used to form dose metrics based on the type and time course of adverse effects. These dose metrics will strengthen the determination of epidemiologic dose-response relationships by reducing misclassification. Findings from this approach can be readily integrated into quantitative risk assessment.

Key Words: exposure assessment, PBPK modeling, dose metric, epidemiology.

INTRODUCTION

The process of risk assessment for a toxic chemical is a complex blending of information from several sources shown in Figure 1, which then fits into the risk management process to determine interventions. Human data are critical in this process for determining what allowable exposures should be, and how much intervention is needed. Unfortunately, it has often been difficult to use epidemiologic data because of the vague specification of dose in epidemiologic dose-response

1080-7039/02/$.50

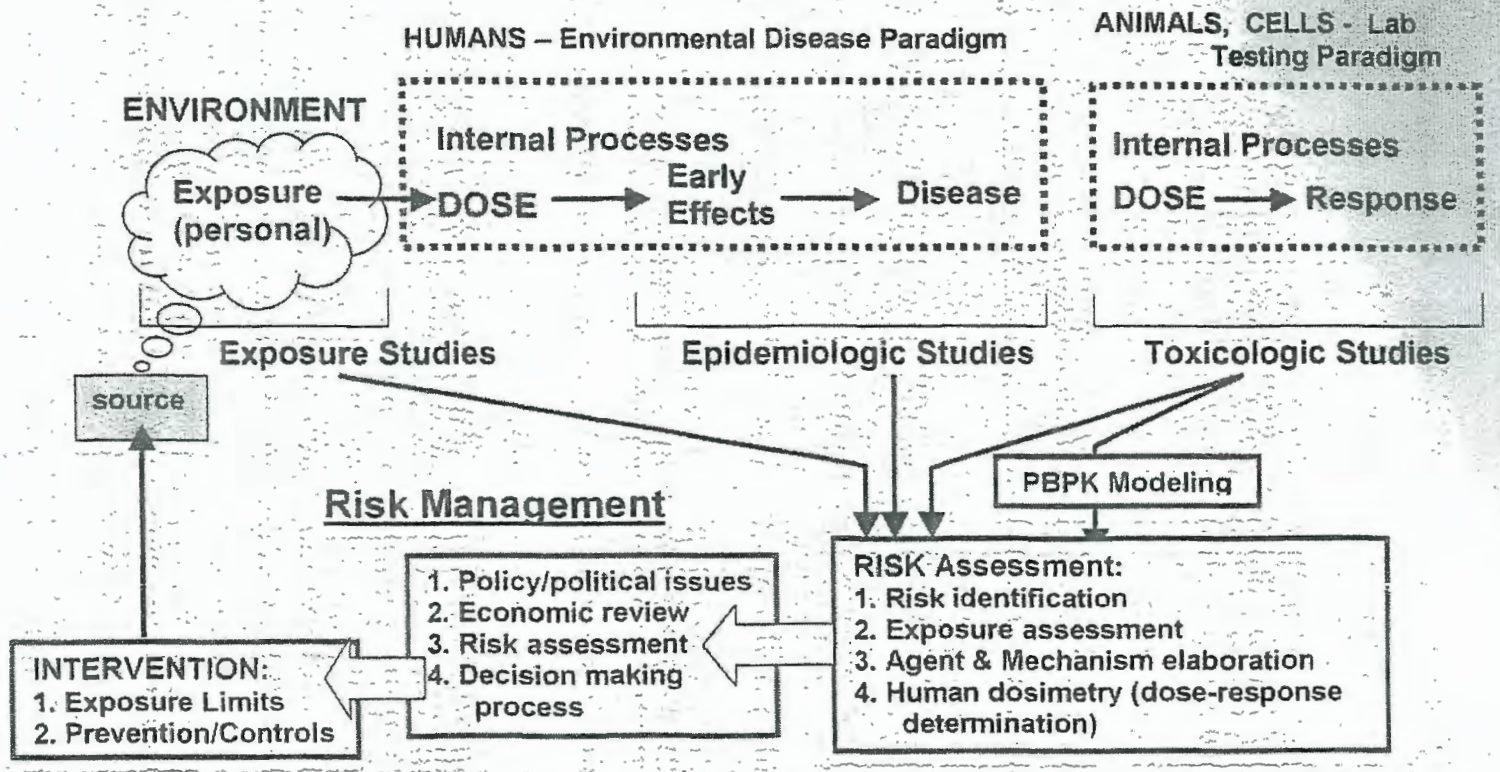


Figure 1. **The relationship between the human paradigm for environmental disease and the range of research approaches used to study exposure-disease processes and their input to risk management.** Note dose is well defined for toxicologic studies but not for epidemiologic studies.

relationships, where the **dose** is a quantity of the chemical taken into the body over a period of time. Animal data also have been used to extrapolate quantitative dose-response relationships for humans, but there are large uncertainties because of the difficulty extrapolating animal metabolism and responses to humans.

The risk assessment-management approach and the process description of environmental disease (summarized in the top left portion of Figure 1) are paradigms formalized by National Academy of Science (NAS) panels (NAS 1983; NAS 1994). In addition to the human disease paradigm and the toxicologic testing model, the diagram also shows the three basic types of studies used to understand the contributing factors to environmental hazards. In the figure there is a clear gap between exposure and epidemiologic studies which leaves out evaluation of human dose and its role in the relationship between exposure and disease. It is proposed that better estimates of dose for epidemiologic studies, will strengthen determinations of risk, and can provide a means to better integrate toxicologic and epidemiologic findings in risk assessment.

Exposure assessment, epidemiology, and toxicology are applied to different parts of the disease paradigm (Figure 1), and have important differences in approach. Exposure assessment is focused on characterization of individual and group exposures for broad classes of materials, such as particulate matter less than 2.5 μm in diameter, and selected chemicals in general and occupational environments. Epidemiology uses human population studies to detect broad factors, such as job title or residence location, to identify those associated with increase risks of adverse responses and/or disease. Toxicology uses laboratory investigations of moderate to high doses of toxic chemicals in animals, and tissues and cells to determine mechanisms of toxic effects. Risk assessment uses statistical and pharmacologic approaches to integrate information from all of these sources to formulate a quantitative dose-response relationships for specific exposures and health risks. Formulation of a risk management strategy will use this relationship plus political, economic, and policy considerations to develop a plan that will ultimately guide interventions to prevent and control exposures. While there is clear overlap and relevant information generated by each type of investigation for risk assessment, there are often uncertainties about how to integrate the information from a specific study. One of the most common problems is how to extrapolate from high-dose animal studies to low exposure human risk (Klaassen *et al.* 1986). Another common problem is how to improve the quantification of dose-response relationships in epidemiologic studies. The relationship between human exposure and dose is central to both of these problems.

The scientific process leading from the identification of a new chemical hazard to the development of a risk assessment can be considered in four parts. First is the risk identification phase, when simple epidemiologic studies and toxicologic bioassays are conducted to detect potential risks. If a study suggests a new hazard, then more focused hazard verification studies may be conducted. The second part is the exposure assessment phase when possible agents and exposed populations are identified. Third is the agent and mechanism elaboration phase, when mechanistic studies are conducted and pharmacokinetic models are developed. The second and third parts can be conducted concurrently, if there are reasonable hypotheses about the agent(s). Given sufficient information and resources the fourth part is the

human dosimetry phase, when epidemiologic studies with full dosimetry are conducted to determine the human dose-risk relationship. The types of studies conducted in each part become progressively more complex and require more resources. There tends to be a large number of hazard identification studies, but only a limited number of human dose-response studies.

This paper addresses the design and concepts for epidemiologic dose-response studies, although many of the concepts can be used to guide hazard verification studies too. A strategy for better quantifying dose in epidemiologic studies and better linkage within risk assessment are strongly needed. The goal of this paper is to discuss how a better linkage of epidemiologic and toxicologic findings might be accomplished using a more precise definition of dose, and to provide some examples.

EPIDEMIOLOGIC DETERMINATION OF DISEASE RISK

Risk is an epidemiologic concept based on the behavior of diseases in populations, and it can only be determined in a population. An individual's personal likelihood of disease is conceptual, such as his risk of leukemia is one in a million, but he either gets leukemia, or not, no matter what the likelihood is. Checkoway and his associates defined risk as "the average probability of developing disease during some time interval." (Checkoway *et al.* 1989) It is well known that several important characteristics of a population can affect the disease risk: age, sex, race, genetic background, and life style variables such as cigarette smoking and alcohol intake (Checkoway *et al.* 1989). If members of a population also have an environmental exposure to a toxic chemical, then that personal characteristic may also modify risk. It is the limited approaches used to characterize population exposures that have caused problems for using those findings in risk assessment.

Epidemiologists have developed standardized methods to adjust for population differences in common personal attributes, such as age and sex, where the effects of these factors are known (see Checkoway *et al.* 1989 for a more detailed discussion). One common approach for epidemiologic detection of an environmental risk of a disease is to identify an exposed population (a cohort) and then compare the disease risk of that population with an unexposed population with similar characteristics of age, race, sex, *etc.* Figure 2 diagrams the basic steps in processing data for this type of analysis. Exposure is assigned to each subject based on his or her personal background using an algorithm that often is very simple, such as a person's job title or residence location define his or her exposure status. If the distribution of disease between the exposed and control populations is sufficiently different and unlikely to be due to chance, then there is evidence that a relationship may exist, but it cannot prove that the relationship is causal. In this analytical context, "exposure" is treated as an event or a stable personal characteristic (a defined distribution with a fixed mean and dispersion) associated with a time period, which varies across subjects. In some cases, the population exposures, such as air concentrations, and changes in the average across time are available or can be estimated. For chronic diseases, very often these data are summarized in a cumulative exposure metric, which is calculated as the average concentration times the duration of exposure. The cumulative exposure is analogous to the administered dose used by toxicolo-

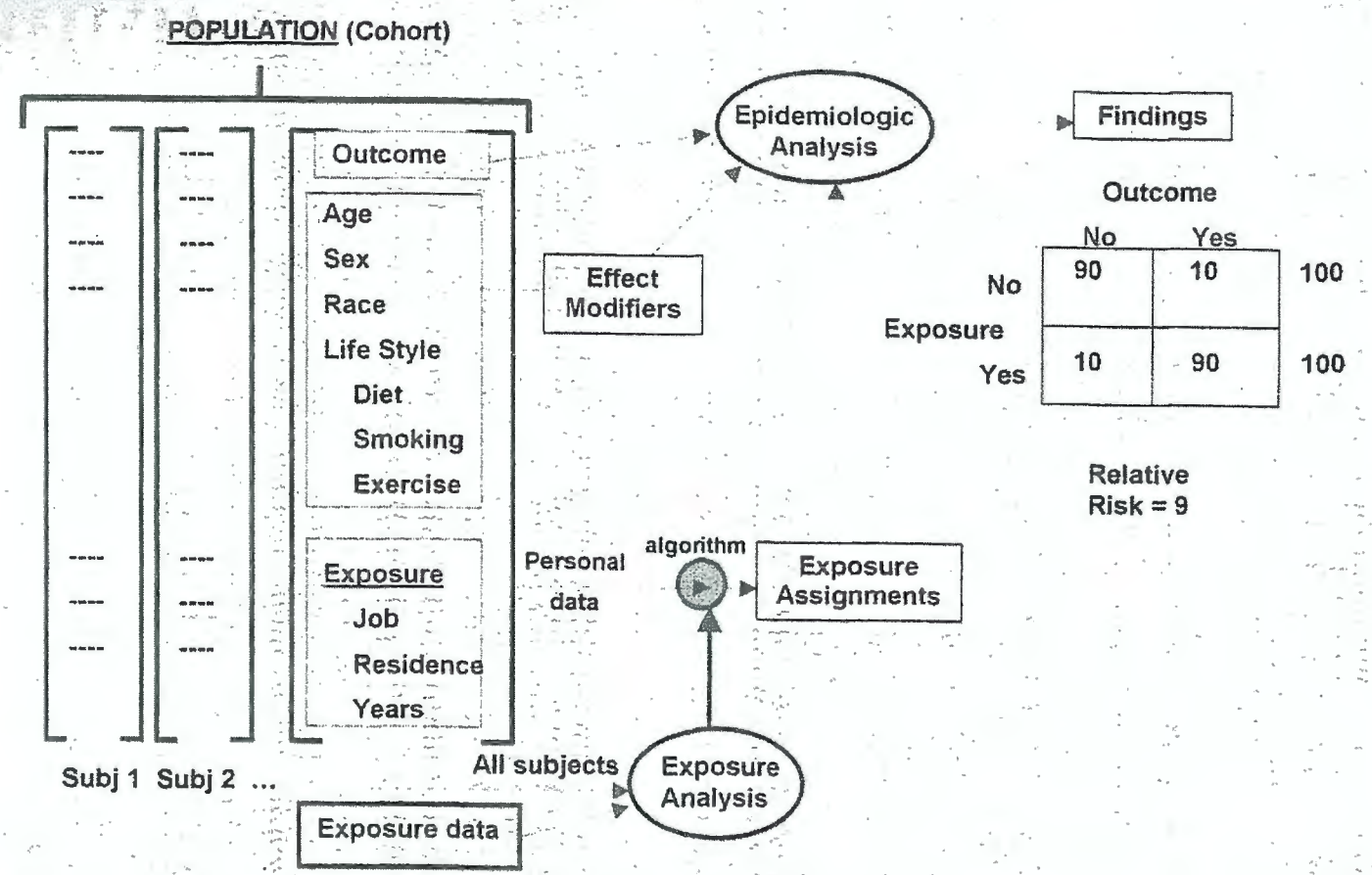


Figure 2. **Standard epidemiologic analysis for risk detection by a cohort study.** The ovals are data processes, and the rectangles are data or results. Exposure analysis provides the algorithm to make the exposure assignments.

gists in laboratory exposures to airborne gases and particles. For the remainder of this paper, "**dose metric**" will be used to describe a summary measure calculated from a concentration-time course, such as cumulative exposure, this metric can be tested to determine if there is a quantitative association with the intensity of a response or the risk of an effect (Kriebel 1994).

As will be discussed in detail below, exposure is not equivalent to dose, even though some investigators have not made clear distinctions. Our goal is to approximate as closely as possible the relevant dose to the target tissues because that will give the strongest relationship with the magnitude or risk of the effect.

The epidemiologic approach is a very powerful way to detect previously unrecognized health risks in populations. It is perhaps the most frequent method by which new environmental health risks have been detected (Checkoway *et al.* 1989). It has several advantages. First, epidemiology can be applied without knowing the agent of the effects or its mechanism of action. For example, a large number of studies with different designs and study populations have shown an impressively consistent, overall increased risk of lung cancer among occupations with low level diesel exhaust exposure (Bhatia *et al.* 1998; Lipsett and Campleman 1999). Second, some types of epidemiological studies can be relatively inexpensive to conduct, especially where data on the population are available from disease registries or other large population tracking devices. Third, epidemiologic investigators can use a variety of inexpensive semi-quantitative dose markers to selected "exposed" groups, or to partition the study population into subgroups that are believed to have different exposures. Different job or work area titles or residence locations are often used as markers for exposure intensity or for probability of exposure, and presumably different doses. Temporal variables that are indicators of exposure duration are common, such as duration of work in a job with exposure, or duration of residence in an environment with contamination, which can be easily determined. The investigators in diesel occupation studies have used job titles, such as "ever worked as a truck driver", as markers for potential exposure with the expectation that differences in potential exposure also represent differences in dose. Simple dose markers must be used with care because they can contain considerable mis-classification unless the exposed population has very high exposures, such as coke oven workers or asbestos insulators, and the duration of exposure is sufficient to accumulate a large dose, which are rarely checked (Smith 1992). It has not been unusual for epidemiologic investigators to conclude that there was no evidence of a "dose-response relationship" between exposure and risk because there was no increase in risk associated with an increase in a simple dose marker, such as years of exposure. Unfortunately this ignores the frequently large dose misclassification in potential exposure and simple temporal dose markers.

While the traditional epidemiological approach is excellent for detecting increased risks in highly exposed populations, it is less useful for identifying the causes of the risk and quantifying dose-response relationships. Even where there is detailed exposure information, it is often not clear how to summarize the data into dose metrics (Kriebel 1994). One of the limitations of environmental epidemiology has been the difficulty of reducing the multidimensional complexity of exposure to meaningful empirical summary dose metrics, although some empirical metrics have been developed (Seixas *et al.* 1993). Because of this, traditionally constructed

epidemiologic studies are often not useful for quantifying how much exposure causes a unit increase in risk. For example, the epidemiologic studies of diesel exposed occupations have shown evidence of increased risk, but the agents and characterization of the dose-response relationships have been so crude that they are not directly useful for risk assessment (HEI 1999). With assumptions about the nature of exposure, sometimes useful inferences about the dose-response relationship can be made. This is a critical limitation on the input needed to define a risk relationship needed for the risk management strategy shown in Figure 1. This lack of detail has also limited the linkage of epidemiological studies with the much more detailed laboratory studies where dose is carefully specified.

One of the objectives of this paper is to propose a method for better quantifying epidemiologic dose-response relationships. To do this we must better characterize exposure and develop suitable methods to formulate meaningful dose metrics.

RISK ASSESSMENT

The risk assessment paradigm is that a period of exposure produces a risk, which is analogous to the toxicologist's applied dose concept (Klaassen *et al.* 1986). As will be discussed in detail later, this is a weak link because there is often not a strong relationship between the external applied dose and the concentration in the target tissue that causes the effects. Although both intensity and duration contribute to risk, risk assessment and management focus on exposure intensity because that is where intervention is practical. Outside of some occupational exposures, it is not practical to limit the duration of an individual's exposure.

In the next three sections, the characteristics of exposure, internal levels of contaminants or their active metabolites resulting from exposure, and dose concepts will be developed. Then in the final section these will be integrated with epidemiologic approaches. The goal is to demonstrate that tools now exist to study human dose-response relationships, and that with better definition of dose in human studies, better dose-response epidemiology can be done, and the findings will integrate better with risk assessment.

CHARACTERISTICS OF EXPOSURE

Exposure is a process, a period of external contact with an environmental contaminant. The defining contact occurs at the point of entry for the toxic substance: air contaminants in the breathing zone near the nose and mouth; food and drinking water contaminants for ingestion; and skin contamination for substances that can be absorbed. Airborne exposures are the most common source of heath risks because the respiratory system efficiently moves airborne materials directly into the arterial blood without many intervening protective processes. The presence of a contaminant in the environment is not equivalent to exposure. Exposure only occurs when the contaminants reach an individual in a suitable form and conditions that they may enter the body. High concentrations of sulfur dioxide leaving a high power plant smoke stack are an environmental problem, but do not constitute exposure until they reach an area where humans can inhale them. As shown in Figure 1, the exposure process is a consequence of contaminants entering

the environment and eventually reaching a point of entry of an individual. Once released, emissions in air and water can be transported, diluted, and chemically modified before they reach the exposure setting. Skin exposure may be a consequence of direct contact with a contaminated surface without transport through the environment, or the result of deposition of air or water contaminants on the skin, *e.g.*, tar globules from an oil spill.

Exposure has four dimensions: composition, physical form (particulate, liquid, or gaseous), environmental concentration (intensity), and their variations across time. Because emissions are often sporadic, highly concentrated at the point of emission, and incompletely mixed with environmental media, exposures are often characterized by considerable heterogeneity and variability over time. Measuring and extrapolating exposures is a major task for epidemiologic studies and frequently takes considerable study resources. In an ideal epidemiologic study, one would wish to measure the exposures of every subject for all potentially relevant agents for the whole time period relevant to the disease. However this is not feasible for most diseases and adverse responses. Even if possible, it is usually not necessary to measure every subject's entire exposure because there are statistical regularities that can be used to reduce the number of measurements needed and make the design more efficient, and simplify the problem of estimating the time profile of exposure for each subject.

When exposures are the result of emissions from a single type of source in physical settings with similar transport and dispersion characteristics, and the activities of the exposed individuals are well defined, then the measured distributions of exposure across time tend to be similar, if not the same. Thus, exposures in this situation are "stationary", which means that the probability distribution across time is a stable, usually lognormal distribution with an approximately fixed geometric mean and geometric standard deviation (Rappaport 1991). However, exposure distributions may also vary among individuals in a common setting, because the individual's activities can influence the exposure process by variations in their actions and the way they conduct activities. For example, two individuals sweeping up household dust in equally contaminated houses may systematically differ in their exposures because one is a vigorous sweeper and the other is more gentle. Even with stable exposure situations, there can also be important variation on different time scales, such as hourly, daily, and seasonally because of variations in the sources and/or transport processes.

Although basic epidemiologic designs, such as the one shown in Figure 2, do not require identification or measurement of toxic agents, a study that includes exposure measurements will require the selection of the materials to be measured, the measurement method, and a sampling strategy. The choice of measurement approach for complex mixtures, such as diesel exhaust can be a major problem, especially where the agent is unknown. Commonly, a marker compound is chosen that is uniquely associated with the mixture and can be easily measured. It is assumed that variation in the marker is proportional to the agent of effects. Unfortunately, it is rare that a unique marker can be found. When emissions from the source of interest predominate, then the marker may be highly useful, but when other emissions are dominant, then the levels of the marker may be only poorly correlated with the agent.

Exposure assessment for epidemiology has the goal of estimating either individual exposures, or more often identifying groups with significantly different exposure distributions, and within which individuals have exposures with the same composition and similar geometric mean and geometric standard deviation values. Note that two individuals in the same exposure situation, will not have identical exposures in each time period, but will have exposures drawn from the same distribution across time, so that their long-term exposures will have approximately the same geometric mean. This becomes important when designing an epidemiologic study to detect effects of an exposure situation. The nature of exposure and its variation will be illustrated with some examples: diesel exhaust where the composition is variable, and butadiene where the concentration is highly variable.

Exposures to Diesel Exhaust in the Trucking Industry

Diesel exhaust is a common air contaminant in occupational and general environments. It is a good example of the complexity of a typical exposure assessment problem. Concern about exposure to diesel exhaust is derived from the observed pattern of elevated epidemiologic risk and the evidence of mutagens and known human carcinogens in diesel particulate (HEI 1995). While diesel exhaust has been treated as if it was a mixture with a fixed composition by the regulatory process of the U.S.Environmental Protection Agency (USEPA) and the state of California, in reality the composition is not fixed.

Diesel exhaust has a complex and variable composition of particle, inorganic gases, and organic vapors (HEI 1995). The composition of the emissions from a given diesel engine depend on the engine type, fuel composition, and how the engine is operated (HEI 1995; Shi *et al.* 2000; Wang *et al.* 2000). Older diesel engines release more emissions and have different composition from those of newer engines because of engine wear and changes in design. Older high sulfur fuels produced more particulate emissions than newer fuels. Idling engines emit moderate levels of particles but high levels of organic vapors, primarily unburned fuel, whereas engines under moderate load operated at steady highway speeds, have both low particulate and vapor emissions. Further, the emitted materials undergo changes with time after release into the atmosphere: particles agglomerate with each other and with ambient particles, and vapors condense on, or evaporate from the particles depending on the volatility of the hydrocarbons. As a result, exposures to emissions from trucks in stop-and-go traffic are qualitatively and quantitatively different from those during highway operations. While all of these factors affect exposure significantly, if we have an exposure situation where the engine type, age, fuel, and operating conditions are defined, then the composition will be reasonably consistent. When determining the health effects of diesel emissions, these different exposure conditions need to be distinguished to avoid confusing the effects of different chemical components with differences in concentration.

Chemical and physical characterization of exposures to diesel exhaust are a difficult analytical problem that required collection of materials in the field and laboratory analysis (HEI 1995; Verma *et al.* 1999). Collection of reactive air contaminants on filters and cryogenic traps can alter the composition (Wongphatarakul *et al.* 1998). Sophisticated separation methods and gas chromatographic-mass spectro-

metric analytic techniques must be used to separate and identify the major components of the hundreds of compounds present (Schauer *et al.* 1996; Fraser *et al.* 2000; Kleeman *et al.* 2000; Schauer and Cass 2000). Non-diesel engines and other combustion sources will also emit many of the same products as diesels, but in different proportions. As a result, an analytical technique called source apportionment may be needed to determine the relative proportions of materials from the different sources (Schauer *et al.* 1996; Schauer and Cass 2000). Since the toxicity of the compounds varies widely, the analytical chemistry alone cannot determine where the hazard lies for a complex exposure.

Given this complexity of composition, exposure and epidemiologic investigators have sought an exposure marker for diesel exposure. An exposure marker is quite different from a dose metric, because exposure is external and dose is internal. Epidemiological researchers have not always made a clear distinction between these two types of markers. Diesel engines are major producers of black soot, elemental carbon (EC), which has been chosen as the exposure marker. However, the relative amount emitted is variable with engine type and fuel, and other combustion sources also produce EC. Where one type of diesel vehicles predominate, EC is a good marker, but where there are few diesels of mixed types, the EC levels are difficult to interpret without additional data.

Exposures to Butadiene in the Petrochemical Industry

In some occupational settings, the exposure composition may be relatively simple, a single contaminant, but variation over time is a concern. This is the case for workers in a petrochemical plant that only produces and processes a single chemical, 1,3-butadiene. Butadiene (BD) is a major petrochemical feed stock, and a common gaseous air contaminant, which is also a suspected human carcinogen (Fajen *et al.* 1990). A typical data-to-day pattern of exposure variation for one individual in a production plant is shown in Figure 3. This distribution shows the characteristic lognormal pattern of exposures that is common: half are below the geometric mean, 1.0 ppm, and there are regular very high exposures associated with infrequent job activities and workplace conditions (Turnbull *et al.* 1990; Ward *et al.* 1995; Sorsa *et al.* 1996).

Typically an epidemiologic exposure assessment focuses on making precise estimates of the mean exposure (not the geometric mean), and identifying factors that modify the mean under various conditions (Armstrong *et al.* 1992). The arithmetic mean of daily exposures (even lognormal values) times the number of exposure days and the duration of each daily exposure, times subject's pulmonary ventilation rate and the fraction retained will give an estimate of the amount of inhaled material retained (the cumulative dose) (Smith 1992). The arithmetic mean of values from a skewed distribution, such as the lognormal distribution, is sensitive to the distribution's skewness toward large values, which appear as outliers relative to values expected for a normal distribution. In this situation, a large number of samples must be collected to observe and define the probability of upper tail exposures, *i.e.*, the infrequent high exposure conditions, which can have a strong effect on the estimates of the arithmetic mean (Rappaport and Selvin 1987), and the asymmetric shape of the distribution makes the arithmetic standard deviation a

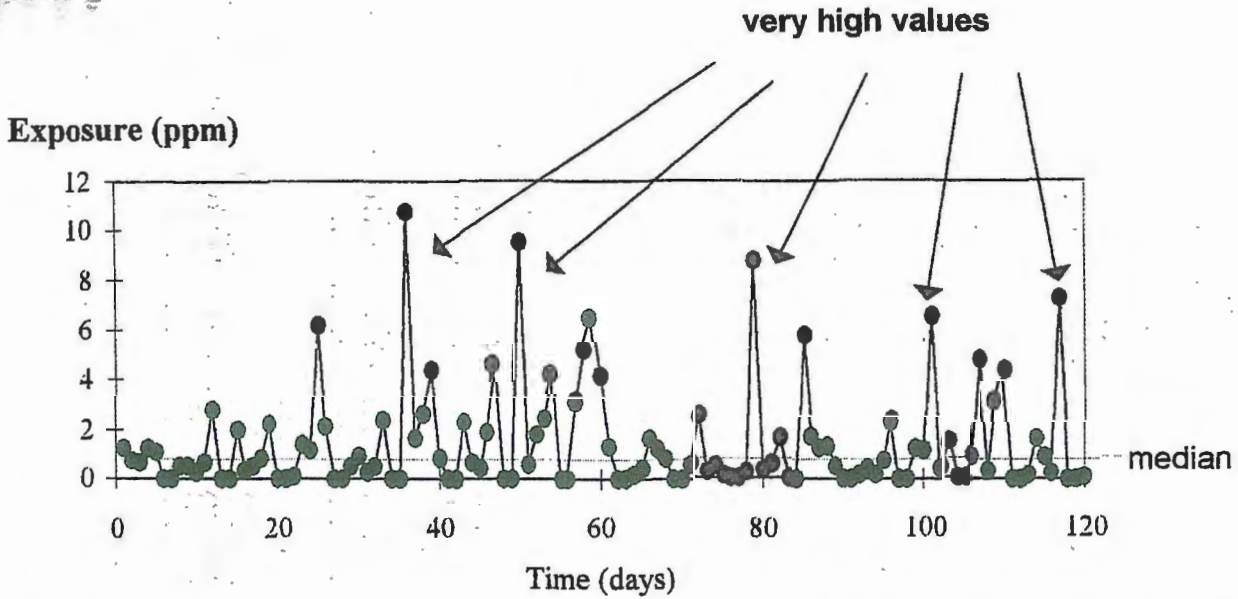


Figure 3. Exposure variability simulated for petrochemical worker based on observed distribution of exposures for butadiene. This assumes a geometric mean exposure of 1.0 ppm (an estimate of the median) and a geometric standard deviation of 3.0. The occurrence of very high exposures is noted.

poor estimator of the probability of the high values. It is critical to recognize that the full distribution needs to be characterized, to obtain precise estimates of the arithmetic mean for calculating the cumulative dose. The frequency and intensity of the highest exposures may be more important for disease risk than the long-term average because they produce high concentrations at the target tissue and may produce disproportionate responses.

The goal is to describe the distribution of exposures across time for groups of individuals with a given set of exposure determinants, *e.g.*, job title, work area, residence location, *etc.* The determinants can affect both the dose related arithmetic mean and dispersion of the exposures. Because we are interested in the time course of tissue concentrations as the proximal cause of effects, exposures need to be defined for pharmacokinetically relevant time periods, *e.g.*, seconds for direct effects on eyes, minutes for anesthetic neurological effects, or months for fibrogenic pulmonary effects of silica dust. Given the time scale of interest, then a representative exposure time profile can be constructed for each subject, which is relevant to his tissue concentration- time course.

In the next section, pharmacokinetic methodology will be presented that can estimate internal concentrations based on an individual's exposure profile. Where the between person variability is small, then each member of a group with the same set of exposure determinants, will have a similar distribution of exposures. However, even though their exposure distributions are very similar, each person will not have the same internal levels because those are affected by differences in personal characteristics, such as age, sex, and race. Given reasonable estimates of the internal levels, then personal dose metrics can be calculated that are quantitatively relevant to disease risk.

INTERNAL CONCENTRATIONS RESULTING FROM EXPOSURE

The internal concentrations resulting from an exposure depend strongly on the route of entry and the physiologic and metabolic characteristics of the individual. The processes that translate external exposures at the point of entry into internal tissue concentrations are diagramed in Figure 4. Some of these are physico-chemical processes associated with solubilization, diffusion, and absorption, and others are physiological processes: respiration, excretion, metabolism, and blood transport to and from the tissues. The time profile of the active agent (parent compound or metabolite) in the internal target tissues is defined by the pharmacokinetics of the substance (also called toxicokinetics for toxic materials), and these processes can be modeled mathematically.

Estimation of Internal Concentrations with Pharmacokinetic Models

Considerable progress has been made in developing mathematical models that can predict the time course of tissue concentrations of whole animals and humans (Gibaldi and Perrier 1982; Gerlowski and Jain 1983; Bailer and Dankovic 1997). The kinetics of materials can be reasonably described by making a few simplifying assumptions. First, the concentration within most organs and discrete tissues behaves like a well mixed compartment, where the incoming materials are rapidly mixed through out the tissue volume and the distribution of materials throughout

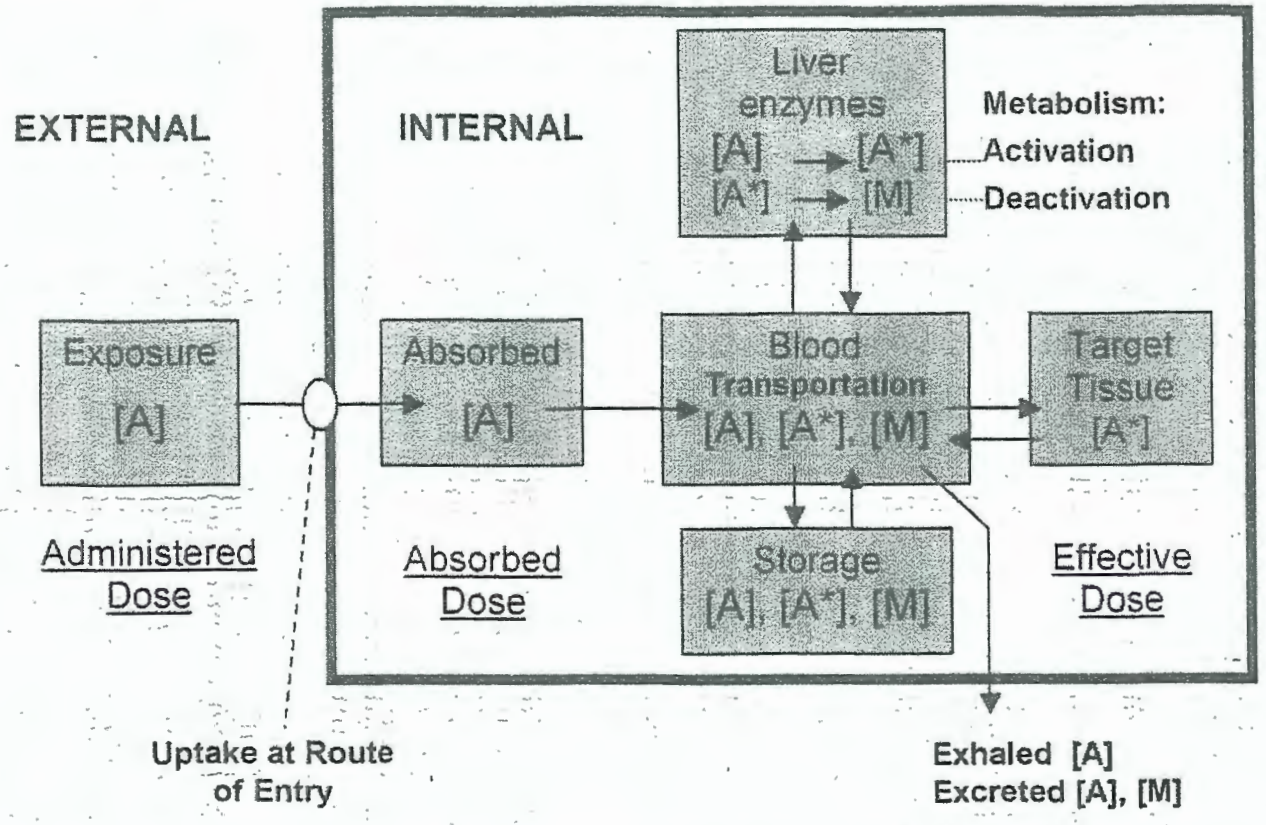


Figure 4. The pharmacokinetic processes that relate exposure to tissue dose: absorption, blood transportation, metabolism, storage, and excretion. Additional processes may also take place for some materials, such as protein binding.

the body is limited by blood flow. Second, each substance has a solubility in the tissues (partition coefficient; $PC_{t/b}$), which is defined by the ratio of concentrations in tissues and blood at steady state, *i.e.*, in venous blood leaving the tissue. Third, the relative blood flow per unit of tissue, its perfusion, is an important kinetic characteristic of the tissue, which determines the rates of uptake and release of the substance by the tissue. Where the $PC_{t/b}$ for a substance is substantially different from 1.0, the perfusion is defined by the ratio of the blood flow divided by the tissue volume times $PC_{t/b}$, where the latter product is the effective volume of tissue for a substance. Fourth, tissues with similar perfusions will have similar uptake and release rates for non-binding, soluble materials. This feature is used to collect tissues into groups with similar kinetic behavior.

For most toxic substances, three or fewer broad tissue groups: well perfused, poorly perfused and body fat, can be usefully identified based on their perfusion time constants. In some cases, it is useful to have the target tissue as a separate compartment, such as the brain for the central nervous system effects. Each of these tissue groups is considered as a "compartment". Each compartment has a blood flow, tissue volume, and a $PC_{t/b}$ for the agent. For most small, non-polar molecules, the uptake and release by the tissues is by passive diffusion, and binding and active transport are not important. If the toxic agent is a metabolite, then additional model components are needed to describe the formation, distribution and removal of the metabolite. Only those physiologic processes relevant to a particular substance and route of entry need to be included in a model. Mathematical models have been developed for clinical applications and for a growing number of environmental exposures (Gibaldi and Perrier 1982; Bailer and Dankovic 1997).

Epidemiologic studies of individuals with genetic polymorphisms in important metabolic enzymes have shown increased risk associated with enzyme variants that alter the amounts of toxic metabolites. For example, individuals who have had long-term exposure to 4-aminobiphenyl and have genetically determined slow acetylation rates, have an increased risk of bladder cancer, because they do not quickly acetylate an intermediate carcinogenic metabolite, N-hydroxy-4-aminobiphenyl (Cartwright *et al.* 1982). Bois and coworkers showed that the normal population variability in three factors: urine pH, N-acetylation rate, and N-hydroxylation rate, could account for a 160-fold difference in bladder cancer risk in a simulated population with a fixed, constant administered dose (Bois *et al.* 1995). Although there is growing awareness of the importance of variation in metabolism among subjects, it is not clear how this can be used to improve traditional dose metrics in epidemiologic studies. However, metabolic information can easily be integrated in PBPK models to estimate individual time courses of tissue levels.

PBPK Model for 1,3-Butadiene

The metabolism, inhalation pharmacokinetics, and tissue dosimetry of butadiene (BD) have been extensively reviewed (Himmelstein *et al.* 1997). A three compartment PBPK model for estimating internal BD levels from an exposure time profile is shown in Figure 5, and the derivation of its parameters is given in Table 1. BD's route of entry is inhalation, and it leaves the body by exhalation and metabolism. In general, metabolism must be considered whenever it affects the formation or

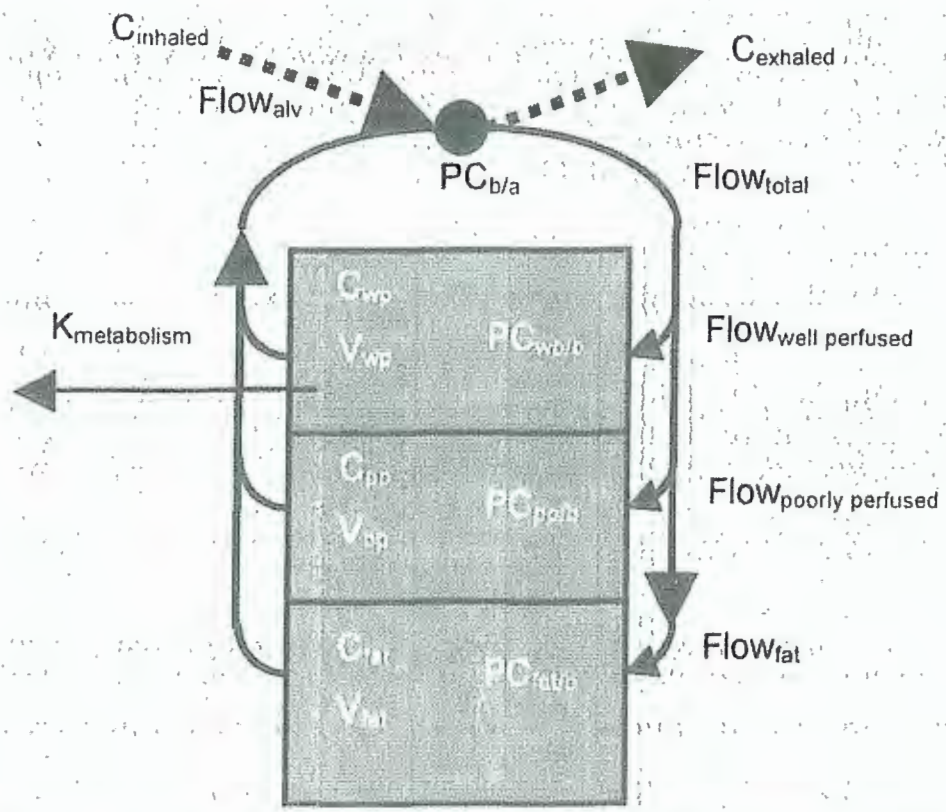


Figure 5. Three compartment physiologically based pharmacokinetic (PBPK) model of an airborne agent absorbed by inhalation. Concentrations, C, in each compartment are estimated by the model based on time course of $C_{inhaled}$ and $C_{exhaled}$

removal of the toxic agent. A simple input/output, mass balance expression was used for the pulmonary exchange of BD, instead of a lung compartment, to simplify the model fitting. BD is oxidized to toxic epoxides (the presumed agents) by P450 enzymes in the liver, kidneys, and lungs (Himmelstein *et al.* 1996), so metabolism was assigned to the well perfused compartment as a whole, instead of only a liver compartment. The target tissues for leukemia are uncertain, but they are likely to be part of the well perfused compartment, which includes the bone marrow and other blood forming organs. BD is highly lipophilic: it has a large partition coefficient for fat ($PC_{fat/blood}$ ~20) and a moderate blood-air coefficient ($PC_{blood/air}$ ~1.3) (Filser *et al.* 1993). As a result, it is important to include a body fat compartment in the model.

While there are a large number of parameters to be estimated for a PBPK model (Table 1), the range of values they can have is limited, and the sums of compartment volumes and blood flows, must equal the body volume (based on weight) and cardiac output (based on alveolar ventilation) respectively. As a result, a PBPK model is highly constrained by pre-existing information about the body and its

Table 1. Parameters for a three compartment PBPK model for 1,3-butadiene.

Parameter	Symbol	Estimation of Value	Parameter Values[a]	Standard Man[b]
Respiratory Activity				
Tidal ventilation	$Flow_{pul}$	(measured for subject)	7.06±1.33 L/min	---
Fraction deadspace	f_{DS}	(algorithm and fitted)	0.48±0.04	---
Alveolar ventilation	$Flow_{alv}$	$Flow_{pul} * (1 - f_{DS})$	3.68±76 L/min	---
Blood Flows (f_i is fraction of flow to tissues)				
Total (cardiac output)	$Flow_{total}$	$Flow_{alv}/1.14$	3.23±0.67 L/min	---
Well Perfused Tissues	$Flow_{wp}$	$Flow_{total} * (1.0 - f_{pp} - f_{fat})$	2.36±0.50 L/min	2.19
Poorly Perfused Tissues	$Flow_{pp}$	$Flow_{total} * f_{pp}$; fitted	0.53±0.18 L/min	0.59
Fat Tissues	$Flow_{fat}$	$Flow_{total} * f_{fat}$; fitted	0.34±0.07 L/min	0.32
Compartment Volumes (total equals body weight, BDW: pi is fraction of volume in tissues)				
Well Perfused Tissues	V_{wp}	BDW * p_{wp} ; fitted	17.4±5.4 L	3.6
Poorly Perfused Tissues	V_{pp}	BDW * $(0.9 - p_{wp} - p_{pp})$	27.6±7.7 L	23.4
Fat Tissues	V_{fat}	BDW * p_{pp} ; fitted	18.4±9.7 L	13.5
Partition Coefficients				
Blood to air	$PC_{b/a}$	(measured for subject)	1.49±0.20	---
Well Perfused Tissues	$PC_{wp/b}$	(taken from literature)	0.8	---
Poorly Perfused Tissues	$PC_{pp/b}$	fitted	0.86± 0.16	---
Fat Tissues	$PC_{fat/b}$	(taken from literature)	20	—

a. The mean _ SD model parameters were estimated by prior measurements for subjects or by fitting as noted above (Mezzetti, *et al.* 2001).

b. The Standard Man values were taken from values published by the ICRP (1975)

physiology. Generally two approaches have been used to obtain parameters: consensus estimates, and fitted values. Consensus model parameters have been developed for the Reference Man and Woman based on a large number of radiological measurements (ICRP 1975; Ellis 1990). The USEPA and researchers have also developed sets of PBPK parameters for risk assessment modeling (Reitz *et al.* 1990). Alternatively, in an approach that is becoming more common, human time course data for breath or blood levels during and after a controlled laboratory exposure can be fitted with one of several biostatistical techniques to estimate the parameters (Bois, Smith *et al.* 1999).

The primary parameters, including metabolic rate, for this model were fitted for a population of 133 test subjects exposed to 2.0 ppm BD for 20 min in a laboratory

experiment using an approved human subjects protocol (Smith *et al.* 2001). Exhaled breath levels of BD during and for 40 min after exposure were measured, as well as the subjects height, weight, blood-air partition coefficient, and other personal characteristics. The goal of this study was to estimate the rate of the first oxidation step in the metabolism of BD: butadiene oxidized to epoxy butene. Even this relatively simple model has six parameters (shown in Table 1), which were estimated by fitting the model using a Bayesian hierarchical approach (individual and population parameters were both fitted simultaneously) with a Monte Carlo-Markov chain modeling method developed by Bois and colleagues (Bois *et al.* 1999). It was not possible to expose humans to sufficient BD to observe the clearance of BD from the fat compartment and directly estimate the effective fat volume. Alternatively, the physiological fat volume was estimated with a regression-based algorithm (Deurenberg *et al.* 1991). The individual fitted PBPK models had a good fit to the observed data. There was considerable person to person variation in the amount of BD retained during this resting exposure (total uptake 1.90±0.82 μg/kg) and the rate of oxidation (0.321±0.076 min^{-1}) for this fixed exposure (Smith *et al.* 2001). Table 1 also compares the fitted parameters to the standard reference parameters used in risk assessment modeling and shows some significant differences in the mean values and larger differences for some individuals within the population tested. Since the population in the BD study was drawn from a cross section of males and females, ages 19 to 60, and four ethnic groups, it is perhaps not surprising.

Thus mathematical tools exist to enable estimation of absorbed dose and even internal concentrations of the parent compound and active metabolites given an exposure time profile, an appropriate PBPK model, and the characteristics of the individual being exposed. The next issue is to determine how to best summarize the tissue-time concentration profile in a way that is quantitatively related to the probability or strength of the response.

RELATING TISSUE CONCENTRATION TO EFFECTS

The type of response and its temporal relationship to the time course of the tissue concentration depends on the mechanism of the effects (Pratt and Taylor 1990). Tissue responses at the cellular level may be immediate and proportional to concentration, such as when some types of cell receptors mediates the response, or there is enzyme inhibition, although the relationship may not be linear and they commonly have a threshold to produce a response. While these responses may be generally immediate in the affected cells, they may not be apparent in the whole organism. In this situation, as long as secondary processes do not intervene, the effect may be reversible and last only as long as the tissue concentration is maintained. Receptor and other initial effects may also cause secondary cellular damage or cell death, and those effects can accumulate during and after the toxic tissue concentration and produce secondary effects. The secondary responses may be proportional to the stimulus or the amount of damage, such as simple fibrogenesis from quartz dust, while some are not, such as immunological responses. Repair processes are also important in determining the magnitude of effects and the time course of the overall effects. Some types of effects are stochastic, such as cancer, which may have probability distributions that are linear or nonlinear with the tissue

concentration depending on the agent. Diseases are often the result of a complex cascade of early, intermediate, and late effects. Each of these effects will have a time course that reflects the underlying pathogenic processes. While it is beyond the scope of this brief discussion, the methods used to measure the effect are also important, as are their selectivity and sensitivity. The same effect evaluated by different methods may appear to have different characteristics, such as how rapidly it appears after exposure is begun.

Despite the wide range of possibilities for mechanistic complexity, the observed time courses of many effects can be described with simple processes operating at a cellular level, or a combination of cellular and organ-level processes (Pratt and Taylor 1990). These process descriptions can be used to model the time course of the effect. This approach of fitting the time course of an effect is analogous to PBPK modeling where relatively simple process models can describe the key features of the physiologic and metabolic processes without including all of the subtleties of an exact model of the physiologic and pharmacologic processes. Many effects can be placed in one of the following broad descriptive categories, sometimes also including a lag or a threshold for the effect: (1) damage and repair processes; (2) stochastic processes; and (3) progressive responses independent of tissue concentration. Some disease effects are combinations of processes, which can be represented by a combination of these process models, such as chemical carcinogenesis, which has been described as a combination of DNA damage and repair processes plus one or more stochastic mutation steps. Some general examples and the relevant dose metrics are discussed below.

Even though the concentration of a toxic agent may be approximately uniform through out a tissue, the cells do not respond exactly the same; there is a distribution of responses. For example, even though the tissue concentration is high enough to kill the cells, they die off in an exponential decay curve. Cellular damage and repair have been successfully represented by a simple process model: changes in a population of cells in an organ or tissue is defined by the balance between the rate of damage and the rate of repair, as shown below.

Change in cell population = (rate cells are repaired) – (rate cells are damaged)

The rate of damage is negative because it reduces the population number. The damage rate depends on the concentration of the toxic material in the target tissue, if the concentration is high enough. The rate of repair is positive and proportional to the number of cells damaged. If cells are killed and there is insufficient time for replacement, then the rate of change in the number of cells during exposure to the agent, equals the initial cell number multiplied by the tissue concentration times a rate constant, K, which is the fraction of cells killed per unit time per unit concentration. This process relationship leads to an exponential decline in cell count with a rate that is proportional to the concentration increment above the threshold, Th.

$$N_{\text{cells after } t} = \int_0^t N_0 \, K \, (C\text{-}Th) \, dt = N_0 \, e^{-K(C\text{-}Th)t} \text{ for } C > Th$$

When longer time intervals are considered, then repair or replacement processes can be considered as well.

This type of model was used by Reitz and later by Smith to model nongenotoxic liver tumor effects of chloroform (Reitz *et al.* 1990; Smith *et al.* 1995). The model was able to resolve apparently conflicting findings of several sets of animal bioassays that had different exposure conditions. They found that dose measures that modeled tissue damage processes, which were sensitive to dose rate, performed better than those based on total administered quantities of chloroform scaled to body size. It may be noted that damage in real life can be more complicated that this. For example, injury from an acute exposure to chlorine gas, is produced not just by killing cells in lung tissues but also by damaging cells to varying degrees, which produces a mixed set of secondary effects that will lag beyond the exposure, such as edema and inflamation. The best dose metric will depend on which outcome is being modeled, an immediate effect of cell killing or a secondary effect. Also, depending on the mechanism of cell mortality, repeated exposures may not produce the same effects because the remaining cells may not be as sensitive to the agent. This is well known for repeated applications of chemotherapy in cancer treatment.

The damage-repair model implies that when a fixed tissue concentration is maintained, after a period of time, a steady state balance between damage and repair may be reached and continuing a tissue concentration (or an exposure) for a longer period of time will not change the observed level of total damage. As a result, the level of damage appears to only depend on the tissue concentration, independent of duration beyond some minimum. An example of this is seen in cigarette smoking, where after an initial period, smokers' symptoms of cough and phlegm depend only on their current rate of smoking. Since it is rare that environmental exposure is actually constant, variations in the time course of exposure, and subsequent tissue concentrations, will affect the apparent relationship between average exposure and response. If there was a lag in the development of the effect, then the dose metric would need to be estimated from tissue concentrations offset by the lag time. Also if the repair process is affected by on- going exposure, *e.g.*, it is inhibited, then the dose metric will need to be appropriately modified. If repair is very slow, or overwhelmed by a high rate of new damage, then the effects will be cumulative, and the relevant dose metric reduces to the simple cumulative type.

For tissue effects where there is no repair and damage is permanent, the tissue dose is related to the cumulative exposure index. A good example is pulmonary fibrosis from dust exposures. Dust is deposited in the alveolar area, where it causes a fibrogenic response. This may be modeled by assuming that each dust particle becomes irreversibly encapsulated in a small amount of fibrotic material, or each mg of dust deposited in the lungs produces K mg of fibrosis after a period of time. The quantity of dust deposited in the lungs, the dose of dust, is the product of average airborne exposure concentration multiplied by the inhalation rate times the fraction deposited multiplied by duration of exposure. Assuming that everyone exposed has approximately the same inhalation rate and fraction deposited, then the average airborne exposure concentration multiplied by duration of exposure is the cumulative exposure dose metric (Smith 1992).

A good example of the stochastic type of process model based on cell turnover kinetics is the two stage cancer model developed by Moolgavkar and his associates (Moolgavkar 1986). They and others have successfully applied that model to a variety

of carcinogenic processes from radioactive materials to chemical carcinogens (Leroux *et al.* 1996; Luebeck and Moolgavkar 1996; Moolgavkar 2000). Even though the model is very simplistic relative to the complex multistep cancer processes, the model gives a good description of the time course and many of the important factors affecting risk.

Each effect that has a different type of time course will have a different form of dose metric. By observing the temporal behavior of a response, it is possible to identify the appropriate category. Given information on variation in the intensity of an effect after beginning and stopping exposure, the repair rate and lag time can be estimated respectively. In some cases, it may be necessary to add a threshold, or combine models to fully describe the observed behavior. Based on these, then an appropriate dose metric can be formulated and its fit to time course data can be tested. A poor fit by a metric implies that the model is not appropriate for the effect.

Even these simple descriptive categorizations can have important implications. In some cases, an hypothesized mechanism cannot fit the observed time course, which can eliminate some otherwise reasonable possibilities, and help guide the search for the underlying mechanism of effects. As mechanistic information develops about the target site, the nature of the cellular response and repair processes, then the dose metric can be modified to better represent the processes. Thus, the appropriate summary dose metric for the target site is a function of the type of response, and there is no single representation that is appropriate for all types of effects.

PROBLEMS WITH CLASSICAL DOSE DEFINITIONS IN EPIDEMIOLOGY

There are several limitations for the application of classical toxicologic dose concepts to human environmental exposures studied by epidemiologic methods, which can be solved by the dose metrics developed from time course data on effects.

- First, human exposures often do not have a clear starting and stopping time point, whereas laboratory exposure studies and clinical dosing do.

There is nearly always some background exposure to general environmental contaminants, and many occupational ones, so "exposure" is more a matter of intensity than presence or absence. For example, particulate matter less than 2.5 μm in diameter ($PM_{2.5}$) is of concern for cardiac and respiratory mortality effects (Schwartz and Neas 2000). However, exposure to $PM_{2.5}$ is nearly universal: outdoors, within homes, and in occupational settings. The airborne concentration of $PM_{2.5}$ varies widely across those settings. The question is: since all periods of the exposure may not contribute to the observed effect, what is the relevant time period for the outcome? If there is a threshold tissue level for the initial effects, or repair processes can handle effects up to some level, then periods of exposure producing tissue levels below those levels are irrelevant to the effect and including them in an epidemiologic dose metric will obscure the dose-response relationship by contributing misclassification to the dose.

- Second, many common human exposures do not have a fixed composition, and the specific agent of effects is uncertain.

Aggregation and averaging daily exposures may obscure differences in effects associated with differences in composition. For example, testing human alveolar macrophages with daily samples of ambient $PM_{2.5}$ showed variable cytokine responses (stimulation of inflammatory responses) per μg across daily samples (Imrich *et al.* 1999). Variations in composition may have a greater effect than variation in exposure intensity. Crude markers of total exposure intensity, such as total $PM_{2.5}$, can hide wide variation in some trace constituents, such as polycyclic aromatic hydrocarbons, with little effect on the total mass. The question is: what are the agents and how might they vary within the total composition and possibly interact? It is better to relate specific agents to effects, than use broad markers of exposure, because they represent stronger tests of hypotheses. Hazard identification studies can reasonably be done with broad indicators of exposure like occupation, but once the hazard seems apparent then the search for agents should be increasingly focused.

- Third, human exposures are uneven across time, which leads to the possibility of dose rate effects.

As shown in the exposure section, it is not unusual that minute-to-minute airborne exposures can vary by as much as an order of magnitude or more. This is not consistent with the laboratory dose paradigm, where the exposure is held as uniform as possible. It raise the concern that human effects may differ across individuals because they experienced different dose rates, but may have the same average exposure. Epidemiologic studies usually report average exposures for groups of subjects. The question is: are there dose rate effects occurring within averaging intervals for dose metrics? The possibility of different dose rates is rarely considered.

- Fourth, the distribution of exposure intensities across an "exposed" population is usually poorly described, usually only by the group mean and SD of a small data set.

In a typical epidemiologic study, groups of subjects are chosen based on some broad characteristic, *e.g.*, their job titles or residence locations, which have assumed or demonstrated differences in average exposures based on limited sampling. It is uncommon that the full distribution of exposure variability across and within exposure classifiers (*e.g.*, job titles) has been described: differences in mean exposure across time for individuals in the same group. Rappaport and Kromhout have noted that for groups chosen by job title alone the large differences among individuals within a group may make apparent differences between job groups are meaningless, even though the overall mean exposures may be statistically different (Kromhout *et al.* 1993; Rappaport *et al.* 1993). With this type of mis-classification, it is unlikely that risk will differ across exposure groups. If differences in risk are observed, it may be that the response is nonlinear and the differences are caused by the few individuals with the highest exposures. Thus the affected individuals are not a random sample from a homogeneously exposed population, rather they are likely to be the subjects whose exposures were much higher than the average. This is not a problem for epidemiologic studies designed for risk detection, unless the uneven nature of

risk across exposure groups is taken as evidence of a lack of dose- response. Thus it is very important to fully characterize the distribution of subjects in exposure groups when the goal is evaluation of dose-response for risk quantification.

These four limitations make it difficult to directly transfer the dose concepts of classical toxicology to epidemiologic studies of toxic material exposures. Human exposures are much more complex than lab studies, and the data analysis strategy needs to deal with the complexity. The alternative approach using dose metrics based on PBPK models and the time course of effects are suitable under all of the limiting conditions named above, when detailed exposure data are available. It is proposed that a PBPK model can be used to estimate individual tissue doses. In some cases this can be done even where repeated measurements on individuals are not available, if it has been previously determined that between individual variation is small relative to the group mean, or the differences among the group mean exposure are very large (Rappaport *et al.* 1993). In addition to accounting for the time course effects of varying exposure, dose metrics based on PBPK models can account for inter-individual differences in physiology and metabolism.

EPIDEMIOLOGIC STUDIES OF HUMAN DOSE-RESPONSE

Epidemiologic determination of quantitative dose-response relationships has not been attempted in many situations. It is the premise of this paper that a special type of epidemiologic study is needed to develop quantitative dose-response data for risk assessment. It is unlikely that this PBPK model based approach can be done as an add-on to a traditional hazard identification design, although some of these studies could be improved by consideration of dose issues. The dose-based studies are difficult, time consuming, labor intensive, and expensive. However, there is a critical need to have human data for common exposures with substantial effects, such as lung cancer from diesel exhaust exposure, which makes this extra effort worthwhile.

Figure 6 shows the structure of a study using PBPK modeling to determine the human dose-response relationship in an epidemiologic context. Comparing this design to the study design for a hazard identification study (Figure 2), the main difference is that there is detailed consideration of physiologic and pharmacokinetic processes by which exposure leads to a tissue dose, and there is a formal estimation process for the dose metric. This study approach can be enhanced by inclusion of biomarkers for exposure and effects that might be used to validate the dose estimates and account for between subject variability in pharmacokinetic and susceptibility factors. One of the strengths of this approach is that it uses data on each subject to personalize the dose metric, such as height, weight, age, sex, and race, which also have implications for dose, when the agent is metabolically activated and/or deactivated. Many of these characteristics are also considered in the classical epidemiologic analysis, but there is an analytical limit to how many can be considered and the structure of those relationships.

The application of this conceptual approach does not require highly detailed information on the mechanism, nor a fully detailed PBPK model. If the tissue target and time course of effect can be hypothesized, then meaningful dose metrics can be formulated. It is not necessary that the mechanism be fully elucidated down to the

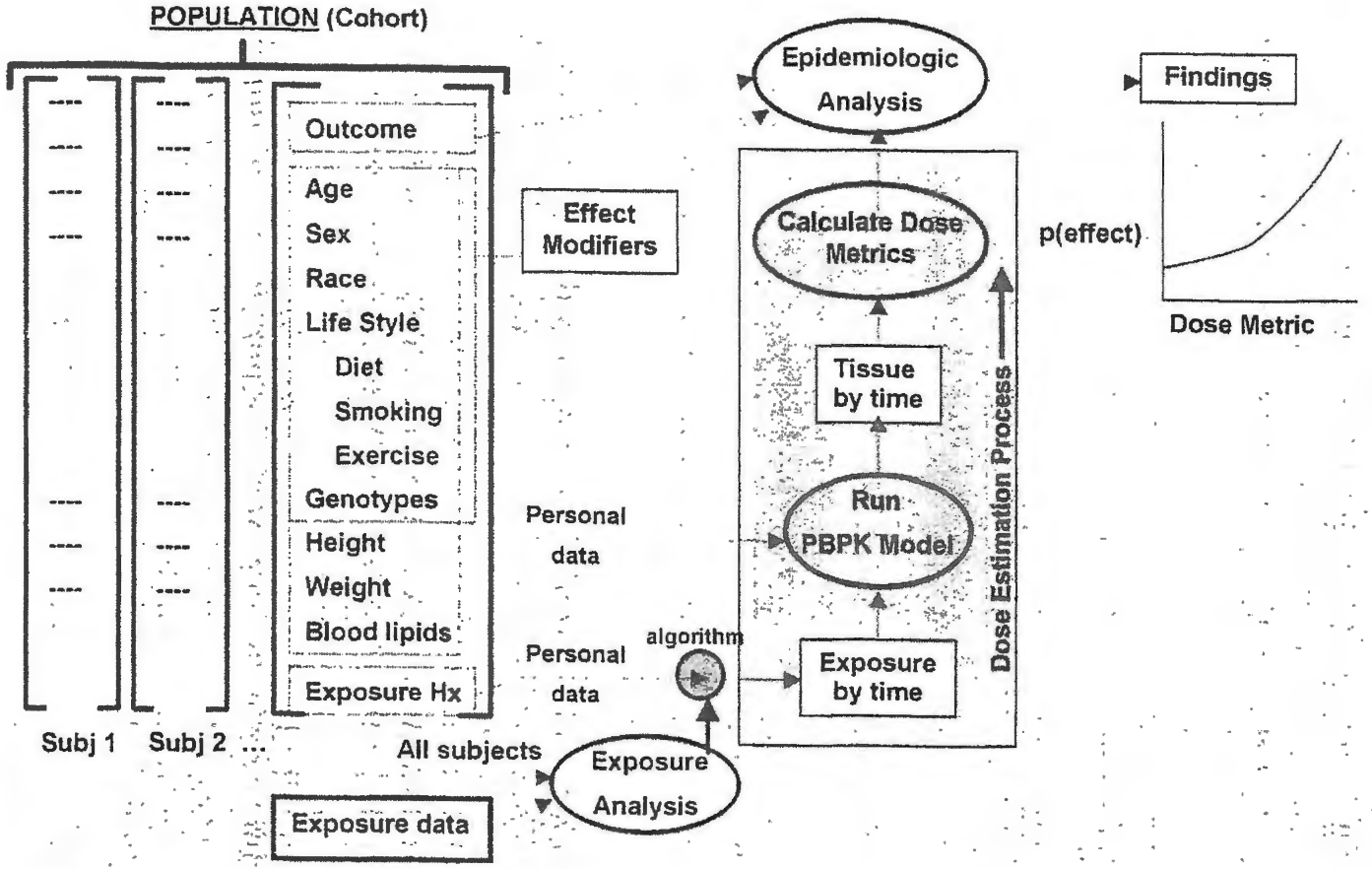


Figure 6. Dose-response analysis using a PBPK model to estimate each subject's dose metric. The ovals are data processes, and the rectangles are data or results. Exposure analysis provides the algorithm to make the exposure profile across time.

molecular level. Descriptive data can be very informative too. It is expected that as mechanistic data are developed then the model of temporal processes can be refined.

It is a premise of this paper that the PBPK dose modeling approach increases the relevant information content of dose metrics over empirical metrics that are based on exposure alone, so unless the statistical noise is very large, some improvement may be expected. How much improvement remains to be determined. This dose modeling approach is best suited for prospective, repeated measures studies, where each subject's personal exposure profile is well defined and there are repeated measures of early indicators of adverse effects. Ultimately this will permit better estimation of risk reductions needed to meet policy goals.

FUTURE RESEARCH NEEDS — IMPROVING THE LINKAGE OF EPIDEMIOLOGY, TOXICOLOGY, AND RISK ASSESSMENT

Better coordination of epidemiologic and toxicologic research is needed, so that risk assessment and risk projection can be validated (Andersen *et al.* 1992). A parallel research strategy needs to be developed where laboratory bioassays of early effects are calibrated against epidemiologic findings with good dosimetry. Mechanistic research is advancing quickly to describe the processes by which toxic materials cause human effects. Time course studies of the development and recovery from adverse effects is another critical need to better structure the models of processes leading to effects and disease. As these are elucidated, then field studies of exposed populations can be conducted to verify their predictive value.

The lack of human validation of estimates of tissue concentration from PBPK models has limited their use. Human volunteer studies are needed to better understand the relationship between exposure and internal levels. There are currently a number of biomarkers, such as hemoglobin and DNA adducts, that can be used to provide estimates of the concentrations of activated metabolites in the blood and tissues, respectively, but the validation of these techniques has been very limited. In some cases, parent and activated materials can be measured directly in blood samples. There is a strong need for more human studies of exposure-dose relationships for key materials, which may be done in the laboratory or some field exposure situations, such as occupational settings.

The wide range of metabolic differences across individuals is likely to be one of the important sources of variation in risks among exposed populations given apparently equivalent exposures. Methods to easily determine an individual's activation/detoxification rates for metabolically activated and/or detoxified agents is strongly needed.

Finally, easy methods to identify active agents for humans effects, such as testing in tissue cultures or test animals with human genes, are strongly needed. Related to this, better data and methods to determine the time course of human diseases are needed that can guide the development and choice of dose metrics. This research needs to be a part of studies of disease mechanisms.

ACKNOWLEDGMENT

The author gratefully acknowledges the thoughtful suggestions and insights contributed by Drs. David Kriebel and Leslie Stayner. This work was partially funded by NIEHS grants No. ES07586 and ES00002.

REFERENCES

Andersen ME, Krishnan K, Conolly RB, *et al.* 1992. Biologically based modeling in toxicology research. Arch Toxicol Suppl 15:217-27

Armstrong BK, White E, and Saracci R. 1992. Principles of Exposure Measurement in Epidemiology. Oxford University Press, NY, NY, USA

Bailer AJ and Dankovic DA. 1997. An introduction to the use of physiologically based pharmacokinetic models in risk assessment. Stat Method Med Res 6:341-58.

Bhatia R, Lopipero P, and Smith AH. 1998. Diesel exhaust exposure and lung cancer. Epidemiology 9:84-91

Bois FY, Krowech G, and Zeise L. 1995. Modeling human interindividual variability in metabolism and risk: The example of 4-aminobiphenyl. Risk Anal 15:205-13

Bois FY, Smith TJ, and Gelman A. 1999. Optimal design for a study of butadiene toxicokinetics in humans. Toxicol Sci 49:213-24

Cartwright RA, Glashan RW, Rogers HJ, *et al.* 1982. Role of N-acetyltransferase phenotypes in bladder carcinogenesis: a pharmacogenetic epidemiological approach to bladder cancer. Lancet 2(8303):842-5

Checkoway H, Pearce N, and Crawford-Brown DJ. 1989. Research Methods in Occupational Epidemiology. Oxford University Press, NY, NY, USA

Deurenberg P, Weststrate JA, and Seidell JC. 1991. Body mass index as a measure of body fatness: age- and sex-specific prediction formulas. Brit J Nutrition 65:105-41.

Ellis KT. 1990. Reference man and woman more fully characterized. Variations on the basis of body size, age, sex, and race. Biol Trace Elem Res 26-27:385-400

Fajen JM, Roberts DR, Ungers LJ, *et al.* 1990. Occupational exposure of workers to 1,3-butadiene. Environ Health Perspect 86:11-8

Filser JG, Johanson G, Kessler W, *et al.* 1993. A pharmacokinetic model to describe toxicokinetic interactions between 1,3-butadiene and styrene in rats: predictions for human exposure. In: Sorsa M, Peltonen K, Vainio H, *et al.* (eds), Butadiene and Styrene: Assessment of Health Hazards. IARC Scientific Publication No 127:65-78. International Agency for Research on Cancer, Lyon, France

Fraser MP, Kleeman MJ, Schauer JJ, *et al.* 2000. Modeling the atmospheric concentrations of individual gas-phase and particle-phase organic compounds. Environ Sci Technol 34:1302-12.

Gerlowski LE and Jain RK. 1983. Physiologically based pharmacokinetic modeling: principles and applications. J Pharmacological Science 72:103-27

Gibaldi M and Perrier D. 1982. Pharmacokinetics. Marcel Dekker, Inc., NY, NY, USA

HEI (Health Effects Institute). 1999. Diesel Emissions and Lung Cancer: Epidemiology and Quantitative Risk Assessment. A Special Report of the Institute's Diesel Epidemiology Expert Panel, pp 1-72. Cambridge, MA, USA

HEI (Health Effects Institute). 1995. Diesel Exhaust: A Critical Analysis of Emissions, Exposures, and Health Effects. Cambridge, MA, USA

Himmelstein MW, Acquavella JF, Recio L, *et al.* 1997. Toxicology and epidemiology of 1,3 butadiene. Crit Rev Toxicol 27:1–108

ICRP (International Commission on Radiological Protection). 1975. ICRP Publication 23: Reference Man. Pergamon Press, Oxford, UK

Imrich A, Ning YY, Koziel H, *et al.* 1999. Lipopolysaccharide priming amplifies lung macrophage tumor necrosis factor production in response to air particles. Toxicol Appl Pharmacol 159:117-24

Klaassen CD, Amdur MO, and Doull J. 1986. Casarett and Doull's Toxicology. Macmillan Publishing Co., NY, NY, USA

Kleeman MJ, Schauer JJ, and Cass GR. 2000. Size and composition distribution of fine particulate matter emitted from motor vehicles. Environ Sci Technol 34:1132-42

Kriebel D. 1994. The dosimetric model in occupational and environmental epidemiology. Occupational Hygiene 1:55-68

Kromhout H, Symanski E, and Rappaport SM. 1993. A comprehensive evaluation of within- and between-worker components of occupational exposure to chemical agents. Ann Occup Hyg 37:253-70

Leroux BG, Leisenring WM, Moolgavkar SH, *et al.* 1996. A biologically-based dose-response model for developmental toxicology. Risk Anal 16:449-58

Lipsett M and Campleman S. 1999. Occupational exposure to diesel exhaust and lung cancer: a meta-analysis. Am J Public Health 89:1009-17

Luebeck EG and Moolgavkar SH. 1996. Biologically based cancer modeling. Drug Chem Toxicol 19:221-43

Moolgavkar SH. 1986. Carcinogenesis modeling: from molecular biology to epidemiology. Annual Reviews in Public Health 7:151-69

Moolgavkar SH. 2000. Multistage models and the A-bomb survivor data: implications for carcinogenic mechanisms? Radiat Res 154:728-9;discussion 730-1.

NAS (National Academy of Sciences). 1983. Risk Assessment in the Federal Government: Managing the Process., National Academy Press, Washington, DC, USA

NAS (National Academy of Sciences). 1994. Science and Judgement in Risk Assessment. National Academy Press, Washington, DC, USA

Pratt WB and Taylor P. 1990. Principles of Drug Action. Churchill Livingstone, Philadelphia, PA, USA

Rappaport SM. 1991. Exposure assessment strategies. In: Rappaport SM and Smith TJ (eds), Exposure Assessment for Epidemiology and Hazard Control, pp 219-49. Lewis Publishers, Chelsea, MI, USA

Rappaport SM and Selvin S. 1987. A method for evaluating the mean exposure from a lognormal distribution. Am Ind Hyg Assoc J 48:374-9

Rappaport SM, Kromhout H, and Symanski E. 1993. Variation of exposure between workers in homogeneous exposure groups. Am Ind Hyg Assoc J 54:654-62

Reitz RH, Mendrala AL, Corley RA, *et al.* 1990. Estimating the risk of liver cancer associated with human exposures to chloroform using physiologically based pharmacokinetic modeling [published erratum appears in Toxicol Appl Pharmacol 1991 Sep 15;110(3):536]. Toxicol Appl Pharmacol 105:443-59

Schauer JJ and Cass GR. 2000. Source apportionment of wintertime gas-phase and particle-phase air pollutants using organic compounds as tracers. Environ Sci Technol 34:1821-32

Schauer JJ, Rogge W, Hildemann M, *et al.* 1996. Source apportionment of airborne particulate matter using organic compounds as tracers. Atmos Environ 30:3837-3855

Schwartz J and Neas LM. 2000. Fine particles are more strongly associated than coarse particles with acute respiratory health effects in schoolchildren [see comments]. Epidemiology 11:6-10.

Seixas NS, Robins TG, and Becker M. 1993. A novel approach to the characterization of cumulative exposure for the study of chronic occupational disease. Am J Epidemiology 137:463-71

Shi JP, Mark D, and Harrison RM. 2000. Characterization of Particles from a current technology heavy-duty diesel engine. Environ Sci Technol 34:748-56

Smith AE, Gray GM, and Evans JS. 1995. The ability of predicted interval dose measures to reconcile tumor bioassay data for chloroform. Regul Toxicol Pharmacol 21:339-51

Smith TJ. 1992. Occupational exposure and dose over time: limitations of cumulative exposure. Am J Ind Med 21:35-51

Smith TJ, Lin Y-S, Mezzetti M, *et al.* 2001. Genetic and dietary factors affecting human metabolism of 1,3-butadiene. Chemical and Biological Interactions 135–136:407-428

Sorsa M, Osterman-Golkar S, Peltonen K, *et al.* 1996. Assessment of exposure to butadiene in the process industry. Toxicology 113:77-83

Turnbull D, Rodricks JV, and Brett SM. 1990. Assessment of the potential risk to workers from exposure to 1,3- butadiene. Environ Health Perspect 86:159-71

Verma DK, Shaw L, Julian J, *et al.* 1999. A comparison of sampling and analytical methods for assessing occupational exposure to diesel exhaust in a railroad work environment. Appl Occup Environ Hyg 14:701-14

Wang WG, Lyons DW, Clark NN, *et al.* 2000. Emissions from nine heavy trucks fueled by diesel and biodiesel blend without engine modification. Environ Sci Technol 34:933-9

Ward EM, Fajen JM, Ruder AM, *et al.* 1995. Mortality study of workers in 1,3-butadiene production units identified from a chemical workers cohort. Environ Health Perspect 103:598-603

Wongphatarakul V, Friedlander SK, and Pinto JP. 1998. A comparative study of PM2.5 Ambient aerosol chemical databases. Environ Sci Technol 32:3926-34

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1295-1301(2002)

Molecular Biomarkers and Epidemiologic Risk Assessment

Paul W. Brandt-Rauf,[1*] Jiin-Chyuan Luo,[2] Tsun-Jen Cheng,[3] Chung-Li Du,[3] Jung-Der Wang,[3] Ramon Rosal,[1] Tamara Do,[1] and Marie-Jeanne Marion[4]
[1]Department of Environmental Health Sciences, The Mailman School of Public Health of Columbia University, 60 Haven Avenue, New York, NY 10032, USA (author for correspondence); Tel(voice):212-305-3959, Tel(fax):212-305-4012; pwb1@columbia.edu. [2]Department of Public Health, Chang Gung Medical College, Taipei, Taiwan. [3]Graduate Institute of Occupational Medicine and Industrial Hygiene, National Taiwan University, Taipei, Taiwan. [4]Unite de Recherche sur les Virus des Hepatites et Pathologies Associees, INSERM, 69424 Lyon, France

ABSTRACT

The use of molecular biomarkers in epidemiologic studies has been advanced as a way to improve risk assessments for occupational and environmental exposures to toxic agents. We have used the detection of two cancer-related, molecular biomarkers of vinyl chloride exposure (mutant *ras*-p21 and mutant p53) to examine workers with equivalent cumulative exposures that would be above or below the current permissible workplace exposure limit for vinyl chloride for differences in the presence of these biomarkers. Workers with cumulative exposures above the current permissible exposure limit (equivalent of >40 ppm-years) have a statistically significantly increased occurrence of both biomarkers in comparison to unexposed controls ($p < 10^{-3}$). Although workers with cumulative exposures of < 10 ppm-years, i.e., well below the current limit, do not have a statistically significantly increased occurrence of these biomarkers ($p > 0.05$), workers with cumulative exposures of 10 to 40 ppm-years, i.e., still below the current limit, are found to have a statistically significant increase ($p < 0.05$). This suggests that the current exposure limit may not be adequately protective and illustrates the potential utility of molecular biomarkers in the refinement of risk assessments for toxic exposures.

Key Words: vinyl chloride, cancer, mutations, exposure limit.

* Corresponding author.

1080-7039/02/$.50

INTRODUCTION

In recent years there has been an explosion of knowledge concerning the molecular pathways by which occupational and environmental toxins produce adverse health effects. This has led to the development of molecular biomarkers that can be used to identify various steps in these pathways *in vivo* in exposed human populations. The use of such biomarkers has been suggested as a way to refine and enhance group and individual risk assessments for toxic occupational and environmental exposures (Hattis and Silver 1993). Rather than relying on traditional estimates of ambient exposures, categorization by clinical disease diagnoses, and stratification by crude external population risk characteristics, the use of molecular biomarkers of dose, effect, and susceptibility, respectively, should be able to provide more precise and mechanistically realistic metrics for the risk assessment process. For example, the use of molecular biomarkers of biologically effective dose (DNA and hemoglobin adducts) have been applied to the risk assessment of genotoxic alkylating agents, such as ethylene oxide (Tornqvist and Landin 1995). More recently, we have attempted to use molecular biomarkers of response (mutant oncoproteins) to examine the risk assessment for workplace exposure to vinyl chloride (VC) (Brandt-Rauf *et al.* 2000). VC provides a particularly good example because considerable detail is available concerning its potential mechanism of action.

VC is a known carcinogen that is rapidly absorbed following respiratory exposure, and it is subsequently metabolized in the liver by the cytochrome P450 2E1 system (ATSDR 1997). The resultant electrophilic metabolites, chloroethylene oxide and chloroacetaldehyde, can form a variety of DNA adducts that are known to be promutagenic, including 7-(2'-oxoethyl)guanine, 1,N^6- ethenoadenine, 3,N^4-ethenocytosine and N^2,3-ethenoguanine (Barbin and Bartsch 1986). Although the oxoethyl adduct accounts for the vast majority of adducts formed, it is rapidly repaired and probably does not contribute to the carcinogenic effects of VC. On the other hand, the less common etheno adducts are poorly repaired and have long half-lives accounting for the production of the specific point mutations identified in VC-related malignancies (Swenberg *et al.* 1992). For instance, 83% of angiosarcomas of the liver (ASL) from VC-exposed workers have been found to contain G→A transitions in the Ki-*ras* oncogene that could be attributed to the generation of ethenoguanine adducts by VC (Marion *et al.* 1991). Similarly, 60% of ASLs from VC-exposed workers have been found to contain A→T transversions in the p53 tumor suppressor gene that could be attributed to the generation of ethenoadenine adducts by VC (Hollstein *et al.* 1994). The occurrence of each of these cancer-related mutations leads to the production of mutant oncoprotein biomarkers (mutant *ras*-p21 protein and mutant p53 proteins and/or auto-antibodies to mutant p53 proteins) that can be detected in the blood of individuals who have tumors that contain the respective mutations (DeVivo *et al.* 1994; Brandt-Rauf *et al.* 1996). Furthermore, these same oncoprotein biomarkers have been identified in several cohorts of workers around the world exposed to VC but without known malignant disease, and they have been found to occur with a significant dose-response relationship with regard to the workers' estimated, cumulative VC exposure at levels above the equivalent current permissible exposure limit used in most western countries of

1 ppm (*i.e.*, cumulative exposures of 40 ppm-years, or 1 ppm for 40 working years) (Smith *et al.* 1998; Li *et al.* 1998; Luo *et al.* 1998, 1999). However, in each of these cohorts the presence of these oncoprotein biomarkers has also been detected in workers with equivalent cumulative exposures that would be below the current permissible exposure limit (*i.e.*, below 40 ppm-years). This has recently prompted us to combine this biomarker data from the different cohorts to examine the prevalence of the biomarkers in various sub-groups of workers with equivalent cumulative exposures below the current permissible exposure limit (Brandt-Rauf *et al.* 2000). This allows us to reevaluate the validity of the risk assessment that forms the basis for the current exposure limit, assuming that the presence of these biomarkers represents a significant cancer risk. The specific issue to be addressed was whether there exist subgroups of workers with equivalent cumulative exposures below the current limit that have statistically significant increases in the prevalence of these VC-induced, cancer-related biomarkers.

METHODS

Two previously described cohorts of VC workers and matched, unexposed controls, one from France and one from Taiwan, were included in the analysis (Smith *et al.* 1998; Luo *et al.* 1998). The combined cohort was comprised of 468 exposed workers, 306 with cumulative exposure > 40 ppm- years and 109 with cumulative exposure ≤ 40 ppm-years, and 155 matched, unexposed controls. Blood samples had been collected from these individuals by routine venipuncture techniques and stored frozen until the time of analysis for the mutant *ras*-p21 and mutant p53 oncoprotein biomarkers. The biomarkers were analyzed as described in detail previously, relying on monoclonal antibodies that are specific for the detection of the mutant oncoproteins (DeVivo *et al.* 1994; Smith *et al.* 1998).

Logistic regression analysis was performed using the unexposed controls as reference with an assigned odds ratio of one, and odds ratios were calculated for various subgroups of exposed workers for the presence of one or both oncoprotein biomarkers with adjustment for previously identified potential confounders including age, smoking and alcohol consumption. Initially, the analysis was performed with the workers stratified into those above and those below the equivalent current permissible exposure limit of 40 ppm-years. Subsequently, the analysis was performed with the workers with equivalent cumulative exposures below the current permissible exposure limit being further stratified into those with 0-10 ppm-years, 10 to 20 ppm-years, 20 to 30 ppm-years, and 30 to 40 ppm- years. Finally, based on the results of the sub-stratification of the workers with equivalent cumulative exposures below the current permissible exposure limit, the workers were grouped into those with < 10 ppm-years of cumulative exposure and those with 10 to 40 ppm-years of cumulative exposures, and these were compared to the workers with cumulative exposures > 40 ppm-years and to the unexposed controls.

RESULTS

Only 16 of the 155 unexposed controls (10%) were seropositive for one or the other of the oncoprotein biomarkers, and no unexposed control (0%) was positive

for both oncoprotein biomarkers. Of the 306 VC workers with cumulative exposures > 40 ppm-years, 122 (40%) were seropositive for one or the other oncoprotein biomarker (adjusted odds ratio = 5.0, $p < 10^{-3}$) and 33 (11%) were seropositive for both oncoprotein biomarkers (adjusted odds ratio = 7.3, $p < 10^{-3}$), both highly statistically significant differences compared to the unexposed controls. Of the 162 workers with cumulative exposures ≤ 40 ppm-years, 33 (20%) were seropositive for one or the other oncoprotein biomarker (adjusted odds ratio = 1.4, p = 0.37) and 3 (2%) were seropositive for both oncoprotein biomarkers (adjusted odds ratio = 4.4, p = 0.06).

Although neither of the latter results were statistically significant compared to the unexposed controls, both odds ratios were elevated, and, in the latter case, the odds ratio approached statistical significance, suggesting perhaps that subgroups among the workers with cumulative exposures ≤ 40 ppm-years could have significantly elevated risks for the occurrence of the biomarkers, particularly when both biomarkers were considered together. Further analysis with stratification of this group into four subgroups with cumulative exposures of < 10 ppm-years, 10 to 20 ppm-years, 20 to 30 ppm-years, and 30 to 40 ppm-years supported this assumption. For example, for the subgroup with cumulative exposures < 10 ppm-years the adjusted odds ratios for the presence of one or both biomarkers were not greatly elevated (both < 2), and neither was close to statistical significance. On the other hand, for the other three subgroups, although the adjusted odds ratios for the presence of one biomarker were similarly not greatly elevated, the adjusted odds ratios for the presence of both biomarkers were considerably higher, indicating that the increased risk for those with cumulative exposures ≤ 40 ppm- years was primarily among those workers with cumulative exposures of 10 to 40 ppm-years. Therefore, further analysis was confined to the cohort stratified by cumulative exposures of < 10 ppm-years, 10 to 40 ppm-years and > 40 ppm-years compared to the unexposed controls, as shown in Table 1. In this case, for the sub-group with cumulative exposures of 10 to 40 ppm-years, the adjusted odds ratio for the presence of one biomarker was only 1.2 and remained statistically insignificant (p = 0.67). However, in this subgroup the adjusted odds ratio for the presence of both biomarkers was 5.7 and was statistically significant (p = 0.045). In fact, this elevated risk was not significantly different from

Table 1. Relationship between molecular biomarkers of cancer-related mutations and estimated cumulative VC exposure.

Exposure in PPM-Years (N)	Both Negative	One Positive	Both Positive	OR[1]	OR[2]
0 (N = 155)	139	16	0	1	1
< 10 (N = 77)	59	17	1	1.9	1.7
10-40 (N = 85)	67	16	2	1.2	5.7*
> 40 (N = 306)	151	122	33	5.0*	7.3*

[1] Odds ratio for one positive vs both negative adjusted for age, smoking and drinking
[2] Odds ratio for both positive vs both negative adjusted for age, smoking and drinking
* $p < 0.05$

that found for the presence of biomarkers in the group with cumulative exposures > 40 ppm-years.

DISCUSSION

These biomarker results suggest that some workers with VC exposures below the current permissible exposure limit have a risk for the occurrence of two VC-induced, cancer-related mutations that is similar to that for workers exposed above this limit and that is significantly greater than unexposed controls. Although we do not know at this time whether these individuals positive for the biomarkers will actually develop cancer, results from studies of other exposed cohorts of workers using banked serum samples suggest that similar biomarkers can have considerable positive predictive value (Husgafvel-Pursiainen *et al.* 1997). If the same proves to be true in these VC workers, this would mean that workers with cumulative exposures of 10 to 40 ppm-years (just like workers with cumulative exposures > 40 ppm-years) could be at an increased risk for cancer and that the current permissible exposure limit is not adequately protective. In addition, since workers with cumulative exposures < 10 ppm-years do not have a higher probability of occurrence of these biomarkers, exposures less than 0.25 ppm over a 40 year working lifetime may not represent a significant risk.

However, additional confounding factors may need to be considered. For example, VC workers who experience both mutations may do so because they are somehow extraordinarily susceptible to the mutagenic effects of the exposure. Other studies have demonstrated that VC workers with genetic polymorphisms in the enzymes responsible for the metabolism of VC (CYP 2E1 and ALDH-2), which could lead to an increased formation of DNA adducts at any given exposure level, are more likely to demonstrate the effects of genetic damage in terms of a significantly increased frequency of sister chromatid exchanges compared to VC workers with normal variants of these enzymes (Wong *et al.* 1998). In fact, in the present cohort thus far, one of the workers with the presence of both oncoprotein biomarkers has been found to have the susceptible variant of the CYP 2E1 gene. In addition, it is possible that genetic polymorphisms in the enzymes responsible for repair of DNA adducts could have a similar effect. Studies are underway to examine these possibilities. Eventually, it may be possible to perform the risk assessment for VC exposure on the basis of the presence of oncoprotein biomarkers stratified by genetic susceptibility due to VC metabolism and/or DNA repair, which would represent a considerable refinement over current approaches.

At the level of the individual worker, however, identification of an elevated risk for cancer based on the occurrence of mutant biomarkers and a susceptible genotype is of little significance unless secondary prevention can be employed to help minimize the risk. One such approach that might be useful with these workers would be to try to correct the specific functional deficits produced by the VC-induced mutations. For example, the mutations in p53 from VC cause common conformational changes in the protein that account for its loss of function for tumor suppression and programmed cell death (apoptosis) (Brandt-Rauf *et al.* 1996). We have shown that mutant p53 in cancer cells can be induced to revert to normal function through interaction with small peptides that represent sequences from the

regulatory domain of the protein, resulting in apoptosis of the cancer cells, but that these peptides have no adverse effect on nonmalignant human cell lines containing wild-type p53 (Kim *et al.* 1999). Such peptides could form the basis for chemotherapy for individuals with tumors that contain p53 mutations and as chemoprophylaxis for individuals who are at risk for the development of such tumors due to the occurrence of p53 mutations from their exposures, such as those VC workers who are positive for the biomarker. High-risk workers could thus be identified and treated early, preventing the development of future cancers. Therefore, ultimately, molecular biomarkers could be useful in contributing to group risk assessment to refine permissible exposure limits improving future primary prevention and in individual risk assessment to identify workers for potential chemoprophylaxis for secondary prevention of the effects of such carcinogenic exposures.

ACKNOWLEDGMENT

This work was supported in part by grants from the U.S. National Cancer Institute (R01-CA69243 and T32-CA09529) and the U.S. Environmental Protection Agency (R-825361 and R-826685), and the U.S. National Institute for Occupational Safety and Health (R01-OH04192).

REFERENCES

ATSDR (Agency for Toxic Substances and Disease Registry). 1997. Toxicological Profile for Vinyl Chloride (Update). DHHS Publication No. TP-92/20. U.S. Department of Health and Human Services, Public Health Service, Atlanta, GA, USA

Barbin A and Bartsch H. 1986. Mutagenic and promutagenic properties of DNA adducts formed by vinyl chloride metabolites. In: Singer B and Bartsch H (eds), The Role of Cyclic Nucleic Acid Adducts in Carcinogenesis and Mutagenesis, pp 345-58. International Agency for Research on Cancer, World Health Organization, Lyon, France

Brandt-Rauf PW, Chen JM, Marion MJ, *et al.* 1996. Conformational effects in the p53 protein of mutations induced during chemical carcinogenesis: Molecular dynamic and immunologic analyses. J Protein Chem 15:367-75

Brandt-Rauf PW, Luo JC, Cheng TJ, *et al.* 2000. Mutant oncoprotein biomarkers of vinyl chloride exposure: Applications to risk assessment. In: Anderson D, Karakaya AE, and Sram RJ (eds), Human Monitoring after Environmental and Occupational Exposure to Chemical and Physical Agents: Proceedings of the NATO Advanced Study Institute, pp 243-8. IOS Press, Amsterdam, The Netherlands

DeVivo I, Marion MJ, Smith SJ, *et al.* 1994. Mutant c-Ki-ras p21 protein in chemical carcinogenesis in humans exposed to vinyl chloride. Cancer Causes Control 5:273-8

Hattis D and Silver K. 1993. Use of biomarkers in risk assessment. In: Schulte PA and Perera FP (eds), Molecular Epidemiology Principles and Practice, pp 251-73. Academic Press, San Diego, CA, USA

Hollstein M, Marion MJ, Lehman T, *et al.* 1994. P53 mutations at A:T base pairs in angiosarcomas of vinyl chloride-exposed factory workers. Carcinogenesis 15:1-3

Husgafvel-Pursiainen K, Kannio A, Oksa P, *et al.* 1997. Mutations, tissue accumulations and serum levels of p53 in patients with occupational cancers from asbestos and silica exposure. Environ Mol Mutagen 30:224-30

Li Y, Marion MJ, Asherova M, *et al.* 1998. Mutant p21ras in vinyl chloride-exposed workers. Biomarkers 3:433-9

Luo JC, Liu HT, Cheng TJ, *et al.* 1998. Plasma asp13-Ki-ras oncoprotein expression in vinyl chloride monomer workers in Taiwan. J Occup Environ Med 40:1053-8

Luo JC, Liu HT, Cheng TJ, *et al.* 1999. Plasma p53 protein and anti-p53 antibody expression in vinyl chloride monomer workers in Taiwan. J Occup Environ Med 41:521-6

Kim AL, Raffo AJ, Brandt-Rauf PW, *et al.* 1999. Conformational and molecular basis for induction of apoptosis by a p53 C-terminal peptide in human cancer cells. J Biol Chem 49:34924-31

Marion MJ, Froment O, and Trepo C. 1991. Activation of Ki-ras gene by point mutations in human liver angiosarcoma associated with vinyl chloride exposure. Mol Carcinogen 4:450-4

Smith SJ, Li Y, Whitley R, *et al.* 1998. Molecular epidemiology of p53 protein mutations in workers exposed to vinyl chloride. Am J Epidemiol 147:302-8

Swenberg JA, Fedtke N, Ciroussel F, *et al.* 1992. Etheno adducts formed in DNA of vinyl chloride- exposed rats are highly persistent in liver. Carcinogen 13:727-9

Tornqvist M and Landin HH. 1995. Hemoglobin adducts for in vivo dose monitoring and cancer risk estimation. J Occup Environ Med 37:1077-85

Wong RH, Wang JD, Hsieh LL, *et al.* 1998. Effects on sister chromatid exchange frequency of aldehyde dehydrogenase 2 genotype and smoking in vinyl chloride workers. Mut Res 420:99- 107

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1303-1313 (2002)

Use of Toxicological Data in Estimating Reference Values for Risk Assessment

Carole A. Kimmel[1]
National Center for Environmental Assessment (8623D), U.S. Environmental Protection Agency, Ariel Rios Building, 1200 Pennsylvania Ave., NW, Washington, DC 20460; Tel(voice):202-564- 3307, Tel(fax):202-565-0078; kimmel.carole@epa.gov.

ABSTRACT

A number of programs within the U.S. Environmental Protection Agency (USEPA) currently set less-than-lifetime exposure limits in addition to the chronic reference dose (RfD) and reference concentration (RfC). A review of procedures within the USEPA for setting reference values suggests that less-than-lifetime reference values should be more routinely developed and captured in the USEPA's online IRIS database where chronic RfDs and RfCs, as well as cancer slope factors, are currently available. A review of standard testing study protocols was conducted to determine what data were available for setting acute, short-term, and longer-term reference values, as well as chronic values. This review was done from the point of view of endpoints assessed for specific organ systems (both structural and functional), life stages covered by exposure and outcome, durations of exposure covered and the outcomes evaluated for each, and evaluation of latency to response and/or reversibility of effects. This review revealed a number of data gaps and research needs, including the need for an acute and/or short-term testing protocol that can be used to set acute and short-term reference values, a strategy for when to conduct more extensive testing based on initial screening data or other information (*e.g.*, chemical class, pharmacokinetics, mode of action), additonal standard testing guidlines protocols to allow more complete assessment of certain organ systems and life stages, development of pharmacokinetic data for different life stages, toxicity related to

1 The views expressed in this paper are those of the author and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of trade names of commercial products does not constitute endorsement or recommendation for use. This manuscript is considered to be a work of the U.S. Government and is therefore not copyrighted.

1080-7039/02/$.50

aging, and latency to response, particularly long-term latency as a result of developmental exposures. The implications of this review are discussed relative to characterizing hazard data for setting reference values, and the potential impacts on uncertainty factors and low-dose extrapolation.

Key Words: toxicology, hazard characterization, reference dose, reference concentration, risk assessment.

INTRODUCTION

Current approaches to risk assessment for health effects other than cancer involve the application of various uncertainty factors (UFs) to a no-observed-adverse-effect level (NOAEL) or a benchmark dose (BMD) to derive a reference value used for setting exposure limits. The reference dose (RfD) and reference concentration (RfC) are values used in various settings by the Environmental Protection Agency (USEPA) for limiting chronic oral or inhalation exposures, respectively. Methodology for deriving chronic RfDs and RfCs has been in place within the USEPA for a number of years (USEPA 1994, 2002), and the reference values along with supporting information are stored in a USEPA database, the Integrated Risk Information System (IRIS) (USEPA 2002). Since 1996, there has been an effort to develop more comprehensive risk assessments as the basis for setting RfDs and RfCs, as well as cancer slope factors, so that the IRIS summary is supported by documentation from the fuller assessment. Both the IRIS summaries and the background documents for the newer assessments are available online from the IRIS database (http://www.epa.gov/iris/index.html).

Also, in recent years, a number of activities have been underway at the USEPA and in other forums to more closely examine the methods that have been used for hazard characterization and dose-response assessment, to make recommendations for improving the process, and to identify data gaps for needed research and information to improve the process. These improvements include, *e.g.*, development of more quantitative alternatives to the NOAEL approach, particularly the BMD methodology (USEPA 2000), revision of the cancer risk assessment guidelines (USEPA 1999a), harmonization of cancer and noncancer risk assessment (Gaylor *et al.* 1999; Butterworth and Bogdanffy 1999; USEPA 1997, 1998a; Bogdanffy *et al.* 2001), exploration of the relationship between exposure duration and toxicity (USEPA 1998b), evaluation of the methodologies for testing and risk assessment to improve the protection of children's health from pesticide exposures (USEPA 1999, 2002b) in response to the Food Quality Protection Act of 1996, and a reevaluation of the RfD/RfC methodology, with a focus on protection of sensitive subpopulations (Kimmel 2001).

In this paper, several issues will be discussed that have been raised by a technical panel (hereafter RfD Technical Panel) established under the auspices of USEPA's Risk Assessment Forum to review the RfD and RfC processes. The purpose of this review is to evaluate the current state-of- the-art for hazard and dose-response assessment with a focus on protection of potentially susceptible sub-populations, to summarize how additional scientific issues not currently being considered can enhance the RfD and RfC process, and to raise issues that should be further

explored or developed for consideration in the process. The RfD Technical Panel has emphasized that the process should not be considered static, but continually evolving with new information and scientific advances incorporated as new reference values are set or as current RfDs and RfCs are reevaluated. The primary focus of this paper will be on (1) the importance of setting reference values for different durations of exposure, (2) the types of toxicity data available for use in hazard characterization and dose-response assessment, and (3) data gaps and research needs for improving the process.

Reference Values for Less Than Lifetime Exposures

As indicated above, the RfD and RfC are chronic exposure values developed by the USEPA to be used for limiting chronic exposure scenarios in humans. Yet many exposures in the environment as well as in occupational settings are acute or short-term repeated or intermittent exposures. Because of this, several offices within the USEPA as well as other government agencies and organizations set acute, short-term, and longer-term reference values, as well as chronic reference values. For example, USEPA's Office of Pesticide Programs sets acute dietary RfDs and sometimes short-term and intermediate dermal and inhalation reference values, especially for residential exposures (USEPA 1998e). The USEPA's National Center for Environmental Assessment is developing methodology for the Office of Air and Radiation to be used for setting acute inhalation reference values (Acute Reference Exposures — AREs) for exposures less than or equal to 24 hours (USEPA 1998c). The USEPA's Office of Water sets 1-day, 10-day, and longer-term drinking water Health Advisories that are nonregulatory standards used in emergency spill or contamination situations (Orme and Ohanian 1991). The Acute Exposure Guideline Levels (AEGLs) (NRC 2000) are set by a National Research Council committee for once-in-a-lifetime short-term exposures to airborne concentrations. AEGLs are used for a variety of emergency situations, rare events such as evacuations, and are used as threshold exposure limits ranging from 10 min to 8 hours. In terms of occupational exposure limits, the American Conference of Governmental Industrial Hygienists (ACGIH) sets short-term exposure limits (STELs; 15 min) (ACGIH 2000) and threshold limit values (TLVs; 8 hour time- weighted averages) for short duration exposure limits that may be repeated frequently for a working lifetime. The National Institute for Occupational Safety and Health (NIOSH) sets RELs (Recommended Exposure Limits) (NIOSH 1992), while the Occupational Safety and Health Administration (OSHA) sets PELs (Permissible Exposure Limits) (OSHA 1999); PELs are enforceable occupational exposure limits. The Agency for Toxic Substances and Disease Registry (ATSDR) sets acute ($\leq$ 4 days), intermediate (15 to 364 days) and chronic ($\geq$ 365 days) Minimal Risk Levels (MRLs), which are defined as an estimate of daily human exposure to a hazardous substance that is likely to be without appreciable risk of adverse noncancer health effects over a specified route and duration of exposure. MRLs are substance-specific estimates intended to be screening levels in the identification of contaminants and potential health effects that may be of concern; they do not define clean-up or action levels (ATSDR 1996).

Thus, a number of program offices within the USEPA as well as other agencies and organizations already set exposure limits for less-than-lifetime exposure. The

possibility of standardizing these values across programs within the USEPA, where possible, is being explored with the intent of deriving acute, short-term, longer-term, and chronic values in risk assessments conducted by the USEPA, and capturing these values in the IRIS database. A set of definitions for these durations of exposure in humans was developed by the RfD Technical Panel. These definitions are compatible with those used by various programs within the USEPA, and they are meant to be flexible because the duration in each case represents a range of time and can be adjusted depending on the exposure scenario of concern.

acute — exposure by the oral, dermal, or inhalation route for 24 hours or less;

short-term — repeated exposure by the oral, dermal, or inhalation route for up to 30 days;

longer-term — repeated exposure by the oral, dermal, or inhalation route for up to approximately 7 years (10% of the lifespan) in humans (up to approximately 90 days in standard laboratory animal species, including rats, mice, and rabbits);

chronic — repeated exposure by the oral, dermal, or inhalation route for up to the average life span in humans (up to approximately 2 years in standard laboratory rodents).

Review of Current Testing Protocols and Data Requirements

In order to set different duration reference values, a variety of types of data are needed, as different duration RfDs and RfCs may need to be based on different types of data and endpoints. The RfD Technical Panel recognized early on in its review of the process that data available from current testing protocols for setting various duration reference values are limited or likely to be lacking altogether, especially for acute and short-term reference values. A review of the current pesticide and toxic substances testing protocols and data requirements (USEPA 1998d) was undertaken as a way of determining the extent of data available for setting various duration reference values and where there are data gaps. This review of standard testing protocols was done with four major areas of focus:

- The endpoints generally evaluated in each testing protocol, as well as those specifically evaluated for several organ systems, including both structural and functional evaluations. Systems reviewed in depth included the reproductive, nervous, immune, and cardiovascular systems.
- The life stages covered by exposure and outcome, from conception to death.
- The duration of exposure used in various protocols, and the outcomes evaluated for each duration.
- The evaluation of latency to response, and/or reversibility of effects.

While some systems are more thoroughly covered by testing protocols and data requirements, *e.g.*, the reproductive and nervous systems, others are not well evaluated, *e.g.*, the cardiovascular and immune systems, unless these are known or suspected target organs.

Figure 1 shows a time line for various life stages with the timing of exposure for various standard testing protocols superimposed upon it (hatched bars). Major endpoints evaluated are shown in the boxes and the timing of evaluation is indicated by arrows or brackets. On an organ system basis, the reproductive system is evaluated for both structural and functional development and alterations, not only in the reproduction and fertility study, but also in the prenatal developmental toxicity study, the developmental neurotoxicity study, the dominant lethal study, and the subchronic and chronic toxicity studies. On the other hand, the cardiovascular system is evaluated only on a structural basis in the prenatal developmental toxicity study, and at necropsy in the acute, subchronic, and chronic toxicity studies in rodents, as well as in the chronic toxicity study in dogs (not shown in the figure).

All life stages, except old age, are covered by exposure in one or another standard testing study protocol. A more careful examination of the evaluations done relative to the various life stage exposures indicates, however, that there are no protocols that begin exposure during development (prenatal and early postnatal development) and follow test subjects into old age. In addition, there are no studies that examine the effects of exposure and outcome during old age, as the current chronic/carcinogenicity study in rodents is a 2-year exposure study that stops short of later old age. This is even more the case in rodents on restricted diets, as they do not age as rapidly and have significantly extended life-spans over animals that are fed ad libitum.

All exposure durations are covered in the various study types, but the current guideline acute toxicity study is designed primarily to establish a median lethal dose (LD_{50}), and does not include the types of outcomes that are needed to establish a NOAEL or BMD for an acute reference value. Data from other studies can be used to supplement the database for an acute reference value, *e.g.*, the response to the initial dose in subchronic studies, the acute neurotoxicity testing study, the prenatal developmental toxicity study, the developmental neurotoxicity study, and the reproduction and fertility study. The prenatal developmental toxicity study, the developmental neurotoxicity study, and the reproduction and fertility study all involve much longer than acute exposures, but developmental effects have been clearly shown in the literature to be inducible by a single exposure; thus, data from these studies are also considered in setting the acute reference value. No specific testing protocols are available for short-term exposure studies, *i.e.*, more than 24 hours up to 30 days. However, data from other studies, as indicated above for acute toxicity, can be used to set a short-term reference value. A good deal of data are available for setting the longer-term reference value, in particular, the adult subchronic study, the subchronic neurotoxicity study, the immunotoxicity study, and the reproduction and fertility study. Data from the prenatal developmental toxicity study and the developmental neurotoxicity study should also be considered in setting the longer-term reference value. The chronic study protocol, particularly the combination chronic/carcinogenicity study protocol, provides a great deal of data for setting a chronic reference value. The prenatal developmental toxicity, developmental neu-

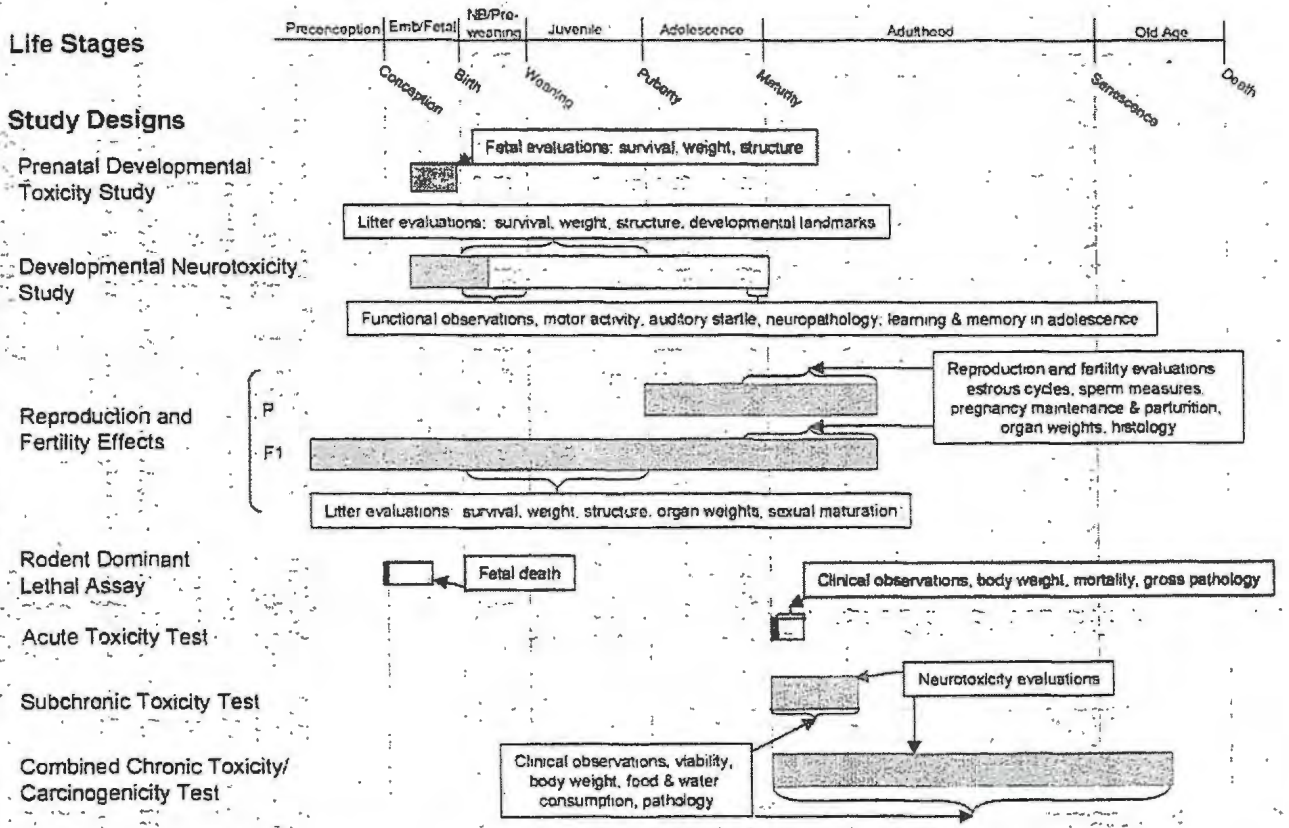


Figure 1. **A life stage view of timing and duration of exposure in standard toxicity testing protocols.** The diagram shows life stages on the horizontal line at the top with the timing of various events indicated by vertical dashed lines. The times between different events are relative and not indicative of any particular species. Listed vertically on the left are the titles of a number of standard testing protocols (USEPA 1998d), and the hatched bars to the right of each title correspond to the exposure periods for each protocol. Major endpoints evaluated in each protocol are shown in the boxes. NB = newborn; P = parental generation; F_1 = first filial generation.

rotoxicity, and reproduction and fertility study data also have been and should continue to be considered in setting chronic reference values. The so-called chronic study in dogs involves a one-year exposure study which is actually a short-term exposure by the definitions indicated above, as it covers only about 1/15th of the life-span of beagle dogs, the species typically used.

The evaluation of latency to response and/or reversibility of effects is minimal in standard testing protocols, and there are only two studies, the developmental neurotoxicity study and the adult delayed neurotoxicity study, that evaluate nervous system function after exposure has been stopped for a short period of time. The chronic rodent assays include an optional recovery period in a subset of animals that entails ending exposure after at least 12 months in a separate group of animals with evaluation 28 days or more after exposure ceases however, this assessment of recovery is not required. In particular, there is no evaluation of latency to response and/or reversibility of effects in animals at older ages following exposures during their early development. With the accumulation of evidence for an impact of early exposures on neurological development and function, immunologic function, diabetes, and heart disease in advanced age, these types of evaluations could provide information of significance for effects with long-term latency to response (Selevan *et al.* 2000).

Data Gaps for Testing and Risk Assessment

The review of current testing approaches from the perspective of life stages, durations of exposure and evaluation, endpoints and systems evaluated, and latency to response and/or reversibility revealed a number of data gaps for testing and risk assessment. These include:

- The need for an acute and/or short-term testing protocol that can be used to derive the type of data needed to set acute and short-term reference values (for excellent examples see Gad and Chengelis 1998);

- A strategy is needed for determining when to conduct more extensive life stage and endpoint evaluations than are currently conducted, or than are reasonable in the initial screening of an agent. For example, a number of additional developmental toxicity testing study guidelines have been suggested for testing pesticides and other agents (USEPA 1999b) in order to more adequately evaluate children's health risks in accordance with the Food Quality Protection Act. These include more specialized developmental neurotoxicity tests, a developmental immunotoxicity protocol, a developmental carcinogenesis testing protocol, a protocol for direct dosing of neonates when there is a question about whether adequate exposure to neonates has occurred, and a battery of tests for screening agents for their potential to interfere with endocrine activity.

- The need for new standard testing guideline protocols to allow more complete assessment of certain organ systems, including an evaluation of both structure and function. For example, the structure of the cardiovascular system is evaluated grossly in the prenatal developmental toxicity study, and histologi-

cally in the subchronic and chronic studies. However, no functional assessment of this system is done in any current testing protocols, yet cardiovascular disease is one of the most prevalent public health problems. In cases where an effect on cardiovascular function has been suspected, *e.g.*, lead, chlorinated fluorocarbon replacements, more detailed evaluations of both structure and function have been conducted, but there are no guidelines for the conduct of such studies. The immune system is evaluated only in young adults, and testing includes assays only for hypersensitivity, humoral immunity, and nonspecific cell-mediated immunity.

- There is a great need for pharmacokinetic data as an aid to study design and interpretation of toxicity data. In addition, there is a need for review and development of pharmacokinetic and pharmacodynamic factors for different life stages, particularly during prenatal development including placental transfer, during postnatal development including lactational transfer, during adolescence and puberty, and related to the aging process.

- In terms of toxic responses during aging, research is needed to more fully characterize those effects occurring at some interval after early exposure (including during prenatal and early postnatal development), as well as those occurring in old age with concomitant exposure during that life stage.

- Research is also needed on how to evaluate latency to response, particularly those effects that occur at some point after exposure ends. As indicated earlier, a number of examples are available in the literature on neurodegenerative diseases, immunological deficits, diabetes, cardiovascular disease, and cancers related to earlier developmental exposures.

Future Considerations

The setting of various duration reference values must be based on some definition of the needed database characteristics. The RfD Technical Panel is considering ways of characterizing the extent of the database available for setting reference values, taking into consideration the types of data available from human studies and literature data, as well as from standard toxicity testing protocols, the types and variety of endpoints evaluated both in terms of structure and function, the life stages evaluated, the durations of exposure included, and other information, such as pharmacokinetics, biological and chemical characteristics including mode of action, and known limits on reserve capacities and repair of agent-induced damage. The use of scientific judgment in characterizing the extent of the database on a continuum from minimal to robust will be extremely important. Studies showing only mortality or severe toxicity are not sufficient to set a reference value.

The implications of setting various duration reference values are currently being explored by the RfD Technical Panel, particularly as they affect the uncertainty factors applied to the NOAEL or BMD (USEPA, 2002c). For example, the consideration of data on a number of different life stages and organ systems at each of those life stages may affect the size of the uncertainty factor used to account for

variability within humans. The development and use of mode of action information is extremely important, particularly in how it can be used for extrapolation to humans, as well as to support a linear or nonlinear approach for low dose extrapolation. Further development of pharmacokinetic and pharmacodynamic data to develop so-called "data-derived" factors is also encouraged (*e.g.*, Renwick 1993; Dourson *et al.* 1998; Zhao *et al.* 1999). Finally, there are several investigators working to develop alternative approaches to the uncertainty factor approach for deriving reference values (*e.g.*, Maull *et al.* 1997; Slob and Pieters 1998; Swartout *et al.* 1998; Brand *et al.* 1999; Gaylor and Kodell 2000). Such efforts will undoubtedly provide new insights into improvements for setting reference values or suggest alternatives to the reference value process.

ACKNOWLEDGMENTS

I thank the members of the RfD Technical Panel, whose hard work has brought us to the point reflected in this paper: Bob Benson, Gary Foureman, Lee Hofmann, Gary Kimmel, Susan Makris, Deirdre Murphy, Edward Ohanian, Jennifer Orme-Zavaleta, Deborah Rice, Jennifer Seed, Hugh Tilson, and Vanessa Vu.

REFERENCES

ACGIH (American Conference of Governmental Industrial Hygienists). 2000. TLVS and BEIs. Threshold limit values for chemical substances and physical agents, biological exposure indices. American Conference of Governmental Industrial Hygienists, Cincinnati, OH, USA

ATSDR (Agency for Toxic Substances and Disease Registry). 1996. Minimal Risk Levels for priority substances and guidance for derivation. Fed Reg 61:25873-81

Bogdanffy MS, Daston G, Faustman EM, *et al.* 2001. Harmonization of cancer and non-cancer risk assessment: Proceedings of a consensus-building workshop. Toxicol Sci (*in press*)

Brand KP, Rhomberg L, and Evans JS. 1999. Estimating noncancer uncertainty factors: Are ratios NOAELs informative? Risk Anal 19:295-308

Butterworth BE and Bogdanffy MS. 1999. A comprehensive approach for integration of toxicity and cancer risk assessments. Regulatory Tox Pharm 29:23-36

Dourson ML, Maier A, Meek B, *et al.* 1998. Boron tolerable intake: Re-Evaluation of toxicokinetics for data-derived uncertainty factors. Biol Trace Elem Res 66(1-3):453-63-600/R-98-051

Gad SC and Chengelis CP. 1998. Acute Toxicology Testing, pp 305-60. Academic Press, San Diego, CA, USA

Gaylor DW and Kodell RL. 2000. Percentiles of the product of uncertainty factors for establishing probabilistic reference doses. Risk Anal 20:245-50

Gaylor DW, Kodell RL, Chen JJ, *et al.* 1999. A unified approach to risk assessment for cancer and noncancer endpoints based on benchmark doses and uncertainty/safety factors. Regul Toxicol Pharmacol 29(2 Pt 1):151-7

Kimmel CA. 2001. U.S. EPA reference dose/reference concentration methodology: Update on a review of the process. Human Ecol Risk Assess 7:117-23

Maull EA, Cogliano VJ, Scott CS, *et al.* 1997. Trichloroethylene health risk assessment: a new and improved process. Drug Chem Toxicol 20:427-42

NIOSH (National Institute for Occupational Safety and Health). 1992. National Institute for Occupational Safety and Health recommendations for occupational safety and health.

Compendium of policy documents and statements. DHHS (NIOSH) Publication No. 92-100.

NRC (National Research Council). 2000. Acute Exposure Guideline Levels for Selected Airborne Chemicals, vol 1. National Academy Press, Washington, DC, USA

Orme J and Ohanian EV. 1991. Health advisories for pesticides. In: Richardson ML (ed), Chemistry, Agriculture and the Environment, pp 429-43. Royal Society of Chemistry, Cambridge, UK

OSHA (Occupational Safety and Health Administration). 1999. OSHA 29CFR 1910.1000. Table Z1. U.S. Government Printing Office, Office of the Federal Register, Washington DC, USA

Renwick AG. 1993. Data-derived safety factors for the evaluation of food additives and environmental contaminants. Food Additives and Contaminants 10:275-305

Selevan SG, Kimmel CA, and Mendola, P., eds. 2000. Identifying critical windows of exposure for children's health. Environ Health Perspect 107 (suppl 3):451-597

Slob W and Pieters MN. 1998. A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: general framework. Risk Anal 18:787-98

Swartout JC, Price PS, Dourson ML, *et al.* 1998. A probabilistic framework for the reference dose (probabilistic RfD). Risk Anal 18:271-82

USEPA (U.S. Environmental Protection Agency). 1994. Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry. EPA-600/8-90-066F. Available from the U.S. EPA Risk Information Hotline at telephone 1-513-569-7254, fax 1-513-569-7159, email RIH.IRIS@epamail.epa.gov and/or the National Technical Information Service, Springfield, Virginia 22161, (703) 605-6000, http://www.ntis.gov

USEPA (U.S. Environmental Protection Agency). 1997. Summary of the US EPA Colloquium on a Framework for Human Health Risk Assessment, vol. 1. EPA-600/R-99-001, available at http://www.epa.gov/nceawww1/colloquium.htm

USEPA (U.S. Environmental Protection Agency). 1998a. Summary of the US EPA Colloquium on a Framework for Human Health Risk Assessment, vol. 2. EPA-600/R-98-155, available at http://www.epa.gov/nceawww1/colloquium.htm

USEPA (U.S. Environmental Protection Agency). 1998b. Report of the Workshop on the Relationship Between Exposure Duration and Toxicity. EPA-600/R-99-081. Summary available at http://www.epa.gov/ncea/smdurto1.htm; report available from NCEA Technical Information Services (202) 564-3261

USEPA (U.S. Environmental Protection Agency). 1998c. Methods for Exposure-Response Analysis for Acute Inhalation Exposure to Chemicals, Development of the Acute Reference Exposure. EPA/600/R-98/051, External Review Draft. Office of Research and Development, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1998d. Health Effects Test Guidelines, OPPTS 870 Series, available at http://www.epa.gov/docs/OPPTS_Harmonized/870_Health_Effects_Test _Guidelines/ Series/.

USEPA (U.S. Environmental Protection Agency). August 11, 1998e. Hazard Identification - Toxicology Endpoint Selection Process. Office of Pesticide Programs, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1999a. Guidelines for Carcinogen Risk Assessment. NCEA-F-0644. July 1999. Review Draft. Risk Assessment Forum, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1999b. Toxicology Data Requirements for Assessing Risks of Pesticide Exposure to Children's Health - Report of the Toxicology Working Group of the 10X Task Force (April 28, 1999 draft), available at http://www.epa.gov/scipoly/sap/1999/index.htm#may

USEPA (U.S. Environmental Protection Agency). 2001. Benchmark Dose Technical Guidance Document EPA-630/R-00-001. Draft, Available at http://www.epa.gov/ncea/

USEPA (U.S. Environmental Protection Agency). 2002a. Integrated Risk Information System (IRIS) online, Available at http://www.epa.gov/ncea/iris.htm

USEPA (U.S. Environmental Protection Agency). 2002b. Determination of the Appropriate FQPA Saftey Factor(s) for Use in the Tolerance-Setting Process. Office of Pesticide Programs, Office of Prevention Pesticides, and Toxic Substances. Available at http://www.epa.gov/oppfead1/trac/science/#10_fold

USEPA (U.S. Environmental Protection Agency). 2002c. A Review of the Reference Dose and Reference Concentration Processes (Exernal Review Draft). Office of Reseach and Development (EPa/6i0/p-oz.002A, May). Available at http://www.epa.gov/eims/eimsapi.detail?deid=51717&partner=ORD-NCEA

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 5:973-83

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1315-1338 (2002)

Linking Pharmacokinetics and Biomarker Data to Mechanism of Action in Risk Assessment

James A. Swenberg,* Nadia Gorgeiva, Amy Ham, Hasan Koc, Eric Morinello, Asoka Ranasinghe, Patricia Upton, and Vernon Walker
Laboratory of Molecular Carcinogenesis and Mutagenesis, Campus Box 7400, School of Public Health, The University of North Carolina, Chapel Hill, NC 27599-7400

ABSTRACT

Incorporating information on metabolism, pharmacokinetics, and DNA and protein biomarkers provides a means to integrate these important factors into the risk assessment process. Such data are useful for species to species extrapolation, high- to low-dose extrapolation and PBPK modeling. In addition, these data are critical for understanding the mode of action for chemical carcinogens. Through the use of mass spectrometry, stable isotopes can be used to unequivocally demonstrate pathways of formation of biomarkers and relationships between exogenous and endogenous processes. This paper reviews what has been learned for two carcinogens, vinyl chloride and butadiene. It is clear that such data play major roles in improving the understanding of how chemicals cause cancer, extending the range of data on exposure-response relationships, and examining interspecies differences and inter-individual differences that may affect susceptibility. As such, it is also clear that these data play a critical role in improving the accuracy of risk assessments.

Key Words: pharmacokinetics, biomarkers, vinyl chloride, 1,3-butadiene, DNA adducts, hemoglobin adducts.

INTRODUCTION

Risk assessment is the process of evaluating data sets pertinent to hazard identification, dose-response, and human exposure in order to characterize the risks associated with exposure to a chemical, process or mixture. This requires integration of knowledge on metabolism, toxicology, carcinogenesis, mutagenesis, and understanding of mechanism from studies in mammalian and lower species. Similar information on humans and data on exposure are necessary in order to make the best estimate of human health risks associated with human exposure. The goal of

* Corresponding author: Tel(voice):919-966-6142, Tel(fax):919-966-6123; James_Swenberg@unc.edu

1080-7039/02/$.50

risk assessment is to *accurately predict* these risks so that the health of exposed individuals is truly protected, while not over regulating useful chemicals. In the absence of good data, the default position is one of being *protective* of public health. When the default position is used, many uncertainties are embedded within the risk assessment. Some of the most important uncertainties that plague accurate risk assessment are (1) high- to low-dose extrapolation; (2) species to species extrapolation; (3) mechanism(s) responsible for the effect; (4) interindividual differences in susceptibility; and (5) poor-quality human exposure data. The goal of this paper is to link data on pharmacokinetics and biomarkers to our understanding of the mechanism of action of two chemicals, vinyl chloride and butadiene, to examine its potential to improve risk assessment.

Toxicology studies for cancer induction usually employ exposures that range from the maximum dose that can be administered to an animal without shortening its life span for any endpoint other than cancer, to doses that are two to ten times lower. These doses are usually well above the range of exposure studied in epidemiologic studies. Risk assessors and modelers are frequently focused on environmental exposures that are often several orders of magnitude lower than either toxicology or occupational studies. The two chemicals discussed in this paper are unusual, in that pharmacokinetic and biomarker data are now available over a broad range of exposures covering the high initial bioassay exposures to within one order of magnitude of current occupational exposures.

Some of the key principles involved in the use of pharmacokinetics and biomarkers are outlined in Figure 1. First, exposure to a chemical occurs by one or more routes. Following absorption, the chemical is distributed to various compartments within the body, where it is metabolized. There are two forms of metabolism, metabolic activation and detoxication. These two pathways compete with each other and are usually enzymatic in nature. Because of this, they can be induced or saturated. Thus, important differences may exist between high and low exposures, single and continuous exposure, different species, and between individuals. This is a critical point for accurate risk assessment, as it is the balance between activation and detoxication that determines many of the endpoints used in biomarker research, *e.g.*, the binding of genotoxic agents to macromolecules. Such biomarkers are typically protein or DNA adducts. Protein adducts are not causally involved in the carcinogenic process. They have the advantage of not being repaired, so a protein accumulates adducts over the life span of the protein. DNA adducts are thought to be involved in the initiation and progression of cancer. However, these lesions in DNA can be repaired and vary at least a 1000-fold in their ability to cause mutations. Finally, even when an adduct is present in DNA, cell proliferation is required to generate a mutation. DNA and protein adducts represent biomarkers of exposure, and DNA adducts may also be an early biomarker of effect.

It is important to recognize that exogenous chemical exposure is not the only cause of mutations. In fact, the spontaneous rate of mutations is quite high. Mutations arise from several endogenous sources, including depurination, DNA polymerase errors, and endogenous DNA damage resulting from alkylation, reactive oxygen species, and lipid peroxidation. Additional exogenous sources of mutations are associated with lifestyle and radiation. All of these types of DNA damage are

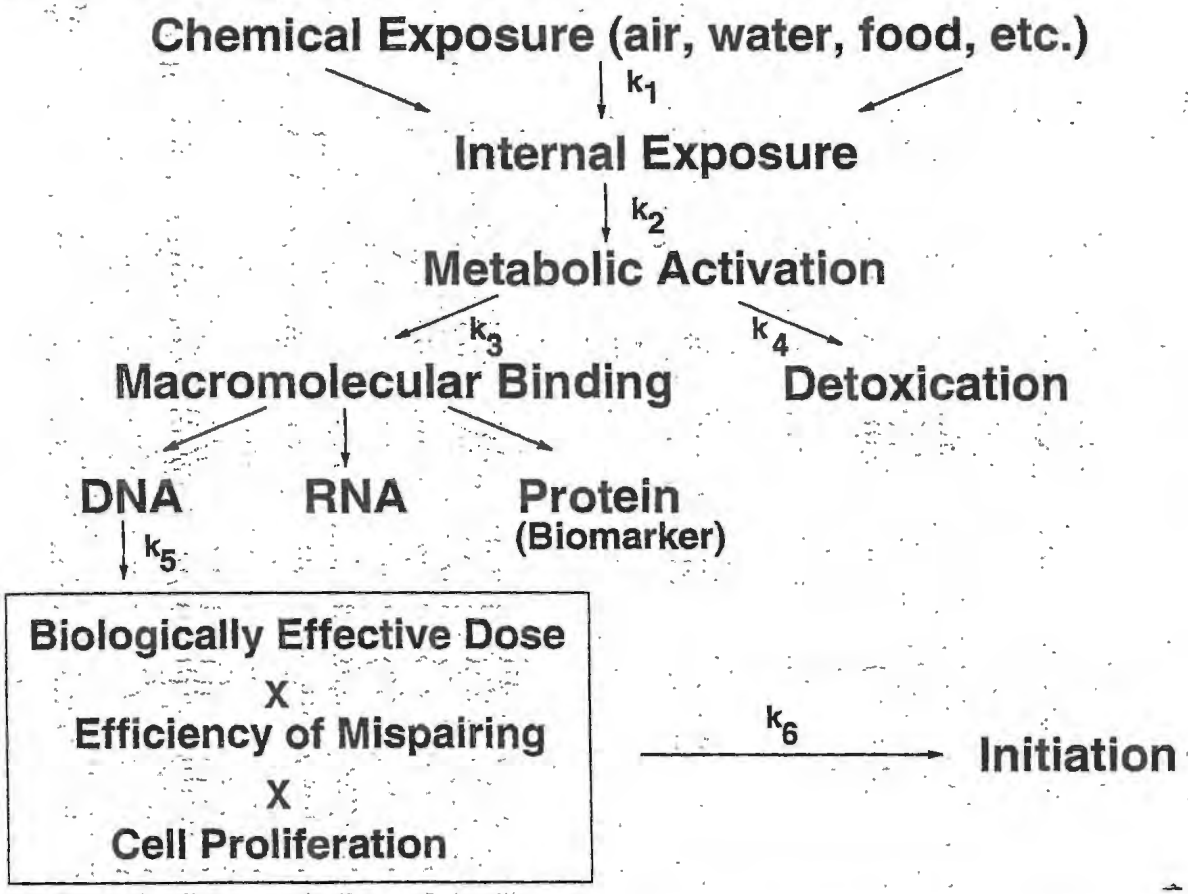


Figure 1. Scheme of some of the key principles involved in the use of pharmacokinetics and biomarkers.

subject to DNA repair. It is the DNA damage that escapes the filter of DNA repair that is available to cause mutations and ultimately cancer.

All of these processes, including metabolic activation, detoxication, DNA repair and cell proliferation are dynamic. They change with the time and extent of exposure, and are confounded by exposures to other natural and synthetic materials. There will not be one rate constant that accurately predicts for a population. Rather, by understanding the dominant pathways, polymorphisms that affect these pathways, and conditions that induce or saturate these pathways, risk assessors will be able to better integrate data from toxicology, pharmacokinetic, biomarker, and molecular epidemiology studies so that more accurate assessments of risk can be made.

VINYL CHLORIDE

Vinyl chloride (VC) is a known human and animal carcinogen that primarily induces hepatic angiosarcomas in humans and experimental animals (IARC 1987). The carcinogenic response in humans has been associated with relatively high occupational exposures (more than 50 parts per million). These exposures occurred prior to the identification of VC as a carcinogen. Current occupational exposure limits are 1 ppm and actual exposure is running about a tenth of a part per million in most plants. In contrast, concerns for environmental exposure are often related to Superfund sites. VC is present in ~133 Superfund sites (http://www.epa.gov/superfund). It is not there because someone dumped vinyl chloride at a site. Rather, it is there because soil microbes metabolize trichloroethylene and perchloroethylene to VC (Owen and Glaister 1990). In contrast to past occupational exposures >50 ppm and current exposures ≤1 ppm, the environmental exposures are present in part per billion amounts (Owen *et al.* 1987). This requires the risk assessments of VC to cover a very large range, roughly six to eight orders of magnitude away from the animal and human exposures that were associated with adverse health effects.

Metabolism and Pharmacokinetics

The metabolism of VC has been studied for over 25 years. Earlier research by Watanabe and Gehring (1976) demonstrated that VC was metabolized by cytochrome P450 in a saturable manner. That is, the metabolic activation had its steepest slope at low exposures and had little or no additional activation at exposures > 250 ppm. Later studies by Guengerich and colleagues demonstrated that CYP-2E1 was responsible for the formation of chloroethylene oxide (CEO), the ultimate carcinogen that induces the DNA adducts of VC (Figure 2) (Guengerich *et al.* 1991; Guengerich 1992). CEO can also rearrange to chloroacetaldehyde, but this process is much slower and does not represent the primary pathway for adduct induction. Detoxication pathways for VC include conjugation to glutathione, hydroxylation by epoxide hydrolase and reduction by alcohol dehydrogenase.

DNA Adducts

The DNA adducts of VC have been previously studied, but only at exposures of 500 to 600 ppm. These studies demonstrated that 7-(2-oxoethyl)guanine (OEG) was

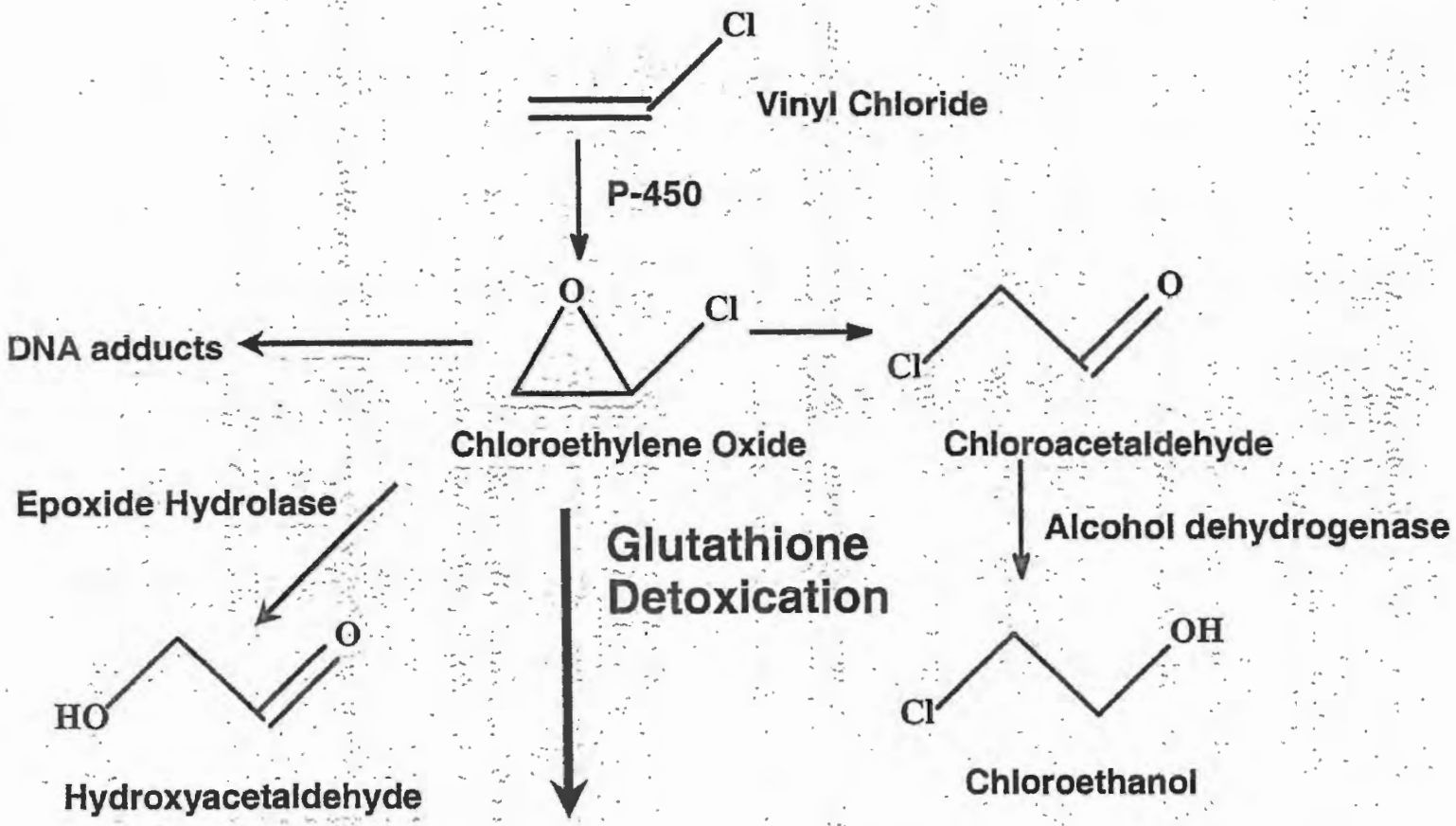


Figure 2. General scheme of the metabolism and detoxication of vinyl chloride.

the major DNA adduct, while N^2,3-ethenoguanine (EG), 3,N^4-ethenodexycytidine (EdC), and 1,N^6-ethenodeoxyadenosine (EdA) were present in much smaller amounts (Fedtke *et al.* 1989; Fedtke *et al.* 1990b; Guichard *et al.* 1996). The etheno adducts differed from OEG, however, in that they actively caused miscoding during DNA synthesis (Barbin *et al.* 1985; Cheng *et al.* 1991; Mroczkowska and Kusmierek 1991; Cheng *et al.* 1991; Singer *et al.* 1991). Highly sensitive and specific methods were developed to measure the etheno adducts. EdA and EdC have most commonly been measured by immunoaffinity/^{32}P-postlabeling (Nair *et al.* 1995; Fernando *et al.* 1996; Guichard *et al.* 1996), while EG was measured by GC/MS (Fedtke *et al.* 1990a; Fedtke *et al.* 1990b; Swenberg *et al.* 1995). The etheno adducts induced by VC are also formed endogenously in unexposed animals and humans (El-Ghissassi *et al.* 1995; Nair *et al.* 1999; Ham *et al.* 2000; Morinello *et al.* 2000). The primary mechanism for this appears to be via 4-hydroxynonenal arising from lipid peroxidation.

Using a newly developed immunoaffinity/GC high-resolution mass spectrometry method for EG (Ham *et al.* 1999), the first molecular dosimetry studies have been conducted on rats exposed to 0, 10, 100, or 1100 ppm VC for 1 or 4 weeks (6 h/day, 5 days/wk) (Swenberg *et al.* 2000a). Table 1 shows the presence of EG in all control animals and the increases associated with the level and length of exposure. While exposure to 10 ppm for 4 weeks results in only a 6-fold increase over the endogenous amount of EG present in control rats, exposure to 100 or 1100 ppm resulted in 25- and 42-fold increases over controls. It is also possible to interpolate between the control EG and that of the 10 ppm rats exposed for 4 weeks to estimate the molecular dose associated with current occupational exposures of 0.1 to 1.0 ppm VC. These exposures would result in 6% and 60% increases over the endogenous levels of EG. The 4-week data at zero, 10, 100, and 1100 parts per million mirrors the supralinear dose response relationship that was demonstrated for metabolism (Watanabe and Gehring 1976; Gehring *et al.* 1978) and the large animal carcinogenicity study of Maltoni *et al.* (1981) (Figure 3). The data are remarkably similar to the PBPK model developed by Clewell *et al.* (1995) that was used in the recent U.S. Environmental Protection Agency (USEPA) risk assessment published in the new IRIS document (USEPA 2000). The old USEPA risk assessment had a factor of 84 $\times 10^{-6}$ cases of cancer per μg VC/M^3 (USEPA 1994). Incorporation of the biomarker and pharmacokinetic data on vinyl chloride has lowered that to 4.4×10^{-6} (USEPA 2000). Similar estimates of risk have been predicted by several different PBPK models (Table 2) (Chen and Blancato 1989; Reitz *et al.* 1996; Clewell *et al.* 1995). The

Table 1. Exposure response data from rats exposed to vinyl chloride for 1 or 4 weeks.

Vinyl Chloride Exposure	EG (pmol/μmol deoxyguanosine)	
(ppm)	1 Week Exposure	4 Weeks Exposure
0	0.075 ± 0.036	0.11 ± 0.053
10	0.196 ± 0.048	0.532 ± 0.106
100	0.68 ± 0.09	2.28 ± 0.18
1100	1.25 ± 0.19	3.8 ± 0.4

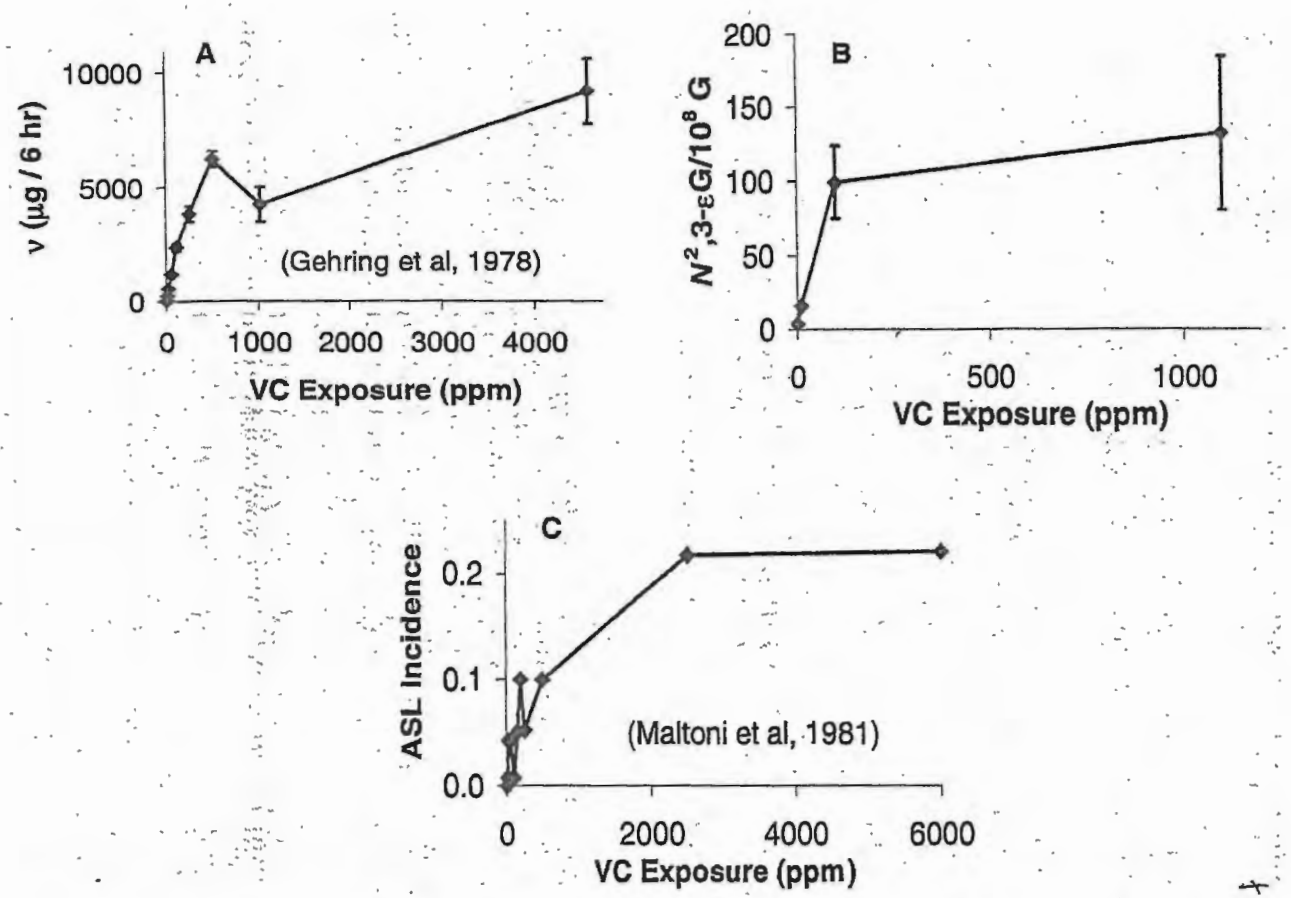


Figure 3. Graphs depicting the similar supralinear dose response relationship in A) the metabolism of VC, B) the molecular dosimetry of the DNA adducts and C) the tumor incidence in the large animal carcinogenicity study.

Table 2. Comparison of estimates of risk predicted by several different PBPK models.

Author(s)	Year	Model	Data	Cancer Risk (per $\mu g/m^3 \times 10^{-6}$)
USEPA	1994	LMS	Rat	84
			Epi	1.4
Chen and Blancato	1989	PBPK/LMS	Rat	0.7-1.4
Reitz *et al.*	1996	PBPK/LMS	Rat	0.6
			Epi	0.3-2.8
Clewell *et al.*	1995	PBPK/LMS	Mouse	1.0-2.3
			Rat	1.6-3.7
USEPA	2000	PBPK/LMS	Rat (f)	4.4

incorporation of more science into the risk assessment process has improved the accuracy in several ways. First, it readily converts animal exposure and biomarker data to human equivalents. It also provides a scientifically based method for route-to-route extrapolation. Finally, it can incorporate biomarker data such as DNA adducts to support the additional safety factor for childhood exposure. In summary, the greater use of science in the risk assessment process clearly increases the degree of confidence in the extrapolations. Several uncertainties still remain, however. No good data exist on the relationship between low exposure (<1 ppm VC) and cancer. Likewise, high quality human exposure data are not available for individuals who have developed angiosarcoma. Finally, there have not been any studies that have incorporated knowledge of endogenous adducts and their impact on risk.

The use of mass spectrometry makes possible direct comparisons between endogenous and exogenous EG adducts in the same animal by exposing the rats to $[^{13}C_2]$-VC. Those EG adducts that were induced by endogenous processes will have a mass of 354, while the adducts arising from the $[^{13}C_2]$-VC will have a mass of 356. Figure 4 shows chromatograms from liver and brain DNA from a rat exposed to 1100 ppm $[^{13}C_2]$-VC. The endogenous EG can be seen in the top panels of both liver and brain. The *m/z* 356 panels show that there is a very large peak in the liver, but no peak in the brain DNA. One of the real controversies in epidemiology for vinyl chloride has been: Does VC induce brain tumors? An increase in brain tumors has been shown in about 50% of the epidemiology studies. As the studies have gotten larger, the evidence has weakened. The most recent IARC update on the epidemiology of VC did not find a causal relationship between VC and brain tumors (IARC 2000). Figure 4 shows that while endogenous EG is present in brain, there is no exogenous EG. This provides a reasonable degree of certainty that VC is not being metabolized or being transported in an active form to the brain. It demonstrates another important utility for biomarker studies to assist in the risk assessment process, *i.e.*, testing the biological plausibility of a target site.

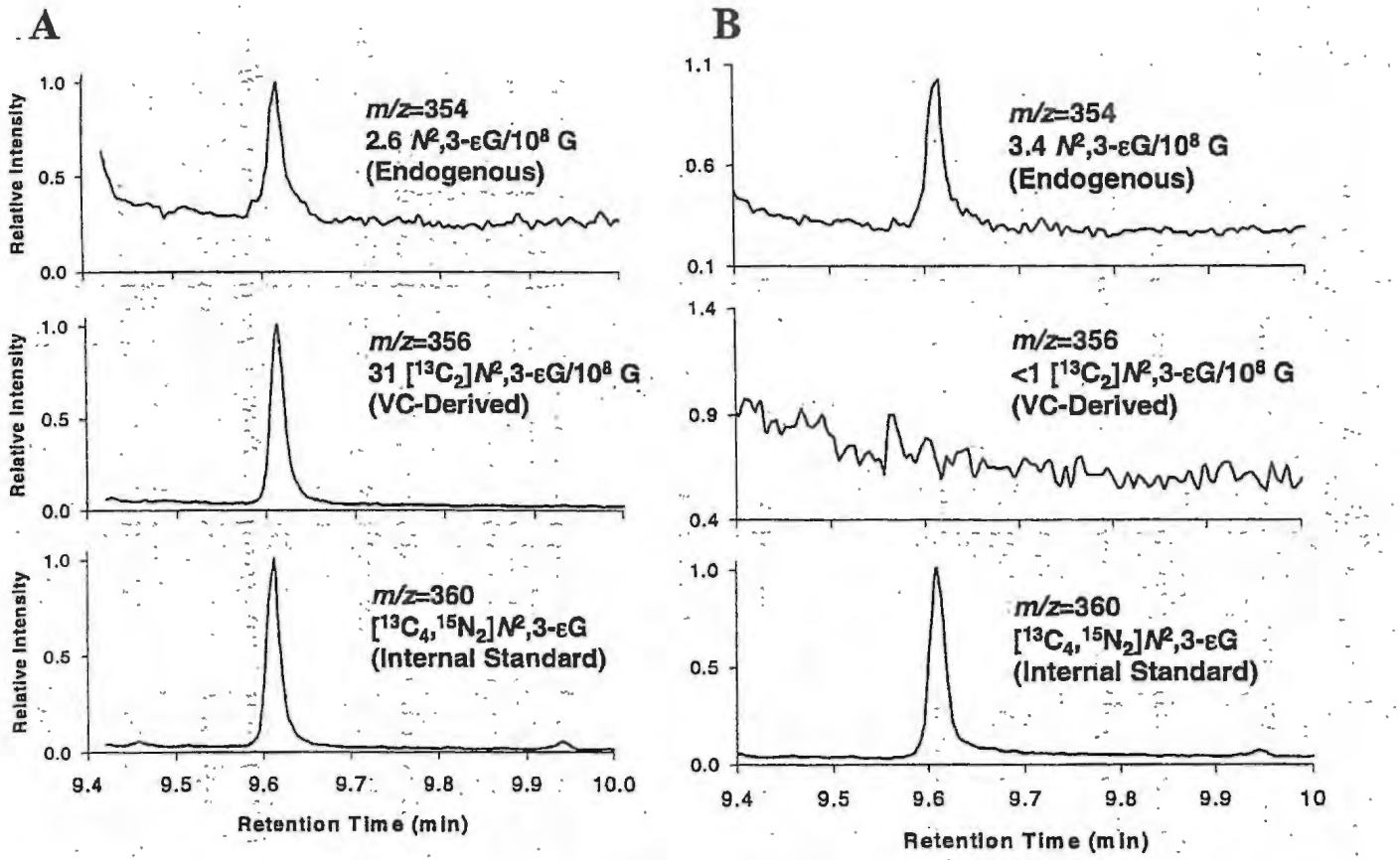


Figure 4. IA-GC/HRMS chromatograms from A) liver and B) brain DNA from a rat exposed to 1100 ppm [$^{13}C_2$]-VC for 5 days showing endogenous EG (top panel), vinyl chloride-derived EG (middle panel) and internal standard EG (lower panel).

Lipid peroxidation appears to be the major factor resulting in the formation of endogenous etheno adducts (Nair *et al.*1999). Recent studies have used [$^{13}C_{18}$]-ethyl linoleate under peroxidizing conditions to examine the source of EG adducts that are formed from oxidative stress (Ham *et al.* 2000). When this was done, it was possible to determine how many of the EG adducts formed from lipid peroxidation ([$^{13}C_2$]-EG) versus 3-phosphoglycoaldehyde, which arises from free radical attack of deoxyribose. Table 3 shows that the vast majority of the EG comes from lipid peroxidation. If a 10-fold increase in [$^{13}C_{18}$]-ethyl linoleate was present, there was an additional fivefold increase in [$^{13}C_2$]-EG. However, there was also an increase in nonlabeled EG, demonstrating that increased redox cycling can also increase the amount of EG coming from the deoxyribose. No similar increase in EG occurred when a 10-fold increase in deoxyribose was present.

Since endogenous EG is clearly present in humans, it raises the question: Why are hepatic angiosarcomas so rare in individuals that have not been exposed to VC? The incidence appears to be less than one per million (Baxter *et al.* 1980; Falk *et al.* 1981). Possible explanations include 1) that the endogenous adducts are located in non-transcribed genes and therefore do not affect the gene product; and 2) there is a sublinear relationship between number of adducts and the induction of angiosarcoma. There are no data that have demonstrated the induction of angiosarcoma where human exposure has been below 50 ppm VC. This data gap has very important implications for environmental exposures to VC, as the removal of this known human carcinogen from Superfund sites is a very expensive procedure. If a means for redirecting such clean-up funds to public health initiatives such as providing health care for children and the indigent, these funds are likely to have much greater impact on public health.

Table 3. N^2,3-Ethenoguanine formed from the reaction of [$^{13}C_{18}$]-ethyl linoleate (EtLA) with deoxyguanosine under peroxidizing conditions.

	Unlabeled EG (EG/10^6dGuo)	[$^{13}C_2$]-labeled EG (EG/10^6dGuo)
10-fold molar excess lipid	75.8 ± 33.9[a] (7%)[b]	1086 ± 518 (93%)
Equimolar lipid and dGuo	24.4 ± 2.7 (11%)	191 ± 57 (89%)
10-fold molar excess nucleoside (thymidine)	21.5 ± 5.7 (89%)	162 ± 50 (11%)
t-BuOOH only-no lipid	11.7 ± 2.1	ND[c]
Control – incubation only	0.2 ± 0.1	ND

[a] Values are expressed as mean ± standard deviation, n=3 for all samples
[b] Percent of total Gua formed
[c] Not detected

BUTADIENE

Butadiene is a very important industrial chemical that is used to manufacture rubber and many other products. It has been classified by IARC as a probable human carcinogen and by the NTP as a known human carcinogen. The primary epidemiologic data driving these classifications is the study by Delzell *et al.* (1996) on styrene-butadiene rubber workers, the SBR process. An increase in leukemia was identified in SBR workers, but no similar increases in leukemia have been found in butadiene monomer worker studies. There have been some reports of increases in lymphoma in monomer workers, however, this has not been consistent across studies.

Butadiene is also a known animal carcinogen, but there are major species differences between mice and rats (reviewed by Himmelstein *et al.* (1997)). Mice are highly susceptible to the induction of neoplasia, compared to rats, developing lung adenomas at exposures as low as 6 ppm BD. The sites of tumors also differ between mice and rats. At high exposures (625 and 1250 ppm BD), all mice developed lymphomas in one year or less (NTP, 1984). At exposures below 200 ppm BD, the site/type of tumor induction changed to hemangiomas/hemangiosarcomas, lung adenomas, mammary and harderian gland tumors (NTP 1993). Rats developed a low incidence of mammary tumors at 1000 and 8000 ppm BD (Owen *et al.* 1987), the only exposures studied.

Metabolism and Pharmacokinetics

The metabolism of butadiene has been studied in rats and mice by several different investigators and was reviewed by Himmelstein *et al.* (1997). The general metabolic scheme is shown in Figure 5. It is generally accepted that BD is first metabolized to 3,4-epoxy-1-butene (EB), a process that is primarily associated with CYP 2E1, but can also be accomplished by additional isoforms including CYP 2A6 and 4B1 (AA Elfarra, personal communication). This electrophilic metabolite can be detoxified by conjugation with glutathione and subsequent excretion in the urine as M2. It can also undergo hydrolysis by epoxide hydrolase (EH) to form butene-diol (BD-diol). BD-diol can also be conjugated with glutathione and subsequently excreted in the urine as M1. It can be further oxidized by cytochrome P450 to the epoxybutane diol (EBD). An alternative pathway for the metabolism of EB is oxidation to the diepoxybutane (DEB), which can be further hydrolyzed to EBD or conjugated by glutathione and excreted as M3. This series of epoxidation and detoxication steps generates three electrophilic metabolites: EB, DEB, and EBD. It is important to understand that quantitative measurements have only been made on the first two of these epoxides and that no measurements of EBD have been made in animals or humans exposed to BD by inhalation. EBD has been measured in *in vitro* studies with human, rat and mouse microsomes (Cheng and Ruth 1993). As will be discussed in detail below, this is important because these metabolites differ in mutagenic potency by a factor of ~200, with DEB being the most mutagenic and EBD being the least. As will be shown below, EBD appears to be the major electrophilic metabolite binding to DNA and hemoglobin, and this effect appears to be most pronounced in humans due to the high activity of EH. Accurate assessment of

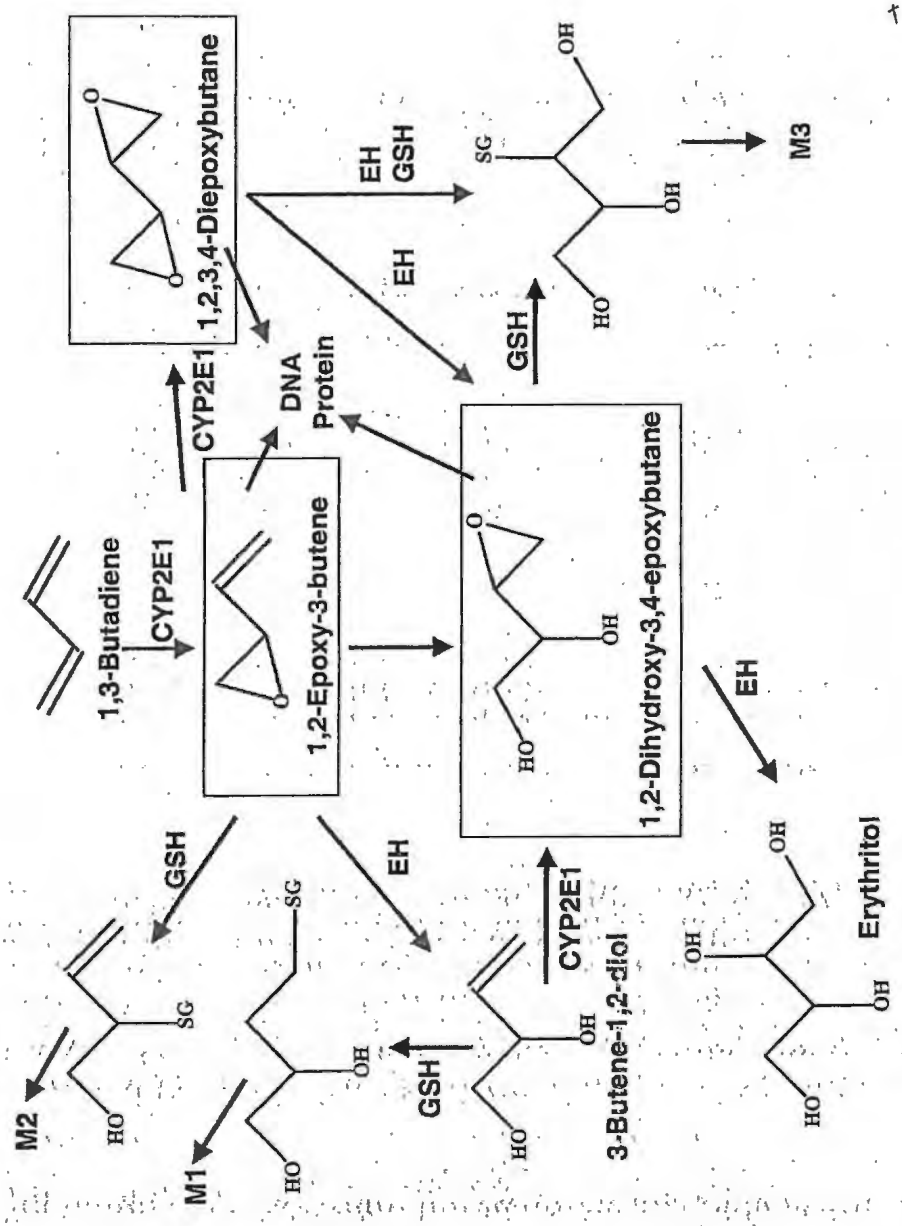


Figure 5. General scheme of the metabolism and detoxication of 1,3-butadiene.

risk clearly requires that quantitative measurements of the formation of these mutagenic epoxides be understood in rats, mice, and humans.

DNA Adducts

The identification and quantitation of DNA adducts formed by BD and its metabolites have been actively studied by several research teams. DNA adducts have been characterized at the *N*7-position of guanine (Citti *et al.* 1984; Tretyakova *et al.* 1997c; Koivisto *et al.* 1998), *N*3- position of thymidine (Selzer and Elfarra 1997) and the *N*1-, *N*3- and N^6-positions of adenine (Leuratti *et al.* 1994; Neagu *et al.* 1994; Tretyakova *et al.* 1996; Tretyakova *et al.* 1997a,b; Tretyakova *et al.* 1998; Koivisto *et al.* 1996). Multiple DNA adducts are formed at each of the above positions by the three epoxides, EB, EBD and DEB (Figure 6). This is further complicated by the presence of diastereomers if nucleosides or nucleotides are being measured. The molecular dosimetry study of Koc *et al.* (1999) compared BD DNA adducts in rats and mice across exposures ranging from 20 to 625 ppm for 4 weeks. It demonstrated that the trihydroxybutyl adducts at *N*7 of guanine (THB-G) were formed in much greater amounts than the hydroxybutenyl adducts at *N*7 of guanine (EB-G) in both rats and mice using LC-MS-MS (Figure 7). In addition, it demonstrated that the exposure response curve for THB-G was supralinear, a result of saturation of metabolic activation. In contrast, the exposure response curve for EB-G was linear. Thus, the ratio of THB-G:EB-G was greatest at low exposure and least at high exposure. The study also showed that similar numbers of adducts were present in all tissues examined, suggesting that the electrophilic metabolites circulate in the blood.

Originally, the THB-G adducts were thought to arise primarily from DEB. The finding that THB-G was more prevalent than EB-G was unexpected, especially in rats, since metabolism studies had shown that DEB was poorly formed in rats, being roughly 1/100th the amount of EB measured. In mice, similar amounts of DEB and EB were measured, yet the amount of THB-G was 25 to 45 times higher than EB-G. It was known that EB and DEB bind equally to DNA (Tretyakova *et al.* 1997c). Assuming that EBD had similar reactivity with DNA, it was possible to calculate the contribution of EBD, since data were available from published metabolism data on EB and DEB in rats and mice. These calculations demonstrated that 95 to 98% of the THB-G came from EBD (Koc *et al.* 1999). While there are still no measurements of EBD *in vivo*, the study strongly suggested that EBD was the main electrophile forming DNA adducts. Shortly thereafter, Koivisto *et al.* (1999) confirmed these findings using ^{32}P-postlabeling. They were able to differentiate THB-G adducts arising from EBD and DEB based on stereoisomers of the nucleotides and demonstrated that 98% of the THB-G adducts came from EBD. They also demonstrated that the DEB-G adduct, with one intact oxirane ring, depurinated at a rate faster than the oxirane ring hydrolyzed to EBD.

Much less information is available on the other DNA adducts of BD. Preliminary data are available on the *N*-3 adenine THB adducts and the N^6-adenine THB adducts. The *N*-3 THB adenine adducts are formed at about 10% of the THB-G adducts, but they are rapidly lost due to depurination and repair. The N^6-adenine THB adducts, which arise as *N*-1 adenine adducts, but rearrange to the N^6-adenine position, are formed at even lower amounts. No information is available on the repair of the

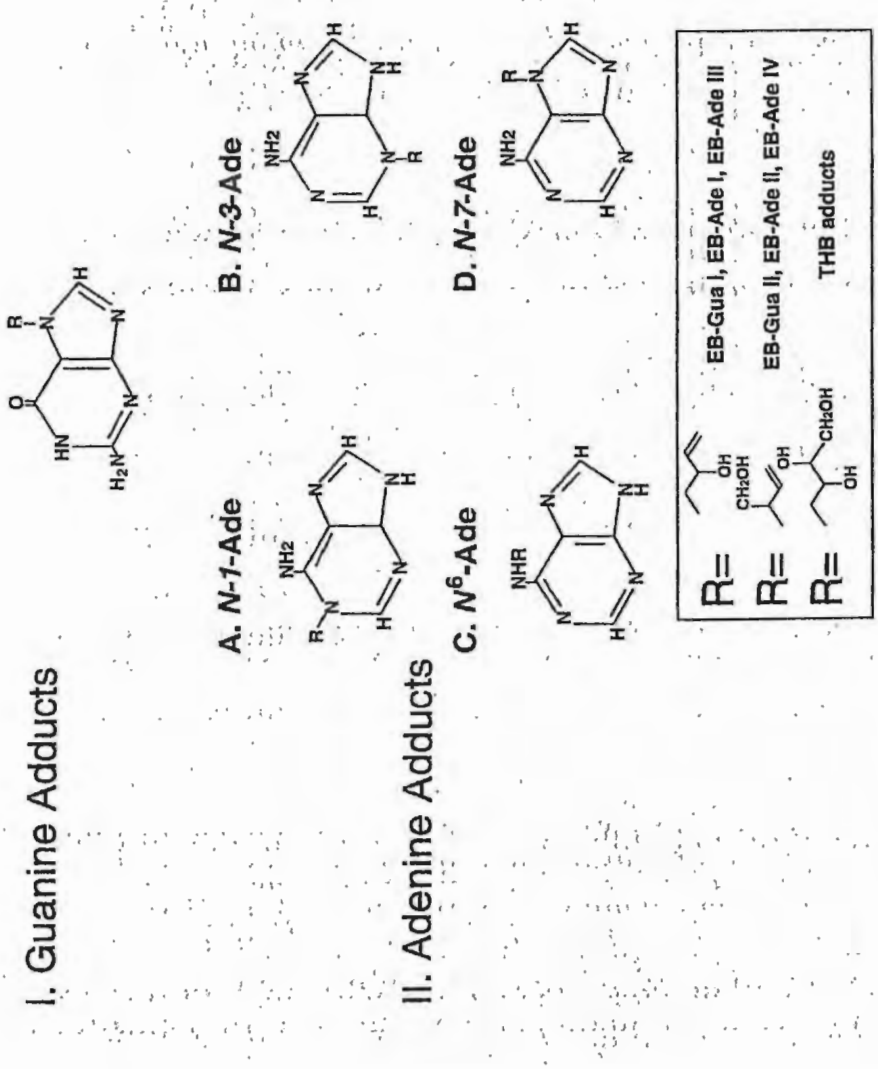


Figure 6. Structures of multiple DNA adducts are formed by the three epoxide metabolites of 1,3-butadiene, EB, EBD and DEB.

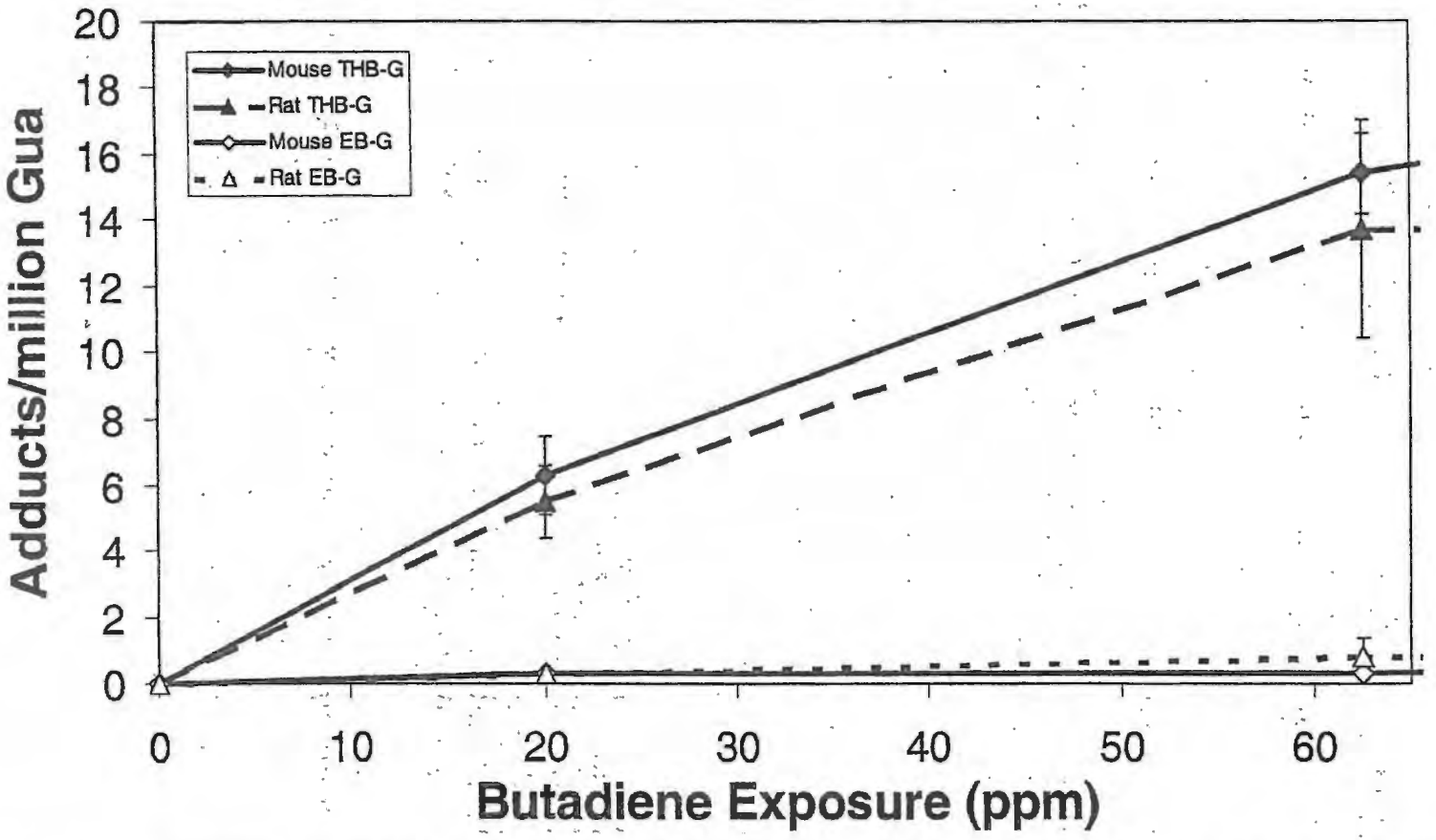


Figure 7. Dose-response curves of the butadiene DNA adducts, trihydroxybutylguanine and hydroxybutenylguanine, in female rats and mice exposed to 1,3-butadiene for 4 weeks.

N^6-adenine adducts (Carmical *et al.* 2000b). This adduct has been reported in butadiene workers at 4 per 10^9 nucleotides by Zhao *et al.* (1998), a level 1000 to 10,000 times lower than endogenous DNA adducts. There are limited studies on crosslinks induced by BD (Jelitto *et al.* 1989; Ristau *et al.* 1990; Vangala *et al.* 1993; Carmical *et al.* 2000a). Unfortunately, there are no *in vivo* quantitative data on the formation of DNA-DNA cross-links, a lesion expected to be highly mutagenic when present as an intrastrand cross-link and highly toxic when present as an interstrand cross-link. These lesions have been demonstrated in a variety of studies using DNA oligomers.

Hemoglobin Adducts

While hemoglobin adducts are not causally related to mutagenic events, they do offer an effective measure of exposure to reactive intermediates of chemicals. They have several advantages for molecular epidemiology studies including that they accumulate over the life of the red cell, which is ~43, 63 and 120 days in mice, rats and humans, respectively (Van Putten 1958). In addition, hemoglobin is more readily available than DNA in human studies. By comparing data in rodents with that from humans, it should be possible to better understand species differences and high to low dose extrapolation. Thus, these data should reduce uncertainties plaguing current risk assessments.

The BD metabolite, EB, has been shown to react with hemoglobin, forming *N*-(2-hydroxy-3-butenyl)valine (MHBVal) adducts (Osterman-Golkar *et al.* 1991). Two major and two minor peaks were identified using a modified Edman degradation and GC-MS. The two major peaks were shown to be the diastereomers resulting from attack of the *N*-terminal valine-NH_2 at C-1 of EB. Adduct concentrations of 1 to 3 pmol/g globin were recorded in humans (nonsmokers) working in a production area with ~1 ppm BD exposure levels (Osterman-Golkar *et al.* 1993). Adducts also were measured in cigarette smokers not occupationally exposed to BD. The reported adduct levels were lower in humans than in mice and rats exposed to 2 ppm BD, and were also much lower than hydroxyethylvaline adducts associated with occupational exposures to ethylene oxide and ethylene. Albrecht *et al.* (1993) reported MHBVal adducts to be five times higher in mice than in rats (17 and 3.5 nmol/g globin, respectively, at 500 ppm, 6 h/day, 5 days), although the diastereomers were not resolved. It is clear that BD exposure results in a supralinear dose-response that is characteristic of saturation of metabolic activation, and that mice have higher amounts of monoepoxide adducts than rats. In pilot studies, we compared male and female rats and mice exposed to 1000 ppm BD for 13 weeks and found that females had higher levels of MHBVal adducts than males (Tretyakova *et al.* 1996). This was confirmed in a larger study (Swenberg *et al.* 2000b) and in subsequent comparisons of rats and mice. All of the hemoglobin adduct studies have utilized the modified Edman degradation method of Törnqvist *et al.* (1986) based on GC-MS measurements using an internal standard of [d_4]-*N*-(2-hydroxyethyl)valine, [^{14}C]-*N*-(2-hydroxypropyl)valine or *N*-(2-hydroxy-3-butenyl)Val-Gly-Gly or an external standard of *N*-(2-hydroxy-3-butenyl)-[$^{13}C_5$]-valine.

A second hemoglobin adduct of butadiene that has been identified is *N*-(2,3,4-trihydroxybutyl)valine (THBVal). This adduct was initially thought to arise from DEB, with subsequent hydrolysis to the trihydroxy adduct. In view of the greater

formation of DEB by the mouse, compared to the rat, and its much greater mutagenicity, it was important that methods be developed so that both quantitative and relative comparisons of hemoglobin adducts could be made between species. Pérez *et al.* (1997) reported the formation of THBVal adducts in hemoglobin of rats and two humans and provided evidence that it was primarily formed by EBD, rather than by DEB. Furthermore, the THBVal adducts were formed in greater amounts than previous measurements of MHBVal adducts. The authors concluded that EBD appeared to be an important metabolite of BD. This issue was further explored in rodents and humans in studies by Swenberg *et al.* (2000b). Using GC/high-resolution MS, they also demonstrated that THBVal was the major hemoglobin adduct in rodents and humans. At high exposures of BD (1000 to 1250 ppm for 90 or 10 days, respectively), the ratio of THBVal:MHBVal adducts was 2 to 6 in rats and mice. In humans, only THBVal adducts could be measured using GC/high-resolution MS. This research also showed that THBVal was present in humans and several other species including rats, mice, monkeys and dogs with no known exposure to BD. The source of these endogenous THBVal adducts remains unknown.

When rats and mice were exposed to 3, 62.5, or 1250 ppm BD for 10 days, the exposure response ratio of THBVal:MHBVal (Figure 8) was very similar to that shown for DNA (Figure 7). At the lower exposures, the ratio was 39:1, while at 1250 ppm BD it was 5.7:1, demonstrating the effect of saturation of metabolic activation of the second oxidation step.

Molecular Epidemiology Studies

A molecular epidemiology study of Chinese butadiene workers that evaluated urinary metabolites, THBVal and a series of genotoxicity endpoints was conducted by Hayes *et al.* (2000). There was an excellent relationship between THBVal and exposure monitoring, and a positive relationship between exposure and urinary metabolites. In contrast, *hprt* mutations, sister chromatid exchange (SCE), aneuploidy and glycophorin A mutations showed no relationship with exposure. Kelsey *et al.* (1995) had previously shown that lymphocytes from GSTT1 null individuals had significantly higher SCEs when exposed to DEB *in vitro* than lymphocytes from individuals expressing GSTT1 and suggested that GSTT1 null individuals might constitute a susceptible worker population (Kelsey *et al.* 1995). When similar *in vitro* studies were conducted using lymphocytes of the Chinese butadiene workers, the lymphocytes from GSTT1 null individuals exhibited increased SCEs. In contrast, the SCEs and all other measures of genotoxicity, THBVal and urinary metabolites showed no difference between GSTT1 null and GSTT1 expressing workers. These results suggest that GSTT1 is involved in detoxication of BD, but that it requires high exposure for it to be a rate limiting detoxication step. High exposures, such as present in the *in vitro* studies, do not exist under the occupational environment of the Chinese BD workers.

A second large molecular epidemiology study by Albertini *et al.* (2001) was the first to examine both THBVal and MHBVal. This study had extensive industrial hygiene measurements of the workers= environment, including individual personal monitors for 60 days. Both hemoglobin biomarkers exhibited excellent exposure responses. The THBVal response is shown in Figure 9. The THBVal:MHBVal ratio was approximately 400 in these

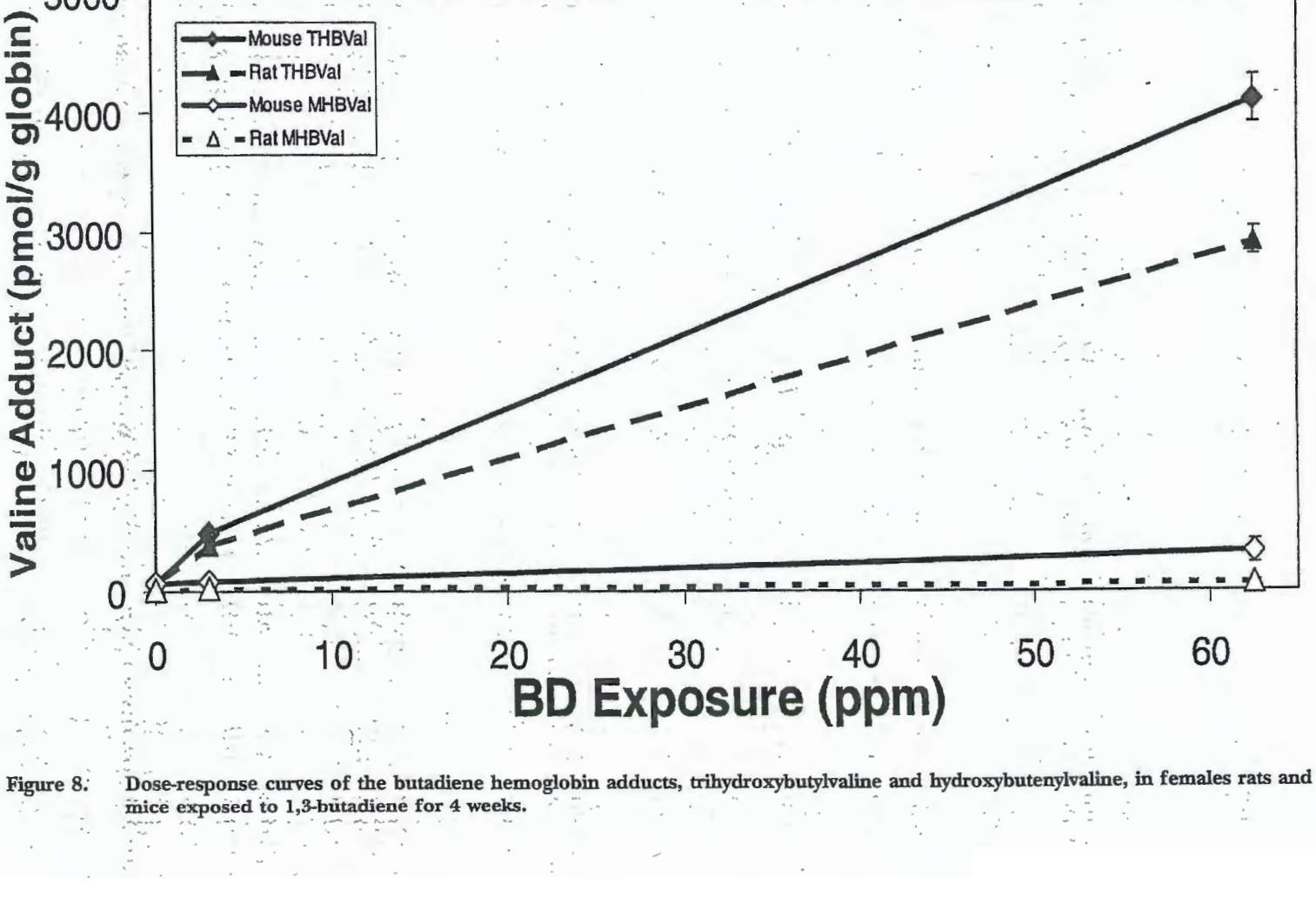


Figure 8. **Dose-response curves of the butadiene hemoglobin adducts, trihydroxybutylvaline and hydroxybutenylvaline, in females rats and mice exposed to 1,3-butadiene for 4 weeks.**

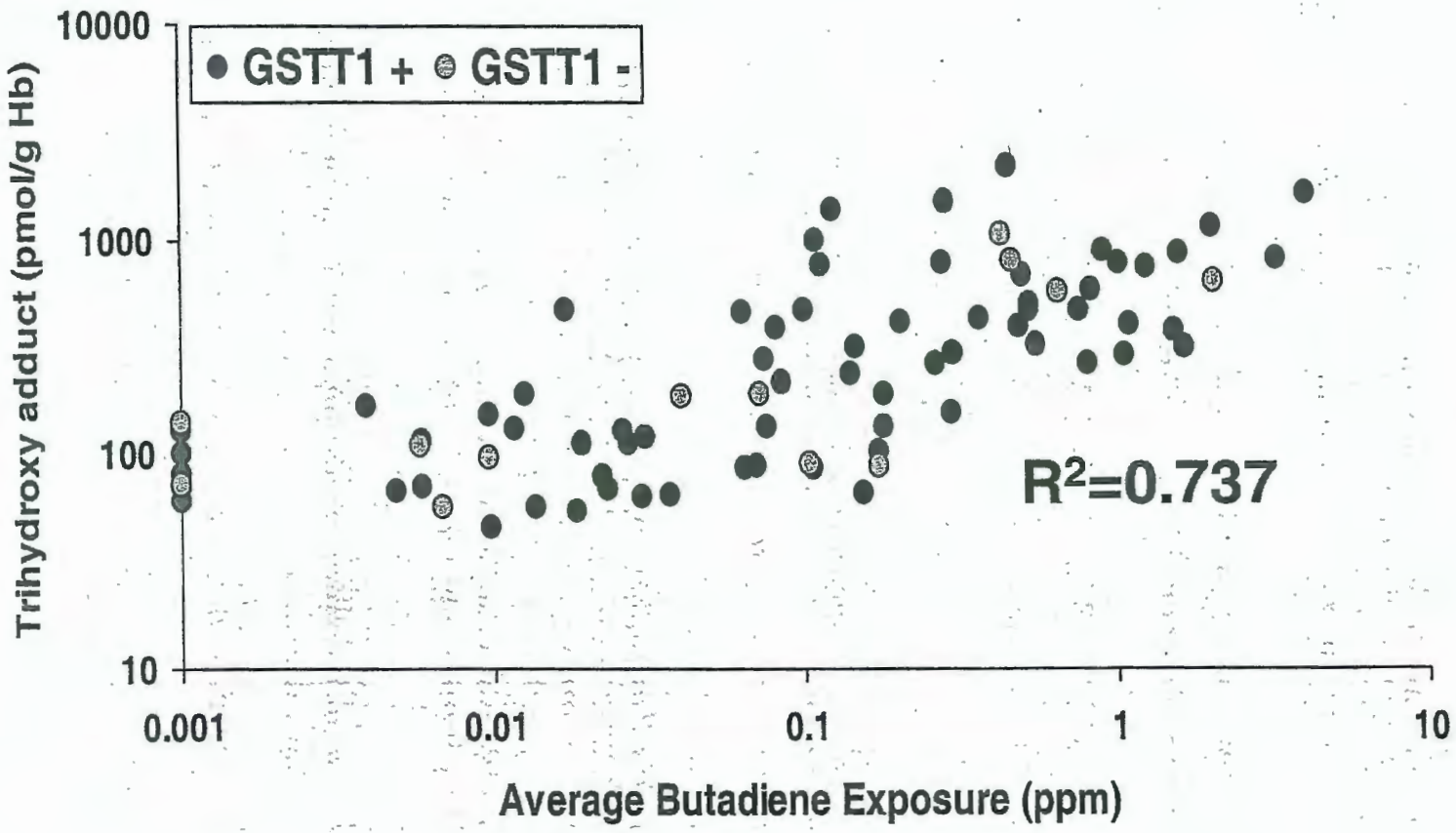


Figure 9. Trihydroxybutylvaline response in Czech workers occupationally exposed to 1,3-butadiene.

workers exposed to ~1 ppm BD. Again, biomarkers for genotoxicity were examined and no increases were associated with BD exposure, while urinary metabolites again exhibited a good exposure response relationship. When either hemoglobin adducts or any of the urinary metabolites were used as the surrogate of exposure, all genotoxicity endpoints remained negative. Likewise, GSTT1 and GSTM1 had no effect on any of the biomarkers. The data in Figure 9 can also be used to place bounds on interindividual differences in metabolism. When 95% confidence limits are placed on these data, they demonstrate that the interindividual differences approximately cover a 10-fold range.

Application of Butadiene Biomarker Data to Risk Assessment

The studies comparing the DNA and hemoglobin adducts of BD with exposure, metabolism and genotoxicity have provided a great deal of insight that is applicable to biologically-based risk assessment. First, the DNA and hemoglobin adduct data strongly support the conclusion that EBD is the major electrophile available for binding to these macromolecules. Obviously, EB is an electrophilic precursor of EBD, but most of the EB must not be accessible for binding. Metabolism, mutagenesis and carcinogenesis data support DEB as the major genotoxic and carcinogenic metabolite. Critical species differences exist in the amount of DEB that can be measured in mice and rats that parallel differences in carcinogenic response. Likewise, DEB is 100 times more mutagenic than EB and 200 times more mutagenic than EBD (Cochrane and Skopek 1994). The biomarker data from humans are consistent with these observations and suggest that EBD is even more readily formed in humans than in rats. Evidence supporting this conclusion includes the established fact that EH is the predominant detoxication pathway in humans, that THBVal:MHBVal ratios are 10-fold greater in humans than in rodents, and that no genotoxic endpoints were associated with BD exposure or biomarkers of exposure in workers even though THBVal adducts were in a similar (Rydberg *et al.* 1996) to 10-fold higher range (Albertini *et al.* 2001) to those of mice exposed to 3 ppm BD, a concentration that had measurable increases in *hprt* mutations (VE Walker, personal communication).

Biomarker studies have also provided insight into the possibility of a sensitive population associated with the GSTT1 null genotype. While it is clear that lymphocytes from GSTT1 null individuals are more sensitive for the induction of SCEs following *in vitro* exposure to DEB, there was no such increase in SCEs or other biomarkers of genotoxicity or exposure in workers exposed to 1 to 3 ppm BD. This most likely reflects high to low dose differences in detoxication, where lacking GSTT1 is a significant factor in the high dose *in vitro* experiments, but other pathways efficiently detoxify DEB and subsequent metabolites *in vivo*. The globin adduct data also demonstrate that there is roughly a 10-fold range for interindividual differences in the metabolism of BD. This study represents an excellent means for providing scientific data for this critical determinant. Another useful application of adducts in risk assessment was demonstrated by regressing data for various endpoints for genotoxicity against that individual=s biologically effective dose, thereby providing an independent mechanism for evaluation that excludes any possible confounding by inappropriate controls. This is a powerful means for evaluating the exposure response relationships for genetic toxicity endpoints that incorporates interindividual differences in exposure, metabolism and susceptibility.

Finally, any review of data for risk assessment should identify critical data gaps and needs. It will never be possible to determine past exposures of individual epidemiology studies. On the other hand, the role of DEB appears to be critical in the risk assessment of BD. It is very important that new biomarkers for DEB be developed so that quantitative comparisons can be made between rats, mice and humans. This will allow important refinements of biologically based risk assessment that will improve the accuracy of the risk assessment and make it predictive of real risk rather than protective of theoretical risks.

REFERENCES

Albertini RJ. 2001. Biomarker Response in Butadiene-Exposed Czech Workers: A Transitional Epidemiology Study. Health Effects Institute Research Report (*in press*)

Albrecht OE, Filser JG, and Neumann HG. 1993. Biological Monitoring of 1,3-Butadiene: Species Differences in Haemoglobin Binding in Rat and Mouse. In: Sorsa M, Peltonen K, Vainio H, *et al.* (eds), Butadiene and Styrene: Assessment of Health Hazards. IARC Scientific Publications, Lyons, France

Barbin A, Laib RJ, and Bartsch H. 1985. Lack of miscoding properties of 7-(2-oxoethyl)guanine, the major vinyl chloride-DNA adduct. Cancer Res 45:2440-4

Baxter PJ, Anthony PP, Macsween RN, *et al.* 1980. Angiosarcoma of the liver: annual occurrence and aetiology in Great Britain. Br J Ind Med 37:213-21

Carmical JR, Kowalczyk A, Zou Y, *et al.* 2000a. Butadiene-induced intrastrand DNA cross-links: A possible role in deletion mutagenesis. J Biol Chem 275:19482-9

Carmical JR, Nechev LV, Harris CM, *et al.* 2000b. Mutagenic potential of adenine N^6 adducts of monoepoxide and diolepoxide derivatives of butadiene. Environ Mol Mutagen 35:48-56

Chen CW and Blancato JN. 1989. Incorporation of biological information in cancer risk assessment: example—vinyl chloride. Cell Biol Toxicol 5:417-44 Cheng KC, Preston BD, Cahill DS, *et al.* 1991. The vinyl chloride DNA derivative N^2,3-ethenoguanine produces G->A transitions in *Escherichia coli.* Proc Natl Acad Sci USA 88:9974-8

Cheng XQ and Ruth JA. 1993. A simplified methodology for quantitation of butadiene metabolites: Application to the study of 1,3-butadiene metabolism by rat liver microsomes. Drug Metab Dispos 21:121-4

Citti L, Gervasi PG, Turchi G, *et al.* 1984. The reaction of 3,4-epoxy-1-butene with deoxyguanosine and DNA *in vitro*: Synthesis and characterization of the main adducts. Carcinogenesis 5:47-52

Clewell HJ, Gentry PR, Gearhart JM, *et al.* 1995. Considering pharmacokinetic and mechanistic information in cancer risk assessments for environmental contaminants: examples with vinyl chloride and trichloroethylene. Chemosphere 31:2561-78

Cochrane JE and Skopek TR. 1994. Mutagenicity of butadiene and its epoxide metabolites: I. Mutagenic potential of 1,2-epoxybutene, 1,2,3,4-diepoxybutane and 3,4-epoxy-1,2-butanediol in cultured human lymphoblasts. Carcinogenesis 15:713-7

Delzell E, Sathiakumar N, Hovinga M, *et al.* 1996. A follow-up study of synthetic rubber workers. Toxicology 113:182-9

El-Ghissassi F, Barbin A, Nair J, *et al.* 1995. Formation of 1,N^6-ethenoadenine and 3,N^4-ethenocytosine by lipid peroxidation products and nucleic acid bases. Chem Res Toxicol 8:278-283Falk H, Herbert J, Crowley S, *et al.* 1981. Epidemiology of hepatic angiosarcoma in the United States: 1964-1974. Environ Health Perspect 41:107-13

Fedtke N, Walker VE, and Swenberg JA. 1989. Determination of 7-(2-oxoethyl)guanine and N^2,3-ethenoguanine in DNA hydrolysates by HPLC. Arch Toxicol Suppl 13:214-8

Fedtke N, Boucheron JA, Turner MJ, *et al.* 1990a. Vinyl chloride-induced DNA adducts. I: Quantitative determination of N^2,3-ethenoguanine based on electrophore labeling. Carcinogenesis 11:1279-85

Fedtke N, Boucheron JA, Walker VE, *et al.* 1990b. Vinyl chloride-induced DNA adducts. II: Formation and persistence of 7-(2'-oxoethyl)guanine and N^2,3-ethenoguanine in rat tissue DNA. Carcinogenesis 11:1287-92

Fernando RC, Nair J, Barbin A, *et al.* 1996. Detection of 1,N^6-ethenodeoxyadenosine and 3,N^4-ethenodeoxycytidine by immunoaffinity/P^{32}-postlabelling in liver and lung DNA of mice treated with ethyl carbamate (urethane) or its metabolites. Carcinogenesis 17:1711-8

Gehring PJ, Watanabe PG, and Park CN. 1978. Resolution of dose-response toxicity data for chemicals requiring metabolic activation: Example — vinyl chloride. Toxicol Appl Pharmacol 44:581-91

Guengerich FP. 1992. Roles of the vinyl chloride oxidation products 2-chlorooxirane and 2-chloroacetaldehyde in the in vitro formation of etheno adducts of nucleic acid bases. Chem Res Toxicol 5:2-5

Guengerich FP, Kim DH, and Iwasaki M. 1991. Role of human cytochrome P-450 IIE1 in the oxidation of many low molecular weight cancer suspects. Chem Res Toxicol 4:168-79

Guichard Y, El-Ghissassi F, Nair J, *et al.* 1996. Formation and accumulation of DNA ethenobases in adult Sprague-Dawley rats exposed to vinyl chloride. Carcinogenesis 17:1553-950

Ham AJL, Ranasinghe A, Morinello EJ, *et al.* 1999. Immunoaffinity/gas chromatography/high resolution mass spectrometry method for the detection of N^2,3-ethenoguanine. Chem Res Toxicol 12:1240-6

Ham AJL, Ranasinghe A, Koc H, *et al.* 2000. 4-Hydroxy-2-nonenal and ethyl linoleate form N^2,3-ethenoguanine under peroxidizing conditions. Chem Res Toxicol 13:1243-

Hayes RB, Zhang L, Yin S, *et al.* 2000. Genotoxic markers among butadiene polymer workers in China. Carcinogenesis 21:55-62

Himmelstein MW, Acquavella JF, Recio L, *et al.* 1997. Toxicology and epidemiology of 1,3-butadiene. Crit Rev Toxicol 27:1-108

IARC (International Agency for Research on Cancer). 1987. Overall evaluations of carcinogenicity: an updating of IARC Monographs volumes 1 to 42. IARC Monogr Eval Carcinog Risks Hum S7:373-6

IARC (International Agency for Research on Cancer). 2000. Update of the follow-up of mortality and cancer incidence among European workers employed in the vinyl chloride industry. Report number 00-001. Lyon, FranceJelitto B, Vangala RR, and Laib RJ. 1989. Species differences in DNA damage by butadiene: role of diepoxybutane. Arch Toxicol Suppl 13:246-9

Kelsey KT, Wiencke JK, Ward J, *et al.* 1995. Sister-chromatid exchanges, glutathione S-transferase theta deletion and cytogenetic sensitivity to diepoxybutane in lymphocytes from butadiene monomer production workers. Mutat Res 335:267-73

Koc H, Tretyakova NY, Walker VE, *et al.* 1999. Molecular dosimetry of N-7 guanine adduct formation in mice and rats exposed to 1,3-butadiene. Chem Res Toxicol 12:566-74

Koivisto P, Adler ID, Sorsa M, *et al.* 1996. Inhalation exposure of rats and mice to 1,3-butadiene induces N^6-adenine adducts of epoxybutene detected by P32-postlabeling and HPLC. Environ Health Perspect 104:655-7

Koivisto P, Adler ID, Pacchierotti F, *et al.* 1998. DNA adducts in mouse testis and lung after inhalation exposure to 1,3-butadiene. Mutat Res 397:3-10

Koivisto P, Kilpelainen I, Rasanen I, *et al.* 1999. Butadiene diolepoxide- and diepoxybutane-derived DNA adducts at N7-guanine: a high occurrence of diolepoxide-derived adducts in mouse lung after 1,3-butadiene exposure. Carcinogenesis 20:1253-9

Leuratti C, Jones NJ, Marafante E, *et al.* 1994. DNA damage induced by the environmental carcinogen butadiene: Identification of a diepoxybutane-adenine adduct and its detection by ^{32}P-postlabelling. Carcinogenesis 15:1903-10

Maltoni C, Lefemine G, Ciliberti A, *et al.* 1981. Carcinogenicity bioassays of vinyl chloride monomer: a model of risk assessment on an experimental basis. Environ Health Perspect 41:3-29

Morinello EJ, Ham AJL, Ranasinghe A, *et al.* 2000. Simultaneous quantitation of N^2,3-ethenoguanine and 1,N^2-ethenoguanine with immunoaffinity/gas chromatography/high resolution mass spectrometry. Chem Res Toxicol (*in press*)

Mroczkowska MM and Kusmierek JT. 1991. Miscoding potential of N^2,3-ethenoguanine studied in an *Escherichia coli* DNA-dependent RNA polymerase *in vitro* system and possible role of this adduct in vinyl chloride-induced mutagenesis. Mutagenesis 6:385-90

Nair J, Barbin A, Guichard Y, *et al.* 1995. 1,N^6-ethenodeoxyadenosine and 3,N^4-ethenodeoxycytidine in liver DNA from humans and untreated rodents detected by immunoaffinity/^{32}P-postlabelling. Carcinogenesis 16:613-7

Nair J, Barbin A, Velic I, *et al.* 1999. Etheno DNA-base adducts from endogenous reactive species. Mutat Res 424:59-69

Neagu I, Koivisto P, Neagu C, *et al.* 1994. Alkylation of Purine Bases with 3,4-epoxy-1-butene, A Reactive Metabolite of 1,3-Butadiene. In: Hemminki K, Dipple A, Shuker DEG, *et al.* (eds), DNA Adducts: Identification and Biological Significance. IARC Scientific Publications, Lyons, France

NTP (National Toxicology Program). 1984. Toxicology and Carcinogenesis Studies of 1,3-Butadiene in B6C3F1Mice. NTP TR 288. NIH Pub. No. 84-2544. U.S. Department of Health and Human Services, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA NTP (National Toxicology Program). 1993. Technical Report on the Toxicology and Carcinogenesis Study of 1,3-Butadiene. NTP TR 433. U.S. Department of Health and Human Services, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA

Osterman-Golkar S, Bond JA, Ward JB, *et al.* 1993. Use of Haemoglobin Adducts for Biomonitoring Exposure to 1,3-butadiene. In: Sorsa M, Peltonen K, Vainio H, *et al.* (eds), Butadiene and Styrene: Assessment of Health Hazards. IARC Scientific Publications, Lyons, France

Osterman-Golkar S, Kautiainen A, Bergmark E, *et al.* 1991. Hemoglobin adducts and urinary mercapturic acids in rats as biological indicators of butadiene exposure. Chem Biol Interact 80:291-302

Owen PE, Glaister JR. 1990 Inhalation toxicity and carcinogenicity of 1,3-butadiene in Sprague- Dawley rats. Environ Health Perspect 86:19-25

Owen PE, Glaister JR, Gaunt IF, *et al.* 1987 Inhalation toxicity studies with 1,3-butadiene. 3. Two year toxicity/carcinogenicity study in rats. Am Ind Hyg Assoc J 48:407-13

Pérez HL, Lähdetie J, Landin HH, *et al.* 1997. Haemoglobin adducts of epoxybutanediol from exposure to 1,3-butadiene or butadiene epoxides. Chem Biol Interact 105:181-98

Reitz RH, Gargas ML, Andersen ME, *et al.* 1996. Predicting cancer risk from vinyl chloride exposure with a physiologically based pharmacokinetic model. Toxicol Appl Pharmacol 137:253-67

Ristau C, Deutschmann S, Laib RJ, *et al.* 1990. Detection of diepoxybutane-induced DNA-DNA crosslinks by cesium trifluoracetate (CsTFA) density-gradient centrifugation. Arch Toxicol 64:343-4

Rydberg P, Magnusson AL, Zorcec V, *et al.* 1996. Adducts to N-terminal valines in hemoglobin from butadiene metabolites. Chem Biol Interact 101:193-205

Selzer RR, Elfarra AA. 1997 Characterization of four N-3-thymidine adducts formed *in vitro* by the reaction of thymidine and butadiene monoxide. Carcinogenesis 18:1993-8

Singer B, Kusmierek JT, Folkman W, *et al.* 1991. Evidence for the mutagenic potential of the vinyl chloride induced adduct, N^2,3-etheno-deoxyguanosine, using a site-directed kinetic assay. Carcinogenesis 12:745-7

Swenberg JA, Ham AJ, Koc H, *et al.* 2000a. DNA adducts: effects of low exposure to ethylene oxide, vinyl chloride and butadiene. Mutat Res 464:77-86

Swenberg JA, Christova-Gueorguieva NI, Upton PB, *et al.* 2000b. 1,3-Butadiene: Cancer, Mutations, and Adducts. Part V. Hemoglobin Adducts as Biomarkers of Butadiene Exposure and Metabolism. Report 92, pp 191-209. Health Effects Institute, Cambridge, MA, USA

Swenberg JA, La DK, Scheller NA, *et al.* 1995. Dose-response relationships for carcinogens. Toxicol Lett 82-83:751-6

Törnqvist M, Mowrer J, Jensen S, *et al.* 1986. Monitoring of environmental cancer initiators through hemoglobin adducts by a modified Edman degradation method. Anal Biochem 154:255-66

Tretyakova NY, Lin YP, Upton PB, *et al.* 1996. Macromolecular adducts of butadiene. Toxicology 113:70-6

Tretyakova N, Lin Y, Sangaiah R, *et al.* 1997a. Identification and quantitation of DNA adducts from calf thymus DNA exposed to 3,4-epoxy-1-butene. Carcinogenesis 18:137-47

Tretyakova N, Sangaiah R, Yen TY, *et al.* 1997b. Adenine adducts with diepoxybutane: isolation and analysis in exposed calf thymus DNA. Chem Res Toxicol 10:1171-9

Tretyakova NY, Sangaiah R, Yen TY, *et al.* 1997c. Synthesis, characterization, and *in vitro* quantitation of N7-guanine adducts of diepoxybutane. Chem Res Toxicol 10:779-85

Tretyakova NY, Chiang SY, Walker VE, *et al.* 1998. Quantitative analysis of 1,3-butadiene-induced DNA adducts in vivo and in vitro using liquid chromatography electrospray ionization tandem mass spectrometry. J Mass Spectrom 33:363-76

USEPA (U.S. Environmental Protection Agency). 2000. Toxicological Review of Vinyl Chloride in Support of Summary Information on the Integrated Risk Information System (IRIS). EPA/635R-00/004. Available at http://www.epa.gov/iris

USEPA (U.S. Environmental Protection Agency). 1994. Chemical Assessments and Related Activities. OHEA-I-127. Washington, DC, USAVan Putten LM. 1958. The life span of red cells in the rat and the mouse as determined by labeling with DFP^{32} *in vivo*. Blood 13:789-94

Vangala RR, Laib RJ, and Bolt HM. 1993. Evaluation of DNA damage by alkaline elution technique after inhalation exposure of rats and mice to 1,3-butadiene. Arch Toxicol 67:34-8

Watanabe PG and Gehring PJ. 1976. Dose-dependent fate of vinyl chloride and its possible relationship to oncogenicity in rats. Environ Health Perspect 17:145-52

Zhao C, Koskinen M, and Hemminki K. 1998. ^{32}P-postlabelling of N^6-adenine adducts of epoxybutanediol *in vivo* after 1,3-butadiene exposure. Toxicol Lett 102-103:591-4

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1339-1353 (2002)

Toxicogenomics and Human Disease Risk Assessment

Kevin T. Morgan,[1]* H. Roger Brown, Gina Benavides, Lynn Crosby,[2] Dirk Sprenger, Lawrence Yoon, Hong Ni, Marilyn Easton, Duncan Morgan,[3] Daniel Laskowitz, and Ronald Tyler[1]

[1]GlaxoSmithKline, Inc., RTP, NC. [2]UNC Curriculum in Toxicology, Chapel Hill, NC/US EPA, RTP, NC. [3]Duke University Medical School, Durham, NC 27599-7400

ABSTRACT

Complete sequencing of human and other genomes, availability of large-scale gene expression arrays with ever-increasing numbers of genes displayed, and steady improvements in protein expression technology can have a great impact on the field of toxicology. However, we are a long way from devising effective standards for human risk assessments based upon these technologies. Current impediments to effective application of these technologies include appropriate normalization procedures (as "there is no fixed point in transcript space"), confirmation of data quality and demonstration of the functional significance of responses observed. Providing risk assessors with statistically and functionally unconfirmed, large-scale gene expression data sets that generally defy interpretation is not an appropriate approach. We propose that a logical process of data generation be developed, with risk assessment in mind from the outset. The basic principles of toxicology should be applied to selection of experimental systems, dose and duration of exposure, along with appropriate statistical analyses and biological interpretation. If mechanistically based interspecies extrapolation of risk is to be undertaken, suitable biochemical or other follow-up studies should be completed to confirm functional significance of transcriptional changes.

Key Words: Transcriptome, toxicogenomics, oxidative stress, risk assessment, *in vitro* toxicology, mechanisms of disease, bioinformatics.

INTRODUCTION

The statement that "new pieces of technology commonly give rise to information overload..." (Nicholls 1999) is clearly true for the revolution that is taking place in

* Corresponding author: Kevin T. Morgan, GlaxoSmithKline, Inc., MAI-T1154, 5 Moore Drive, Research Triangle Park, NC 27709; Tel(voice): 919-483-1178, Tel(fax): 919-315-8391; ktm42807@gsk.com

1080-7039/02/$.50

technological developments in molecular biology. A relatively small gene expression data set, such as that published on the response of pancreatic β-cells to high glucose levels (Webb *et al.* 2000), sends one searching through texts on intermediary metabolism in order to put these data in proper perspective. Considerable 'on line' assistance may be obtained with interpretation, if one can negotiate links to the growing number of databases, such KEGG or EPD, available on the Internet. An understanding of these pathways and regulatory circuits is essential for mechanistic interpretation of such data, and will be critical for assessment of human risks (Klaassen 1996) based upon them.

The scientific and popular press has been replete with promises concerning the 'genomic revolution.' Such promises have resulted in a backlash of skepticism, as exemplified by the following quote from an extensive review of this subject in 1999 (Cole *et al.* 1999): "The scientific literature contains more reviews about arrays than primary research papers applying them." The field has since moved from promises to increasingly solid data sets, many of which are available on the Internet (Iyer *et al.* 1999; Lee *et al.* 2000; Wang *et al.* 1999; Webb *et al.* 2000; Wen *et al.* 1998). There are also an increasing number of tools with which to explore gene expression responses to disease states, including chemical toxicity. Such tools include a range of transcript (mRNA) (Collins 1999) and protein expression platforms, data analysis software, and 'laboratories on a chip.' We shall make no attempt to review these tools, most of which are readily found on the Internet along with extensive descriptions on their respective web sites. When selecting a system, however, it is critical to consider data quality, bioinformatics support, and cost. These new, and much touted (Farr and Dunn 1999; Nuwaysir *et al.* 1999), technologies have become an integral component of the evolving discipline of toxicogenomics.

TOXICOGENOMICS

Toxicogenomics is probably developing more rapidly than any other area in safety assessment during the last 40 years. The following discussion is intended to provide some insight into the potential utility of gene expression analysis in toxicology, and more specifically, with respect to pharmaceuticals. These technologies are applicable equally to chemical risk assessments associated with environmental exposure. The field of toxicogenomics, which focuses largely on transcript data, includes or touches upon many areas of research, especially pharmacogenetics, proteomics, and biochemical toxicology. This discussion of toxicogenomics is biased towards the authors' direct experience with responses of the transcriptome to toxic insults, and we recognize that it will be somewhat incomplete and speculative. As we learn more about gene expression, new aspects of application will be recognized and some thought to be currently feasible may not work out.

In general terms, toxicity is an adverse alteration of morphology or function. Toxic responses that occur within a few minutes are not likely to be caused by or have an impact on gene expression. On the other hand, toxic responses that occur over one or more hours are likely to have some impact on gene expression, or to be a result of changes in gene expression, or both. Some of the gene expression changes that occur as toxicity develops are expected to be unique to

the mechanism of toxicity (*e.g.*, free radical production, inhibition of cellular respiration). These will be referred to as mechanism associated gene expression changes. Other gene expression changes are expected to be unique to specific types of toxicity (*e.g.*, apoptosis, oncosis, and nongenotoxic [epigenetic] oncogenesis) but common amongst mechanisms that cause the same type of toxicity. These gene expression changes will be referred to as toxicity-type associated gene changes. Other gene expression changes are expected to be adaptive, in response to changes in such things as blood pressure, local blood perfusion rates, and the nutrient environment of the target cell population. These gene expression changes will be referred to as adaptive gene expression changes. Determination of which patterns of gene expression are unique to a mechanism of toxicity, which are unique to a type of toxicity (but common amongst mechanisms that cause that type of toxicity), and which are adaptive, will allow development of gene expression-based toxicity screens, diagnostic assays, and surrogate markers.

Hopefully a manageable number of critical genes can be identified for each mechanism and type of toxicity. This would make development of toxicology screens, diagnostic assays, and surrogates much easier than if the pattern of change for a large number of genes, *e.g.*, thousands, must be analyzed to recognize the mechanism and type of toxicity. If only a few critical genes can be identified for each mechanism and type of toxicity, transgenic reporter systems can be developed for *in vivo* and *in vitro* (cell culture) studies. This will allow high throughput rapid analysis. It is generally recommended, however, that direct measurement of the respective cell function (*e.g.*, redox state) would be preferred to indirect indicators, such as transcriptional changes.

If critical "core" genes, responsive to specific mechanisms of toxicity and the various types of injury exist, then they have probably been during evolution and are expressed in many tissues. Each mechanism or type of toxicity would then be expected to result in a unique set of transcriptional, and consequently translational, responses that will be shared amongst animal tissues and species. For example, a core set of transcriptional events occurs in homologous genes in association with oxidative stress (Scandalios 1997), apoptosis (Jehn and Osborne 1997) in humans, rats, and even *Caenorhabitis elegans*, regardless of the tissue. On the other hand, each species and tissue will likely have subsequent adaptive responses that are, in certain details at least, unique to the species/tissue. As a result, interspecies comparisons may be useful in differentiating critical unique core gene expression changes from nonspecific adaptive changes.

To date, toxicity screens have been essentially confined to low throughput *in vivo* assays and moderate to high-throughput *in vitro* assays that measure selected indicators of defective cell function or death. Gene expression changes, developed for a few critical genes that reliably predict or identify established mechanisms of toxicity, would be readily amenable to the development of rapid high throughput toxicity screens, diagnostic assays, and toxicity surrogates. These screens and diagnostic assays, which could be applied to human subjects (*e.g.*, blood, skin and hair samples) have considerable potential for determination of dose and kinetic responses, interspecies extrapolation, and mechanistically based risk assessments for toxic chemicals.

OXIDATIVE STRESS AS A TEST CASE FOR TRANSCRIPTOMICS

Our research group (Toxicogenomics-Mechanisms, GlaxoSmithKline, Inc. Research Triangle Park, NC) was created for the task of assessing the value of gene expression array technology as a tool for toxicologists studying potential drug candidates. The work briefly referred to in this article was derived from studies carried out (see acknowledgements) using a combination of Clontech® Human Stress Arrays and a commercial Real Time Polymerization Chain Reaction (RT/PCR) platform (TaqMan™) for confirmation of array data. A clear example of the ability of these arrays to detect almost complete absence of expression of the ApoE gene in ApoE knockout mice is shown in Figure 1. The research was directed toward applying transcriptomics to the more efficient detection of oxidative stress, as a test case for this technology. Oxidative stress is a component, and potential cause of, many disease states (Armstrong 1998), including infection with the Human Immunodeficiency Virus (HIV), diabetes mellitus, certain idiosyncratic drug reactions, and chemical toxicity. Oxidative stress affects us all, as it appears to play a major role in aging as a consequence of cumulative damage to our DNA and other critical macromolecules (Guyton *et al.* 1997; Haffner 2000; Scandalios 1997). Mild oxidative stress results from 'leakage' of reactive oxygen species (ROS) during normal respiration in mitochondria, and thus is tightly coupled with bioenergetics.

A number of antioxidant systems have developed to counter the threat of oxidant damage associated with respiration and other sources of ROS (such as xenobiotic metabolism and peroxisomal function), including transcriptional responses, resulting in changing patterns of gene expression. The latter systems, which have been reviewed in detail (Scandalios 1997), play a critical role in the maintenance of the correct reduction-oxidation (redox) balance in cells and tissues. One key molecule in the maintenance of redox balance is glutathione (Anderson 1998), for which the maintenance of the appropriate balance of the reduced (GSH) *versus* the oxidized (GSSG) states is critical to cell function and survival. Determination of the intracellular GSH:GSSG ratio provides one of many means of assessing cellular redox state and detecting oxidative stress (Armstrong 1998).

In oxidative stress, redox balance and bioenergetics come together to provide an interesting test case (*vide infra*) for its detection using transcriptomics. The relationship between energy dependent synthesis of GSH and its energy dependent maintenance in the correct redox state is shown in an extremely simplified form in Figure 2. For more detail, see reviews listed in the bibliography (Armstrong 1998; Gille and Sigler 1995; Guyton *et al.* 1997; Haffner 2000; Saran *et al.* 1998; Scandalios 1997). The use of oxygen to oxidize fuel has associated risks that persist throughout life and contribute to age-related diseases, such as cataracts. The presence of antioxidant gene responses can be used to detect the induction of this stress in toxicological studies. We carried out investigations in rat mesothelial cells using Clontech Gene Expression Arrays™ (Figure 3). These arrays revealed gene expression changes consistent with the oxidative nature of this compound (Crosby *et al.* 2000b). Related studies were undertaken using a larger number of compounds, both oxidative stressors and nonoxidative stressors, and a human cell line (HepG2) and Clontech Human Stress Arrays combined with RT/PCR (Morgan *et al.* 2002). From this work a set of seven genes was selected for TaqMan™ analysis, and used successfully to

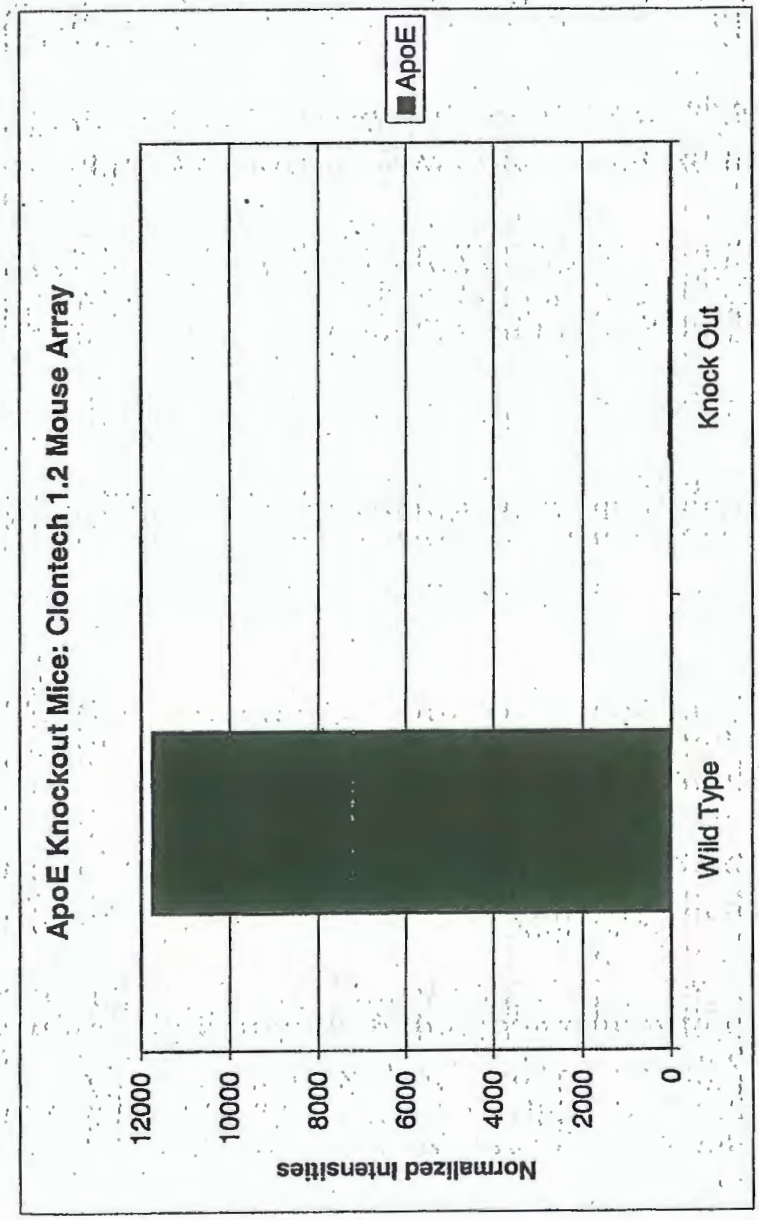


Figure 1. Comparison of wild type *versus* ApoE knockout mice on a 'relatively' inexpensive gene expression platform (Clontech, Nylon Arrays). These samples from the cerebral cortex were hybridized to the Mouse 1.2 Atlas array (~1200 genes) and the strongest signal difference by far was the ApoE gene, which was clearly knocked out. Not all data are this clear!

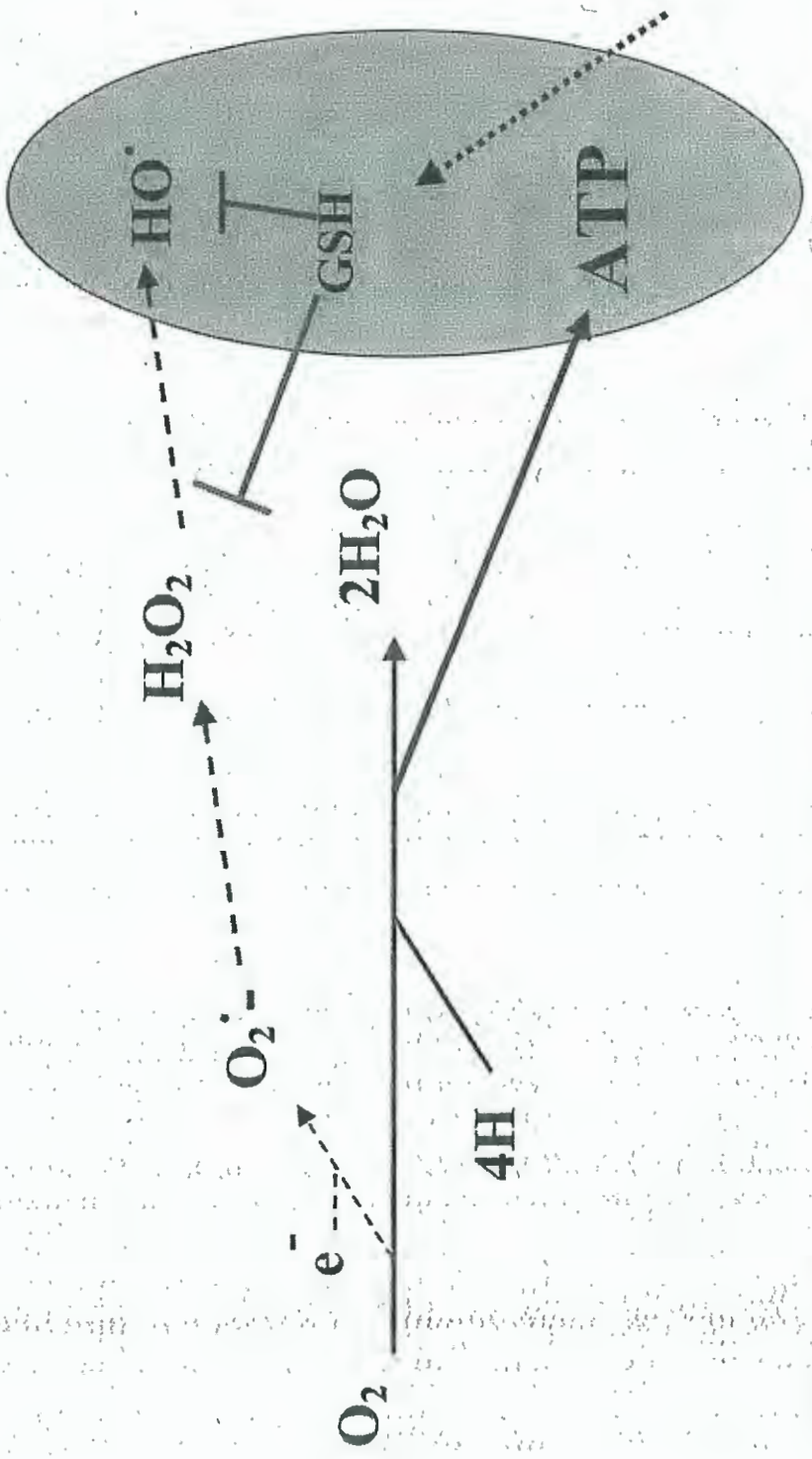


Figure 2. **A brief recapitulation of the relationships between energy generation systems (ATP from mitochondrial oxidative phosphorylation, as key example), ROS leakage, and maintenance of redox balance through the energy-dependent generation of reduced glutathione (GSH).** The oxygen leakage as ROS ranges from 0.5 to 1.5% of oxygen consumed. These systems are in a state of dynamic balance throughout life.

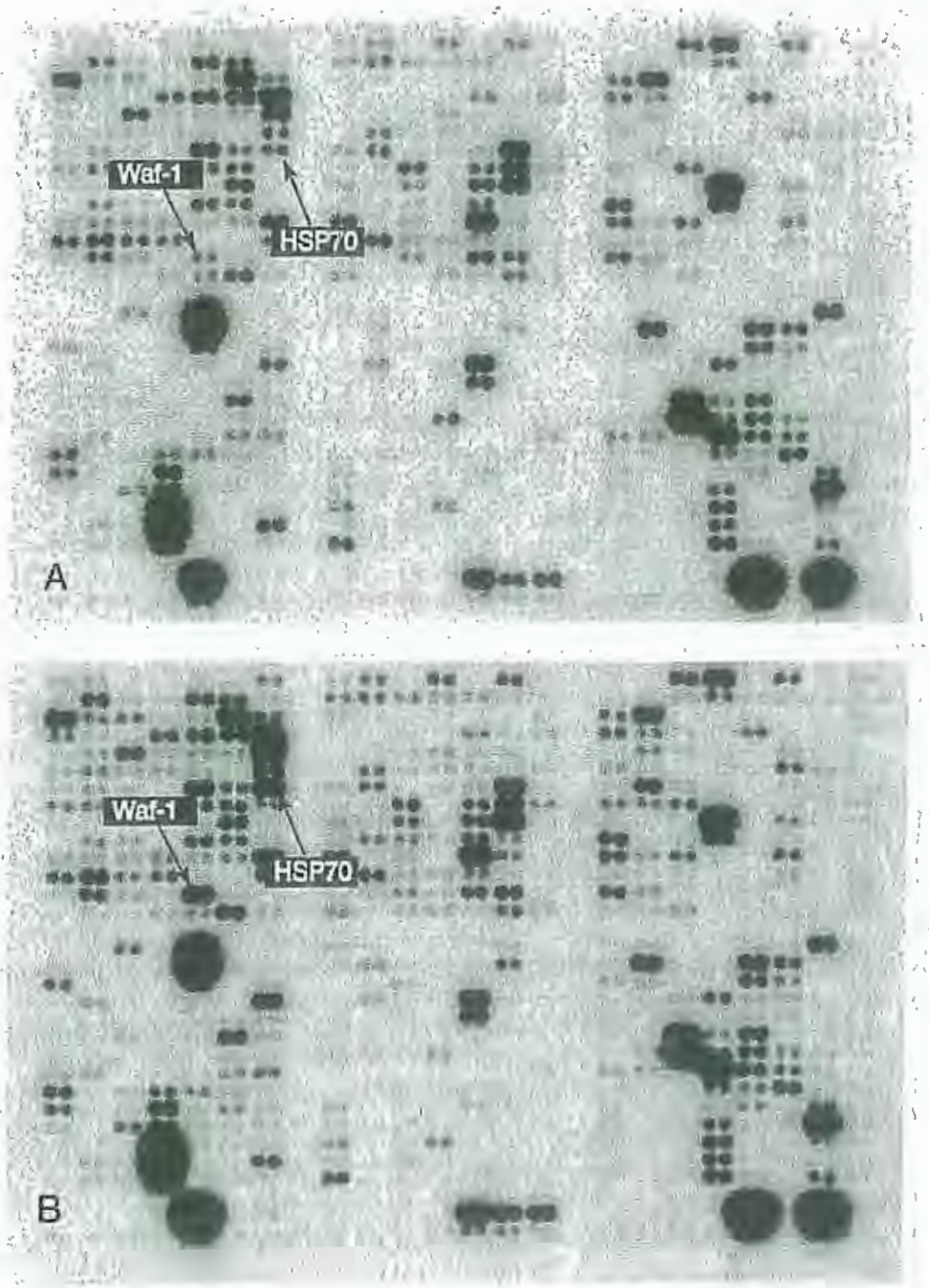


Figure 3. **Clontech gene expression arrays showing gene expression levels for 588 genes in rat mesothelial cells.** A. Untreated control. B. Cells exposed to 5 m*M* potassium bromate, an oxidizing agent. Note up regulation of the cyclin dependant kinase inhibitor (Waf1) and the heat shock protein (HSP70) following exposure to this oxidant. Also note the relative stability of the transcriptome in cells exposed to a highly toxic concentration of this compound. From Crosby *et al.* (Crosby *et al.* 2000a)

detect the induction of oxidative stress induced by exposure to potassium bromate (Figure 4). Oxidative stress was confirmed by quantitation of oxidized glutathione (GSSG) using HPLC and found to correlate well with up- or down-regulation of these genes on a standard exposure concentration *versus* response curve in HepG2 cells exposed to the oxidative stressor, sodium selenite (Shen *et al.* 1999) (Figure 5). Experience gained with studies of oxidative stress in this laboratory lead us into the complex process of interpreting gene expression array data sets. This interpretation process is in its' infancy, and we have much to learn.

INTERPRETATION OF ARRAY DATA BASED ON KNOWLEDGE OF UNDERLYING MECHANISMS

Gene expression array data sets can be overwhelmingly large and methods are developing for their interpretation, ranging from clustering procedures (Herwig *et al.* 1999) to gene-by-gene examination of the data. The detection and use of transcriptional biomarkers can be optimized by a thorough knowledge of the relationships between toxic insult and the associated transcriptional response observed. Changes in transcriptional activity of marker genes is but one component of a complex series of events. Expression arrays show transcript levels, while up and down stream events are not apparent. The fact that the transcript level is increased or decreased provides no direct information on flow of chemical substrates through their respective cellular pools. Furthermore, transcript signal levels provide little insight into related enzyme activity or protein and mRNA half-lives. If this information is needed, transcripts provide initial clues upon which to base further experimental work. When it comes to the meaningful interpretation of array data, "the devil is in the details."

As example of a specific case, control of GSH synthesis, is provided in Figure 6. The synthesis of GSH is regulated at the chemical, transductional and translational levels: for detailed review see Lu (1999). The major controlling factors appear to be the availability of cysteine and expression of the enzyme complex γ-glutamylcysteine synthetase (γ-GCS). We have found that upregulation of γ-GCS is a fairly reliable indicator of oxidative stress (1999) associated with GSH depletion and/or reduced GSH:GSSG ratio (Morgan *et al.* 2002)). This correlation is probably a consequence of the critical nature of levels of this transcript in GSH synthesis. Research directed toward finding and confirming sets of genes for the diagnosis of a wide range of mechanisms of toxicity promises to be a fruitful area of research for toxicogenomics.

RISK ASSESSMENT

With respect to risk assessment, it is important to distinguish "mechanism of action" from "mode of action." Transcript profiling can certainly aid the latter but the former is absolutely dependant on "one gene at a time" biochemical toxicology and molecular biology to determine the role of transcriptional responses in altering phenotype. It is important that regulatory scientists are aware that the technology in isolation has limited value in mechanism-based research.

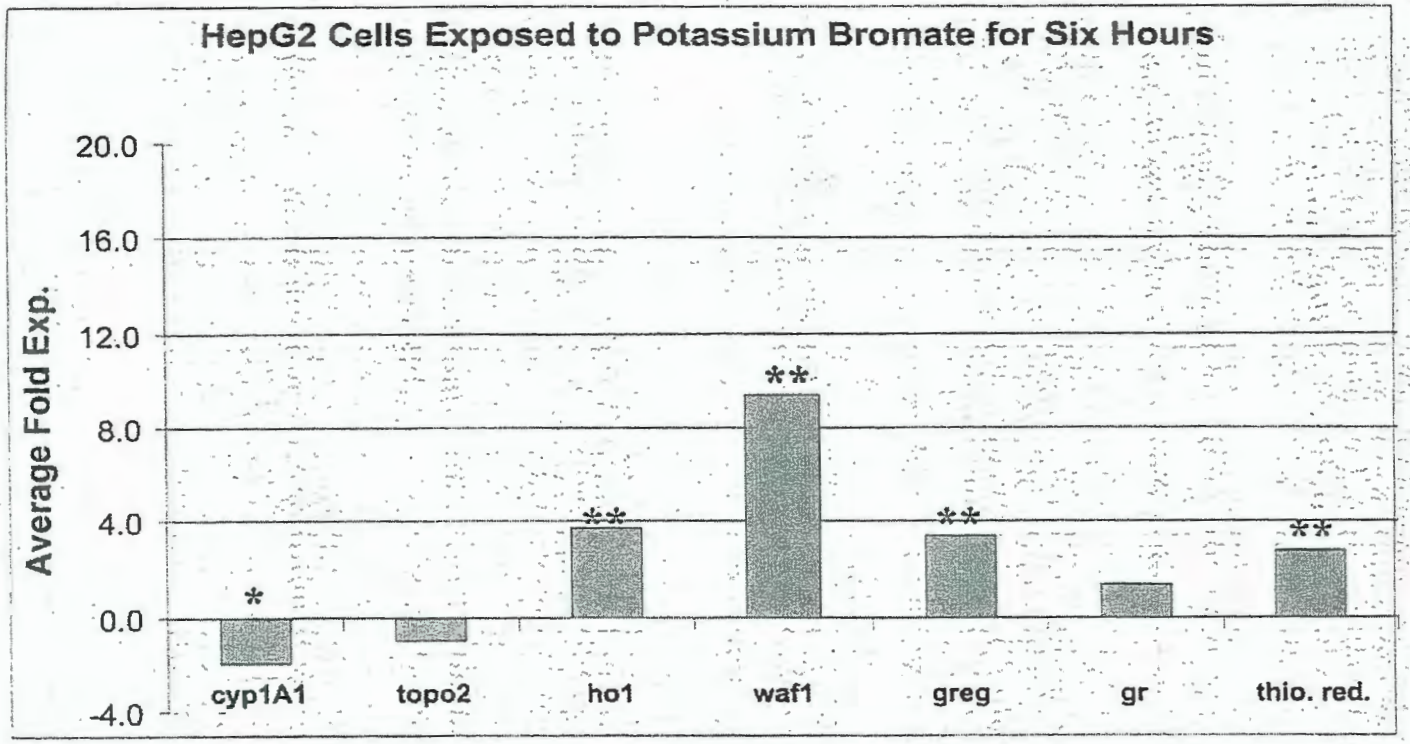


Figure 4. **Gene expression responses in HepG2 cells exposed for six hours to 2.5 m*M* potassium bromate under standard cell culture conditions.** These responses are considered to represent evidence of oxidative stress, consistent with previous studies in rodent mesothelial cells (Crosby *et al.* 2000a), and not seen in chemicals that failed to induce oxidative stress (INSERT). Three control and three treated cultures were used and data are expressed as mean fold change of treated *versus* control; * $p<0.05$, ** $p<0.01$. Note down regulation of ROS-generating (*CYP1A1*) and DNA repair (*TOPO2A*) genes, along with up-regulation of 'antioxidant' genes (heme oxygenase-1, γglutamylcysteine synthetase regulatory subunit (greg), glutathione reductase and thioredoxin reductase. Upregulation of the cyclin-dependent kinase inhibitor (waf1) is associated with cell cycle arrest in the cultures, probably in response to redox-induced activation of p53.

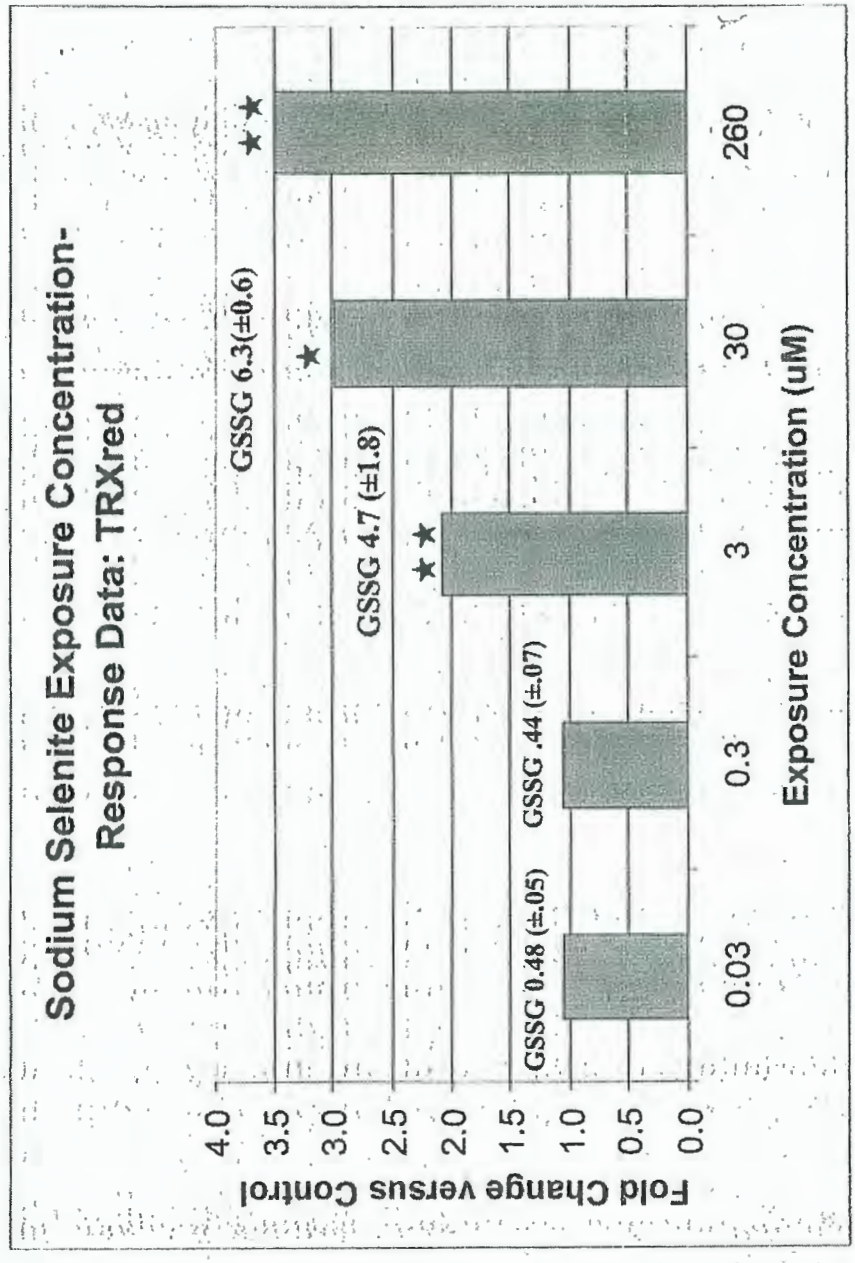


Figure 5. **Exposure concentration-response for thioredoxin reductase (TRXred) mRNA in HepG2 cells exposed to sodium selenite for 6 h.** For control *versus* treated cultures, * $p<0.05$, ** $p<0.01$. Oxidized glutathione (GSSG) measured by HPLC (INSERT), expressed in n*M*(±SD)/ml sample (~30 × 10^6 cells in ~1.2 ml 10% metaphosphoric acid) indicating oxidation of glutathione, and thus oxidative stress. Note clear correlation between expression of the 'antioxidant' gene, TRXred, and oxidative stressed as indicated by oxidation of glutathione.

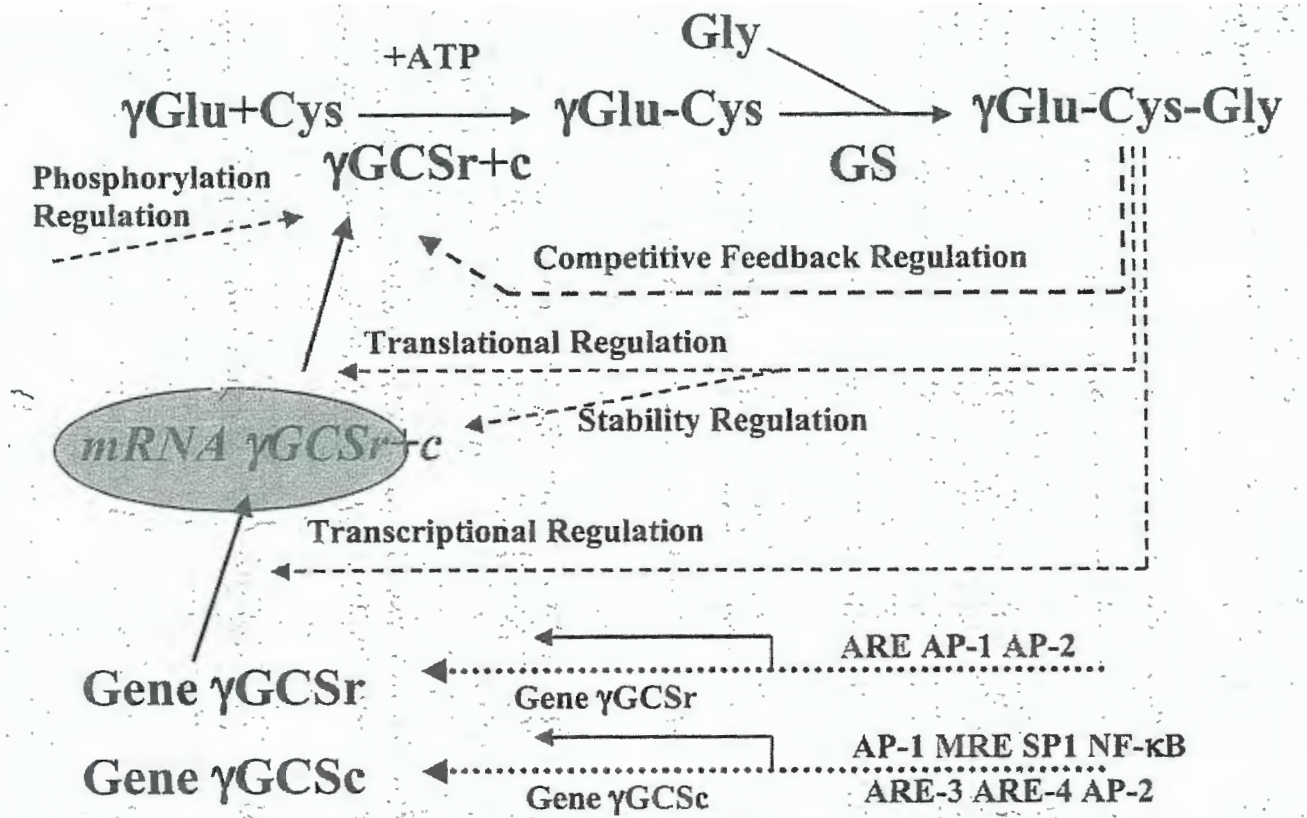


Figure 6. **Regulatory circuitry reported for γ-glutamylcysteine synthetase, regulatory (γ-GCSr) and catalytic (γ-GCSc) subunits; adapted from Lu (1999).** GS = glutathione synthase. The red circle covering mRNA γ-GCSr indicates information provided by a gene expression platform, which is clearly a small part of the total of events related to GSH synthesis. Transcript levels do not portray the related control circuitry and chemistry . Nor does the transcript signal give any indication of the dynamic state of the transcript; it just gives an indication of the amount of mRNA, in this case γ-GCSr, present at a given time.

In the short-term, transcriptomics and proteomics will probably be of most value for the hazard identification (Faustman and Omenn 1996) phase of risk assessment, while genetics can provide tools that permit the detection of susceptible subpopulations. For compounds about which little or nothing is known, the application of gene expression arrays or target gene expression screens can prove very enlightening, revealing unexpected responses by the target cell, tissue or organism. Data derived from arrays may reveal areas requiring further enquiry, such as effects on cell cycle or DNA damage. The identification of fingerprints, such as those discussed above for oxidative stress, will be a key task for toxicologists working in this area for the foreseeable future.

When screening large numbers of chemicals for potential toxicity and dose-response relationships, gene expression arrays will remain cumbersome in the near future. It is more likely that small gene sets, known to detect mechanisms of interest, will be incorporated into higher throughput gene expression-based screens. Irrespective of the gene expression technology applied to toxicity assessments, there will be a need for assays to confirm a random sample of 1 to 10% of these results. Protocols for assays to confirm selected biochemical responses and related cell functions will need to be developed as regulatory requirements if gene expression technology is to enter the mainstream of the risk assessment process. In this regard, the use of RT/PCR has come into somewhat more general use, and is becoming accepted as a valid alternative to Northern analysis for validation of responses (Crosby *et al.* 2000a). It is expected that new technologies will lead to continual improvements, including greater sensitivity and specificity, reduced sample requirements, combined with the ability to obtain a direct measure of the copy number of the transcripts of interest. Reduction in cost of such techniques, and large-scale expression arrays, will allow smaller laboratories to make use of them, opening such use to the careful scrutiny needed for their application to risk assessment. The application of these techniques in academic settings will make them available to the next generation of biology students, including those destined to become the risk assessors of the future.

Gene expression analyses are currently being investigated both *in vitro* and *in vivo*. Their effective application to intact human and non-human subjects will provide a range of transcriptional biomarkers of exposure and/or toxicity. Furthermore, the differentiation of species-specific mechanisms of toxicity will be markedly accelerated by these methods. As we gain experience with these tools, human risk assessments can only improve, as long as we base their use on solid experimental design (Steel and Torrie 1980) combined with appropriate application of statistics.

THE 'FACTORIAL FALLACY' AND OTHER REASONS FOR SKEPTICISM

A number of challenges have been raised concerning gene expression technology, including statements to the effect that we will never be able to interpret these data sets because the number of potential combinations of gene expression states is too large to analyze or interpret effectively. The latter reasoning we have dubbed "the factorial fallacy." Basically, it is proposed that, if each gene can be changed or not (altered or not in response to treatment), for n genes there are thus at least n! (n factorial) possible expression patterns. For the seven gene oxidative stress test

this would assume 7!, or 5,040 different states. For 100 genes this would grow to ~9.3^{157} states (a very large number indeed). With a genome of 30,000 genes, the number of potential states reaches 'astronomical' dimensions. This assumption is based upon the fallacy that these genes function independently, which is clearly not the case.

Basic biochemistry texts (Murray *et al.* 1996) make it very clear that certain metabolic, and thus transcriptional, states are mutually exclusive. For instance, the studies of transcriptional responses of pancreatic β-cells to glucose, using large-scale expression arrays (Webb *et al.* 2000), provided a classic example. There were many changes observed in response to glucose, including down-regulation of phosphoenol pyruvate carboxykinase (PEPCK). PEPCK plays a key role in cellular synthesis of glucose (gluconeogenesis), while a surplus of glucose was supplied to these cells, resulting in increased breakdown of glucose by glycolysis. If increased synthesis of glucose (associated with upregulation of PEPCK) occurred in combination with increased use of glucose, within a single cell, a futile cycle would result. Synthesized glucose would be consumed to produce energy, which would then be consumed to produce more glucose. Many metabolic switches prevent such futile cycles, and gene expression states supporting futile cycles are not options for a viable transcriptome, making the factorial argument invalid for gene expression patterns. An examination of Figure 4 shows, in fact, the remarkable stability of a transcriptome in cells that are under severe oxidative stress, many of which will be dead or dying within a few hours. If a factorial combination of gene expression profiles were possible, the expression patterns in Figure 4 would be much more clearly distinguishable as the transcriptome moved from one of a multitude of unrelated states to another.

Contrary to the factorial argument, it turns out that in response to a range of different stimuli, the complex "music of the genes" is orchestrated through a few simple underlying patterns of gene expression change (Holter *et al.* 2000). Wave-like changes in gene expression occur following alteration of cellular environments (Iyer *et al.* 1999). This might be expected, for instance in the case of oxidative stress, as a series of events occurs. Once the stress is imposed the cell has to (1) down-regulate certain functions in order to (2) upregulate the antioxidant machinery, concurrent with (3) cell cycle arrest if sufficient oxidative DNA, protein, or lipid damage occur, followed by (4) upregulation of repair pathways, and finally by (5) continued health or death by apoptosis, which also has a gene expression component. Thus, as in most things, for gene expression studies of toxic responses, (except, of course, exposure concentration) 'timing is everything.' Learning to interpret gene expression patterns combined with adequate training for the next generation of risk assessors will be key to the successful application of these techniques.

DEVELOPING EDUCATIONAL AND TRAINING APPROACHES

Interpretation of the large data sets generated by mRNA and protein expression arrays falls within the reach of the rapidly evolving science/art of bioinformatics. The complexity and scope of these data sets tends to overwhelm all but the most dedicated observer. As a direct consequence of the innate power of mathematics there is a risk that mathematical approaches (Baldi and Søren 1998) will excessively

mold and define the discipline of bioinformatics. If biologists are to exert appropriate levels of influence on the application of "omics" to toxicology and human disease risk assessment, they need to find ways of training a new generation of toxicologists and risk assessors in the field of bioinformatics, while encouraging a balanced approach. This new generation will have the interesting challenge of interpreting the massive data sets derived from new technologies while integrating these interpretations with data provided by tried and true disciplines, such as anatomy, physiology, biochemistry, pharmacology, and pathology. The 'old' disciplines are easy to forget when you are excited by the promise of these new technologies. Finding and developing the right mixture of mathematical and 'intuitive biology' skills in risk assessors of the future will be key to the success of this endeavor.

ACKNOWLEDGMENTS

We thank the many people who made our work possible, with special thanks to Charles Qualls for recommending studies of oxidative stress. Extensive intellectual and/or technical support was provided by Thomas Kepler (NCSU), Marilyn Easton, Hong Ni, and Karim Hyder. Lynn M. Crosby was supported by a UNC Curriculum in Toxicology Postdoctoral appointment.

REFERENCES

Anderson ME. 1998. Glutathione: an overview of biosynthesis and modulation. Chem-Biol Inter 111-112:1-14

Armstrong DE. 1998. Free Radical and Antioxidant Protocols. Humana Press, Totawa, NJ, USA

Baldi P and Søren B. 1998. Bioinformatics, The Machine Learning Approach. The MIT Press, Cambridge, MA, USA

Cole KA, Krizman DB, and Emmert-Buck MR. 1999. The genetics of cancer - a 3D model. Nature Genetics 21:38-41

Collins FS. 1999. Overview of DNA chip technology. Nature Genetics 21:1-60

Crosby LM, Benavides G, Yoon L, *et al.* 2000a. Morphologic analysis correlates with gene expression changes in cultured F344 rat mesothelial cells. Toxicol Appl Pharmacol 169:205-21

Crosby LM, Hyder KS, DeAngelo AB, *et al.* 2000b. Gene expression array technology reveals underlying mechanisms of mitotic arrest and apoptosis induced in rat mesothelial cell cultures by potassium bromate. In 39th Annual Meeting of the Society of Toxicology, vol 54 (suppl), p 357 (Abstract #1674). Oxford University Press, Philadelphia, PA, USA

Farr S and Dunn RTI. 1999. Concise review: gene expression applied to toxicology. Toxicol Sci 50:1-9

Faustman EM and Omenn GS. 1996. Risk assessment. In: Klaasen (ed), Casarett and Doull's Toxicology, The Basic Science of Poisons, pp 75-88. McGraw-Hill, Health Professions Division, NY, NY, USA

Gille G and Sigler K. 1995. Oxidative stress and living cells. Folia Microbiol 40:131-52

Guyton KZ, Gorospe M, and Holbrook NJ. 1997. Oxidative stress, gene expression, and the aging process. In: Scandalios G (ed), Oxidative Stress and the Molecular Biology of Antioxidant Defenses, J vol 34, pp 247-72. Cold Spring Harbor Laboratory Press, NY, NY, USA

Haffner SM. 2000. Clinical relevance of the oxidative stress concept. Metabolism 49:30-4

Herwig R, Poustka AJ, Muller C, *et al.* 1999. Large-scale clustering of cDNA-fingerprinting data. Genome Res 9:1093-105

Holter NS, Mitra M, Maritan A, *et al.* 2000. Fundamental patterns underlying gene expression profiles: simplicity from complexity. Proc Natl Sci USA 97:8409-14

Iyer VR, Eisen MB, Ross DT, *et al.* 1999. The transcriptional program in the response of human fibroblasts to serum. Science 283:83

Jehn BM and Osborne BA. 1997. Gene regulation associated with apoptosis. Crit Rev Eukaryo Gene Express 7:179-93

Klaassen CD (ed). 1996. Casarett & Doull's Toxicology, The Basic Science of Poisons. McGraw-Hill, NY, NY, USA

Lee CK, Klopp RG, Weindruch R, *et al.* 2000. Gene expression profile of aging and its retardation by caloric restriction. Science 285:1390-93

Lu S. 1999. Regulation of hepatic glutathione synthesis: current concepts and controversies. FASEB J 13:1169-83

Morgan KT, Ni H, Brown HR, *et al.* 2002. Application of cDNA microarray technology to *in vitro* toxicology and the selection of genes for a real time RT-PCR-based screen for oxidative stress in Hep-G2 cells. Toxicol Pathol (*in press*)

Murray RK, Granner DK, Mayes PA, *et al.* 1996. Harper's Biochemistry. Appleton & Lange, Stamford, CT, USA

Nicholls P. 1999. The mitochondrial and bacterial respiratory chains: From MacMunn and Keilin to current concepts. In: Sergio Papa FGaJMT (ed), Frontiers of Cellular Bioenergetics, pp 1-22. Kluwer Academic/Plenum Publishers, NY, NY, USA

Nuwaysir EF, Bittner M, Trent J, *et al.* 1999. Microarrays and toxicology: the advent of toxicogenomics. Mol Carcinogen 24:153-9

Saran MM, Christa; Bors, Wolf 1998. Radical functions in vivo: a critical review of current concepts and hypotheses. Radical Functions *In Vivo* 53:210-27

Scandalios JG. 1997. Oxidative Stress and the Molecular Biology of Antioxidant Defenses. Cold Spring Harbor Press, NY, NY, USA

Shen H, Yang C, and Ong C. 1999. Sodium selenite-induced oxidative stress and apoptosis in human hepatoma HepG2 cells. Int J Cancer 81:820-8

Steel RGD and Torrie JH. 1980. Principles of experimental design. In: Principles and Procedures of Statistics, pp 122-36. McGraw-Hill, Inc, NY, NY, USA

Wang Y, Rea T, Bian J, *et al.* 1999. Identification of the genes responsive to etoposide-induced apoptosis: application of DNA chip technology. FEBS Letters 26:269-73

Webb GC, Akbar MS, Zhao C, *et al.* 2000. Expression profiling of pancreatic β cells: glucose regulation of secretory and metabolic pathway genes. Proc Natl Sci USA 97:5773-8

Wen X, Fuhrman S, Michaels GS, *et al.* 1998. Large-scale temporal gene expression mapping of central nervous system development. Proc Natl Sci USA 95:334-9

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1355-1374 (2002)

Simplicity vs. Complexity in the Development of Risk Models for Dose-Response Assessment

D. Krewski,[1]* K. P. Brand,[1] R.T. Burnett,[2] and J.M Zielinski[2]
[1]Department of Epidemiology and Community Medicine, McLaughlin Centre for Population Health Risk Assessment, University of Ottawa, Ottawa, Canada;
[2]Healthy Environments and Consumer Safety, Health Canada, Ottawa, Canada.

ABSTRACT

Deliberations over chemical safety rely heavily upon the interpretation of toxicological and epidemiological evidence. Dose-response modeling plays an important role, allowing true effects to be more effectively discerned from background processes (qualitative determinations) and playing an even more instrumental role in quantitative judgments (*e.g.*, the estimation of potency or acceptable dose). We survey five relatively distinct topic areas (ranging from the assessment of genotoxicity, to epidemiological studies of respirable particles) where dose-response models have been applied, and explore the varying degrees of complexity used in modeling. We choose a descriptive approach, presenting each topic area as a case study. The survey reveals a wide spectrum of complexity both within and between topic areas. The question of 'what level of model detail is appropriate?' is widely debated in various disciplines. Notably, the policy context of dose-response assessment (with its attendant real-world stakes) involves some added considerations. We therefore try to emphasize the role of each modeling application's outcome in affecting a decision process. The case studies provide a first step in identifying issues related to model parsimony that are unique to the dose-response, and policy realm.

Key Words: dose-response modeling, parsimony, regulatory toxicology, risk assessment, mutagenicity, developmental toxicity, carcinogenicity, cardiorespiratory health.

I. INTRODUCTION

Toxicological and epidemiological data continue to serve as key sources of information for the assessment of population health risks. Well-established frame-

* Corresponding author: Dan Krewski, Professor and Director, McLaughlin Centre for Population Health Risk Assessment, Institute of Population Health, University of Ottawa, One Stewart Street, Room 320, Ottawa, ON K1N 6M5 Canada; Tel(voice):613-562-5381, Tel(fax):613-562-5380; dkrewski@uottawa.ca

1080-7039/02/$.50

works for health risk assessment developed by the U.S. National Research Council (NRC 1983) and the U.S. Presidential/Congressional Commission on Environmental Health Risk Assessment and Risk Management recognize these sources of data both for hazard identification and risk characterization, and the subsequent implications of these data for risk management.

An important component of risk characterization is dose-response or exposure-response assessment. Whereas dose-response assessment requires information on the dose of the reactive metabolite reaching the target tissue, exposure-response assessment is based on oral, dermal, inhalation or other relevant measure of exposure. While both dose-response and exposure-response models are discussed in this article, the former term will often be used in general discussion for simplicity.

Dose-response assessment contributes to risk characterization in several ways. First, the development of a suitable dose-response model provides a convenient way of describing how risk varies with dose. Such an overall description of the dose-response relationship is often the first step in quantitative risk assessment, and provides the risk assessor with a general understanding of the relationship between dose and response (Moolgavkar *et al.* 1999). Second, a dose-response model can be used to estimate key indicators of risk, such as the benchmark dose (Gaylor *et al.* 1998), which in turn provide a basis for the establishment of exposure guidelines. Third, more complex dose-response models can be used to describe temporal aspects of risk (Goddard *et al.* 1995), modifying effects of important covariates, and possibly the mechanisms by which toxic substances lead to the induction of adverse health effects (Goddard and Krewski 1995).

In this article, we explore the balance between simplicity and complexity in the construction of dose-response models for risk assessment. This balance will depend on the objectives, the availability of appropriate methodologies to achieve these objectives, and the quality and depth of the available data. For certain applications, a comparatively simple model may be sufficient to address the issue of interest. For other applications, however, the use of more complex models may be required.

Our exploration of simplicity vs. complexity in dose-response assessment will be based on an examination of a number of examples in which risk models have been developed to describe both toxicological and epidemiological data. In Section 2, we discuss both statistical and biologically based dose-response models for the Ames *Salmonella* assay, which is widely used to evaluate the potency of chemical mutagens. In Section 3, the modeling of developmental toxicity experiments is described. These particular models take into account both multivariate outcomes (specifically, embyrolethality and fetal malformations) and correlated binary responses.

Risk models that have been used to describe dose-response relationships in carcinogenesis are described in Section 4. This is a well-developed area in risk modeling, including simple procedures for estimating upper bounds on cancer risk assuming that the dose-response curve is linear at low doses to complex biologically based models that take into account cellular kinetics and genetic alterations involved in carcinogenesis. Our final example (Section 5) involves the use of new dose-response models that have been developed to describe the relationship between exposure to particulate matter in urban air and cardiorespiratory mortality, taking into account spatial patterns in the data. These particular models were developed as part of a comprehensive reanalysis of epidemiological data from a

study of over 150 metropolitan areas in the United States using data on more than 550,000 subjects originally collected by the American Cancer Society. Our conclusions concerning the balance between simplicity versus complexity in dose-response modeling are presented in Section 6.

II. MODELING THE AMES *SALMONELLA* ASSAY

The Ames test is designed to determine whether a chemical can induce mutations in various strains of *Salmonella* (Krewski *et al.* 1992). The bacteria are genetically engineered to be incapable of synthesizing an amino acid, histidine, which is essential for survival. Mutations reverting the bacteria's genotype to its original form (thereby restoring their ability to synthesize histidine), can be detected by their ability to proliferate. Increases in the frequency of micro-colonies (a manifestation of the restored ability to proliferate), called revertants, are indicative of mutagenicity. A comparison of revertant counts between treated and untreated plates is used to evaluate mutagenicity.

Qualitative judgments of mutagenicity follow for agents causing significantly more revertants in treated plates, and their potency is assessed. If an agent is deemed to be mutagenic it is likely to be more stringently regulated, not only because of concern about its mutagenic effects, but also because of concern about its potential carcinogenicity. Mutagenicity is thought to be an indicator of carcinogenic potential, lending plausibility to a genotoxic mode of carcinogenesis. This mode of carcinogenicity is of particular concern because it is thought to involve a linear no threshold response, implying some elevated risk no matter how low the dose.

A number of dose-response models have been proposed to describe the relationship between dose and the number of revertants found in the Ames test. The more simplified approaches for mathematically describing how a chemical's genotoxicity varies with dose, were purely empirical. More complicated approaches were directed towards improved modeling of either the central tendency of the observed dose response trends or their variance (error structure). The simplified models implied a monotonic relationship between dose and response, and yet a large fraction of data sets suggested otherwise (Margolin *et al.* 1981; Lewtas *et al.* 1992). Specifically, the observed data exhibited a domain of monotonic increasing response (at lower doses) followed by a domain of monotonic decreasing response (at higher doses). The simplified models also presumed a Poisson distribution of counts at each dose and yet the data suggested extra-Poisson variation (overdispersion). We will not consider the efforts to model extra-Poisson variation in this brief review, but will instead focus on efforts to better model the central trend of the dose response relationships (Margolin *et al.*1981).

As an example of the simplified approach, Bernstein *et al.* (1982) modeled the Ames test outcome using a purely empirical model of the dose response relationship and assumed a Poisson 'error-structure.' They fitted a simple exponential model to Ames test outcome, and used a somewhat ad hoc procedure to address those dose-response outcomes showing inverted-U shaped curves: dropping all higher doses until the responses in the remaining doses could be adequately fit by a monotonic function (one-hit).

Margolin *et al.* (1981) proposed a quasimechanistic model as an alternative to the simplified models. It invokes assumptions about how mutagenic and toxicologic processes govern colony formation, and involves an additional modification to account for the observed overdispersion.

Their model assumes that sufficient histidine exists to permit growth through k generations, that a bacterium has a constant probability of reverse mutating in the first m generations, a constant (and possibly different) probability of mutating in the next k-m generations, and zero probability of mutating thereafter. A constant probability is also assumed for cytoxicity, though it is presumed to apply for only the first t generations. Further, the opposing processes are assumed to be independent.

Using these assumptions Margolin developed formulae to model $P_D(m,t)$, the probability that a colony is observed in a plate exposed to dose D, assuming that mutation and cytotoxicity are possible for m and t generations respectively. These assumptions define a broad class of dose response functions, with special cases corresponding to specific values of m, k and t. In the case of $m = k = t = 1$, one obtains,

$$p_D(1,1) = p_M(D)\ (1 - p_T(D))$$

where the specific functions for cyto-toxicity and mutagenesis can be defined using standard empirical dose response functions. For example, the assumption of single hit kinetics implies

$$p_M(D) = 1 - \exp(-H_M(D))$$

and

$$p_T(D) = 1 - \exp(-H_T(D))$$

where H_M and H_T are simple polynomials in dose.

Krewski *et al.* (1993) have extended this model, by allowing the maximum number of cell divisions to depend on the total supply of histidine in the plate agar rather than assuming that each cell's histidine supply is strictly localized. This approach yielded another class of models for describing the Ames test outcome. For the case of m=k=1, the expected number of colonies at dose D is approximately

$$\mu(D) = (\beta_0 + \beta_1 D)\ \exp(-\beta_2 D)$$

where the β's are obtained by maximum likelihood fits to the data. This model can be conceptualized as a simple linear function in dose (for mutation), with an exponential attenuation factor to capture the impact of cytotoxicity on preventing cell proliferation.

The low dose mutagenic potency can be expressed as $\beta_1 - \beta_1 * \beta_2$; however if cytotoxicity is assumed to have a threshold, the potency would reduce to β_1. Instead of defining a threshold, Krewski modified the exponential attenuation factor, by putting a supra-unity power on dose,

$$\mu(D) = (\beta_0 + \beta_1 D)\ \exp(-\beta_2 D\theta), \text{ where } \theta > 1.$$

Empirical analysis found that a θ of 2 fit most datasets quite well (Leroux and Krewski 1993), thereby suggesting that cytotoxicity does vary sublinearly with dose.

This example illustrates the use of both simple empirical models and more complex biologically based models to describe mutation rates in the Ames *Salmonella* assay. The initial slope of the dose-response curve is often linear at low to moderate doses, reflecting a direct relationship between dose and mutation frequency prior to the induction of nonlinear cytotoxic effects at high doses. If the risk assessment objective is simply to obtain an indicator of mutagenic potency such as the initial slope of the dose-response curve, then simple models will suffice. The more complex biologically based models will however be of value in exploring mechanistic hypothesis about the way in which mutations occur in simple bacterial test systems.

III. DOSE-RESPONSE MODELS FOR DEVELOPMENTAL TOXICITY STUDIES

Developmental toxicity experiments carried out in animals represent an important tool for the identification of chemicals that pose a threat to the developing fetus (Krewski *et al.* 1999a). Several studies typically use rats or mice, with a control group and on the order of three to four dosed groups, each of which includes 20 to 30 impregnated dams. Dosing generally starts during major organogenisis (early in gestation), and the dams are followed up until just before term. At this time they are sacrificed so that the uterus can be examined, revealing the number of implants, resorptions, stillbirths, and various types of malformations. Body weight measurements are also taken. The results are then used to ascertain whether the test chemical had an impact on these various developmental indicators, and if so their relative potency. Statistical models are involved in this process, and in the more quantitative task of characterizing a dose response relationship (Catalano and Ryan 1994; Krewski *et al.* 1999a).

There are at least three special features of developmental studies that have inspired modeling extensions. These include the presence of multiple outcomes, a hierarchical structure across the endpoints, and inter-litter correlation. Simplified models tend to ignore one or more of these features, and therefore may result in theoretically incorrect results. More sophisticated modeling approaches, make attempts to capture the features, but require more detailed input information, and more complicated computational techniques.

The hierarchical nature of the developmental studies refers to a conditional dependence structure that exists across some of the indices. For example, only surviving pups are at risk for malformation. Thus the malformation endpoint must be conditioned on the outcome of the fetal-viability end point.

The existence of multiple endpoints, including, fetal death, fetal body-weight, and (several types of) malformation, has necessitated efforts to consider the covariance structure across endpoints. Some analyses make the simplifying assumption that the endpoints (*e.g.*, body-weight and malformations) are independent, while other more sophisticated analyses attempt to allow for dependence, by modeling the endpoints jointly. A joint analysis can have significant advantages, especially if the different endpoints are the manifestation of a single common mechanism; it

exploits more of the available information. Even if the endpoints result from distinct biological/pharmacological mechanisms, a joint analysis of the endpoints will be a more sensitive endpoint than any of the endpoints in isolation. However, there are still practical issues about how to interpret significance, risk, and safety, jointly across the endpoints.

Krewski *et al.* (1995) used the Weibull models

$$\pi_i(d) = 1 - \exp(-a_i - b_i d^{v_i})$$

$(a_i \geq 0, b_i \geq 0)$ to describe the dose-response for embryolethality and malformations, respectively, $\pi_i (i = 1, 2)$; where π_1 is the probability of any malformation in a live fetus, and π_2 is the probability of a prenatal death, and the remaining terms are endpoint specific fitted parameters. Overall toxicity is defined as the occurrence of either of these endpoints and can therefore be expressed as,

$$\pi_3(d) = 1 - [1 - \pi_1(d)][1 - \pi_2(d)]$$

The issue of inter-litter correlation has also drawn attention. Each dam typically has on the order of 10 or 15 pups, and littermates tend to have similar adverse effects. Neglecting this inter-litter correlation can result in underestimation of variance, and an overstatement of statistical significance. Some approaches allow for and attempt to estimate the inter-litter correlation. (Pragmatic issues arise, such as whether to insist on a common correlation across all litters, or whether to allow the correlation to depend upon, among other things, dose.)

IV. DOSE-RESPONSE RELATIONSHIPS IN CARCINOGENESIS

Cancer risk assessment plays a prominent role in regulatory determinations of safety. The qualitative judgment of whether a chemical is a carcinogen has a strong influence on such deliberations. Once a chemical is deemed carcinogenic, quantitative characterizations of its dose response relationship can also be influential, supporting estimates of risk for a given exposure of interest or an estimate of safe or acceptable dose given a prespecified level of acceptable risk.

A variety of modeling approaches has been used for dose response assessment (Moolgavkar *et al.* 1999). As in the case of mutagenicity these approaches evolved from strictly empirical models, to quasimechanistic models attempting to redress unappealing properties of the first models, and finally to more and more biologically based models.

A. High- to Low-Dose Extrapolation of Rodent Bioassay Data

High to low dose extrapolation, a special application of dose response modeling, is required to project low dose implications from high dose data. Two characteristic features of high to low dose extrapolation have encouraged a wide range of sophistication in approach. First, the evidence does not support a strong preference across the array of plausible dose-response models. Second, that same array of models will typically imply a widely divergent set of low dose implications. A polarizing tension results because the evidence is so ineffective in choosing among models, and yet that

choice matters immensely. The ineffectiveness is not so surprising once one considers how far removed the observed data are from the domain of interest (the experimental data is typically several orders of magnitude higher than the domain of interest to human exposures), and the limited relevance of ancillary data.

Several forces may be seen at work in the various extrapolation approaches that have been employed. Efforts to promote transparency favor restricting the choice via some rationale, thereby limiting the opportunities for inconsistency. Efforts to err on the side of caution favor a choice more likely to over- than underestimate the risk. Others may have a different concern, fearing that erring on the side of caution could have undue opportunity costs. Such concerns are typically manifest in calls to "sticking as close to the data as possible" when doing so would appear to favor less stringency. Calls are made to better accommodate features in the observed data, or better reflect the implications of ancillary data. Finally, some may advocate more simplified models dispensing with complexities incommensurate with relative crudity of the evidence (Krewski *et al.* 1991; Olin *et al.* 1995).

These forces, which are not mutually exclusive, may help interpret the variety of approaches that have been used for high to low dose extrapolation. Early approaches were empirically based and deferred to pre-existing models, such as the probit, log-probit and Weibull models (see Zeise *et al.* 1987 and Olin *et al.* 1995 for good reviews). After grappling with the dilemma of widely divergent low dose implications, efforts to standardize and restrict the choice started to appear (Anderson *et al.* 1983; Olin *et al.* 1995). These did not go unchallenged. Some expressed concerns that the restrictions conflicted with efforts to stick closely with the data, others stressed concerns about the possibility of an excessive opportunity (*e.g.*, control) cost (Sielken 1987). One prominent restriction of choice was the strong preference for the linearized multistage model, a model thought to have some biological basis, but more importantly enjoying a rare consensus that its predictions were more likely to over, rather than understate, the true risks (Anderson 1983).[1] While some saw this property as a virtue, others, citing the potential for unduly burdensome control costs, did not.

As an example of those favoring parsimony, Krewski *et al.* (1991) proposed a simple approach to low dose extrapolation of carcinogen bioassay data that involves simple linear extrapolation from a point well down on the dose response curve, but still within the experimentally observable risk range, such as the BMD05. For most datasets, this simple approach leads to unit risks close to those based on the Armitage-Doll model, thereby avoiding the complexity of the latter model.

Further simplification can be achieved by exploiting the high correlation between the maximum tolerated dose or MTD (the highest dose used in carcinogen bioassay) and the unit risk (Krewski *et al.* 1993). This relationship can be used to predict the unit risk once the MTD is determined, although completion of the bioassay is needed to establish evidence of carcinogenic potential (Figure 1). Gaylor and Gold (1998) show that preliminary risk estimates of a virtually safe dose corresponding to a one in a million lifetime risk can be obtained simply by dividing the MTD by a factor of 700,000.

---

[1] One should not overinterpret this consensus. The estimate is in fact not by any means guaranteed to overstate risks. Too many contingencies hold, to make a useful statement.

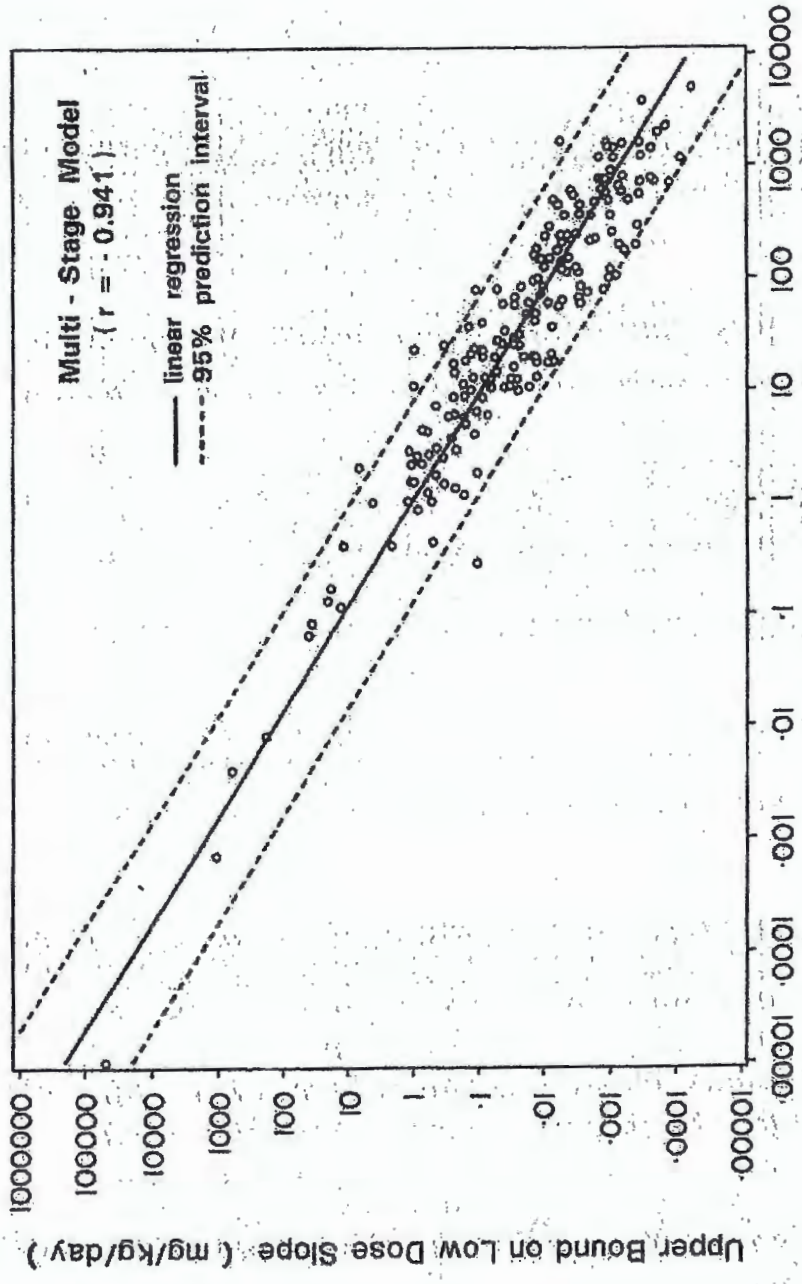


Figure 1. **The Maximum Tolerated Dose (MTD), used in rodent bioassay experiments, is plotted against the potency index (q1*) estimated from that experiment.** There is clearly a strong relationship across a wide range of MTDs. The predictive interval implies a roughly two-order of magnitude span between the upper and lower confidence limits for q1*. (Excerpted from Krewski *et al.* 1993.)

B. Two-Stage Models for Radiation Carcinogenesis

The two-stage model of cancer is based on the theory that a cancer cell is produced following the occurrence of two mutations in a single stem cell, with cells that have sustained the first mutation (initiated cells) subject to a birth-death process leading to clonal expansion (Moolgavkar and Luebeck 1990). Kodell *et al.* (1991) refer to agents that increase the first or second stage mutation rate as initiators or completers, respectively, with agents increasing the clonal expansion rate being promoters.

The two-stage model has been employed in a number of applications in cancer risk assessment, and provides a useful biologically based approach to cancer risk modeling (Moolgavkar *et al.* 1999). Moolgavkar *et al.* (1993) used the exact form of the two-stage model to describe the interaction between radon and tobacco smoke in the induction of lung cancer in underground miners as described below. The transition intensity functions rates corresponding to the first and the second mutation rates were modeled as linear functions

$$\nu(d_s, d_r) = a_0 + a_s d_s + a_r d_r$$

and

$$\mu(d_s, d_r) = b_0 + b_s d_s + b_r d_r,$$

where d_s and d_r represent the level of exposure to tobacco smoke and radon, respectively. (Both d_r and ds may vary with age.) The rate of promotion was modeled as the nonlinear function,

$$(\alpha - \beta)(d_s, d_r) = c_0 + c_{s1}(1 - \exp[-c_{s2} d_s]) + c_{r1}(1 - \exp[-c_{r2} d_r])$$

with β/α held constant. This functional form allows for saturation of the effects of both radon and tobacco smoke. Since no effect of radon or smoking on the second mutation rate was observed, b_s and b_r were set to zero. With the identifiability constraint $a_0 = b_0$, only nine parameters were actually estimated.

The relative risks of lung cancer at age 60 are given in Table 1 (based on the fitted model) for both single and joint exposures to radon and cigarette smoke. Also shown is Thomas' index of synergy considered by Zielinski *et al.* (2001), which indicates that the relative risks are supra-additive ($S > 0$) but submultiplicative ($S < 1$).

Zielinski *et al.* (2001) used the parameter values obtained by fitting the exact form of model to the Colorado miners data as a starting point in for further numerical investigation of interactive effects between two carcinogens. Since the parameters b_s for the second mutation rate were equal to zero, they set $b_s = a_s$ and $b_r = a_r$ to include agents that can increase the second stage mutation rate.

Patterns of interaction between two carcinogens acting on a single component (initiation, promotion or completion) were investigated under a scenario in which exposure to either carcinogen started at age 15 years of age and continued at the same level through to 80 years of age. Duncan's index of synergy S based on the age-specific hazard is shown in Figure 2 for four different carcinogen combinations. The

Table 1. Relative Risk of Lung Cancer due to Exposure to Radon and Tobacco Smoke Based on Exact Age-Specific Hazard[a].

Radon Exposure[b] (WLM/month)	Tobacco Exposure[c] (cigarettes/day)	Relative Risk due to Radon	Relative Risk due to Tobacco	Relative Risk Combined Exposure	Index of Synergy S
1	10	1.3	5.4	6.6	0.68
1	30	1.3	10.2	12.3	0.70
1	40	1.3	11.8	14.4	0.73
50	10	12.9	5.4	28	0.21
50	30	12.9	10.2	46.8	0.23
50	40	12.9	11.8	55.3	0.25

[a] Relative risk evaluated at 60 years of age
[b] Exposure to radon between 30 and 40 years of age
[c] Cigarette smoking between 15 and 60 years of age

interaction between two initiators conforms closely to additivity for all t. The case of constant lifetime exposure to two completers (not shown) follows a pattern similar to that for two initiators. Unlike the cases of exposure to two initiators or two completers in which the risks are close to additive, a near multiplicative relative risk relationship arises with joint exposure to an initiator and completer. (Note that for large t, the relationship becomes supra-multiplicative.)

In general, Zielinski *et al.* (2001) found the temporal patterns of interaction to be qualitatively similar regardless of whether the relative risk was based on age-specific hazard or cumulative probability. For joint exposure to two initiators or to two completers, the values of the index of synergy were calculated to be near zero, reflecting an additive relative risk relationship. For joint exposure to two promoters, the relative risk relationship was found to range from supra-multiplicative ($S > 1$) in younger age groups to sub-additive ($S < 0$) at older ages.

These results differ notably from those reported previously by Kodell *et al.* (1991) for the approximate form of the two-stage model, which predicts much higher values of the index of synergy S than the exact form of the model when promotion is involved. The fact that maximal discrepancies between the approximate and exact forms of the two-stage model occur when one or both agents is a promoter is not surprising, since the approximation uses the difference between the birth and death rates (rate of clonal expansion of initiated cells) as a sufficient statistic.

Biologically based cancer risk models continue to evolve as our understanding of the process of carcinogenesis increases (Moolgavkar *et al.* 1999). Extensions include allowances for stochastic stem cell growth (Denes and Krewski 1995), clonal expansion at multiple intermediate stages (Zheng *et al.* 1997), and the incorporation of information on the number and size of premalignant clones in fitting the two-stage model in longitudinal studies (Dewanji *et al.* 1999). Tan (1991) has presented many related extensions as well.

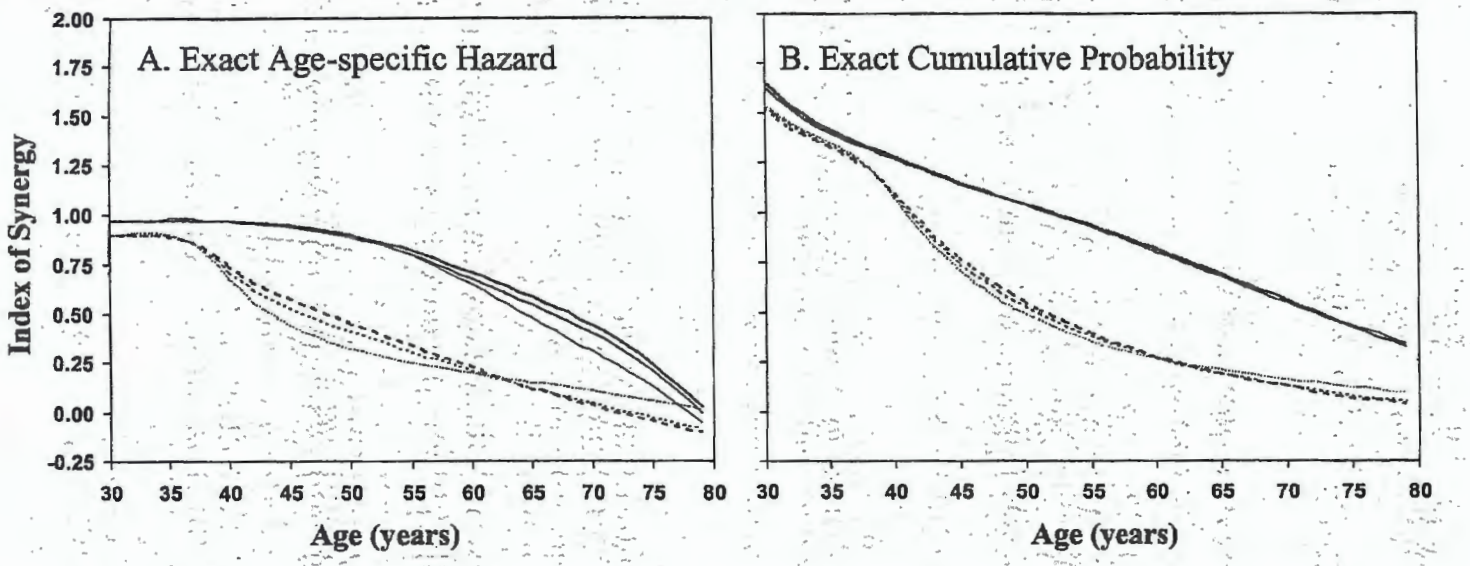


Figure 2. **Index of Synergy Between Radon and Tobacco Based on Exact Age-Specific Hazard (A) and Exact Cumulative Probability (B). (Excerpted from Zielinski *et al.* 2002.)**

C. Empirical Models for Radon and Lung Cancer

Underground miners exposed to high levels of radon in the past have been shown to be at excess risk of lung cancer (Lubin *et al.* 1995, 1997). Based on the results of a series of epidemiological studies of lung cancer in underground miners, the International Agency for Research on Cancer has classified radon as a human carcinogen (IARC 1988).

Released from many rocks and soils, radon is found at varying concentrations in indoor air in residences. Since radon has been shown to increase lung cancer risk among miners exposed to higher levels of radon, exposure to residential radon may also present some level of risk. Residential radon lung cancer risks were considered by the most recent Committee on the Biological Effects of Ionizing Radiation (BEIR VI).

The BEIR VI Committee (NRC 1999) adopted an empirical approach to risk estimation similar to that employed by Lubin *et al.* (1994). A dosimetric approach based on projections of risk from A-bomb survivor studies was not pursued because of differences in the type of radiation and radiation exposure patterns. Biologically based risk models, such as those considered by Moolgavkar *et al.* (1993), were not implemented partly because of the complexity of such models and partly because of uncertainties in the mechanism of radon carcinogenesis.

The starting point for development of the BEIR VI risk models was the combined analysis of 11 studies of underground miners conducted by Lubin *et al.* (1994). In total, 2674 lung cancer deaths were observed in the 68,000 miners included in these 11 cohorts. Using these data, the Committee developed the following model to describe the excess relative risk (ERR) of lung cancer associated with exposure to radon:

$$ERR = \beta\, w^{*}\phi_a\, \gamma_z$$

This model expresses the ERR, which represents the multiplicative increase in the excess lung cancer risk due to exposure to radon, as a function of an exposure-response parameter (β) and past exposure to radon (w*), modified by multiplicative factors representing attained age (ϕ_a) and either exposure concentration or exposure duration (γ_z). These two models are referred to as the exposure-age-concentration and exposure-age-duration models, respectively. Excluding exposures occurring within the last 5 years as not biologically relevant to cancer risk, cumulative exposure $w^{*} = w_{5\text{-}14} + \theta_{15\text{-}24} w_{15\text{-}24} + \theta_{25+} w_{25+}$, is partitioned into temporal exposure windows $w_{5\text{-}14}$, $w_{15\text{-}24}$, w_{25+}, defining exposures incurred 5 to 14 yr, 15 to 25 yr, and 25+ years previously. The coefficients $\theta_{15\text{-}24}$ and θ_{25+} represent the relative contributions of exposures 15 to 24 yr and 25 yr or more previously.

Estimates of the parameters for the exposure-age-concentration model and the exposure-age-duration model are given in Table 2. Within the three exposure-time windows, more recent exposures are given more weight. The ERR declines with attained age under the both the exposure-age-concentration and exposure-age-duration models. The effect of a given exposure level increases with decreasing exposure rate, indexed by either exposure concentration or duration of exposure. Because both the exposure-age-concentration and exposure-age-duration models

Table 2. Parameter estimates from BEIR VI models based on pooled updated miner data. Figure Captions

Exposure-age-duration model*		Exposure-age-concentration model*	
βx100	0.55	βx100	7.68
Time since exposure windows			
$\theta_{5\text{-}14}$	1.00	$\theta_{5\text{-}14}$	1.00
$\theta_{15\text{-}24}$	0.72	$\theta_{15\text{-}24}$	0.78
θ_{25+}	0.44	θ_{25+}	0.51
Attained age			
$\phi_{<55}$	1.00	$\phi_{<55}$	1.00
$\phi_{55\text{-}64}$	0.52	$\phi_{55\text{-}64}$	0.57
$\phi_{65\text{-}74}$	0.28	$\phi_{65\text{-}74}$	0.29
ϕ_{75+}	0.13	ϕ_{15+}	0.09
Duration of exposure		Exposure rate (WL)	
$\gamma_{<5}$	1.00	$\gamma_{<0.5}$	1.00
$\gamma_{5\text{-}14}$	2.78	$\gamma_{0.5\text{-}1.0}$	0.49
$\gamma_{15\text{-}24}$	4.42	$\gamma_{1.0\text{-}3.0}$	0.37
$\gamma_{25\text{-}34}$	6.62	$\gamma_{3.0\text{-}5.0}$	0.32
γ_{35+}	10.2	$\gamma_{5.0\text{-}15.0}$	0.17
		γ_{15+}	0.11

* Parameters estimated on the basis of the model $RR = 1 + \beta \times w^* \phi_a \times \gamma_z$ fit using two-stage method, where $w^* = w_{5\text{-}14} + \theta_2 w_{15\text{-}24} + \theta_3 w_{25+}$. Here, the subscript $_a$ denotes categories of attained age and subscript $_z$ denotes categories of either exposure duration categories (in years) or radon concentration categories (in WL).

provided equally good fits to the data, both models are of value in predicting radon lung cancer risks.

Several assumptions are required to extrapolate the BEIR VI risk models to the general population. First, it is assumed that the linear exposure-response relationship observed for miners also held at lower residential exposures. Support for this assumption is provided by consistency of the slope of the exposure-response relationship with a meta-analysis of residential radon case-control studies conducted by Lubin and Boice (1997). Further mechanistic support for this assumption is the direct damage to DNA associated with alpha particles released from radon progeny. Second, the modifying effect of exposure rate on risk is assumed to be negligible at residential exposure levels. Empirical support for this assumption is provided by the consistency of risk estimates based on the BEIR VI risk models with estimates based on a simple constant relative risk model, without the modifying effects of attained age or exposure concentration or duration of exposure.

Estimates of the lung cancer risk associated with residential exposure to radon are subject to a number of uncertainties, including uncertainty in the risk model, uncertainty in the extrapolation of the risk model to the general population, uncertainty in the relationship between exposure and dose, uncertainty in residential radon exposures, and uncertainty in the demographic data required for projecting population risk. In addition to identifying specific sources of uncertainty, the

BEIR VI committee conducted a formal analysis of uncertainty using the general methods developed by Rai and Krewski (1998). These results have subsequently been extended by Krewski *et al.* (1999b).

V. PARTICULATE AIR POLLUTION AND CARDIORESPIRATORY MORTALITY

Cohort study designs are often used to assess the association between community-based ambient air pollution concentrations and longevity (Dockery *et al.* 1993; Pope *et al.* 1995; Abbey *et al.* 1999). In one such study, volunteers of the American Cancer Society enrolled more than 1.2 million people in September of 1982 throughout the United States (Thun *et al.* 1995). Information on history of disease, demographic characteristics, and mortality risk factors were obtained from respondents. Vital status was monitored through the end of 1989. The association between concentrations of sulfate particles and longevity was examined in 144 communities for white members of the cohort, totalling 509,292 subjects (Burnett *et al.* 2001). The mean age at enrollment was 56.7 years, 5% of subjects were younger than 40 years, 5% were older than 75 years, and 56.3% of subjects were women. During the course of the seven years of follow-up, 39,474 (7.8%) subjects died. Standard statistical computing software programs (*e.g.*, SAS 1997) can be used for analysis if the assumption of statistical independence between subjects is appropriate.

Health responses, however, often cluster by community, indicating that responses of subjects within the same community are more similar than responses of subjects in different communities. Failure to account for all the variation between community health outcomes even after controlling for mortality risk factors can lead to downward biased estimates of the uncertainty in the air pollution effect (Ware and Stram 1988). Additional bias can occur if the community mortality rates display spatial autocorrelation. Failure to account for spatial autocorrelation can also yield downward biased estimates of uncertainty in the air pollution effect on mortality and may suggest incomplete control for potentially confounding community-level factors with the variables of primary interest, such as air pollution (Miron 1984).

As part of a comprehensive re-analysis of the data from the Harvard Six-cites Study (Dockery *et al.* 1993) and the American Cancer Society Study (Pope *et al.* 1995), Krewski *et al.* (2000) developed new spatial methods for the analysis of data of this type. These methods included allowance for spatial autocorrelation within cites and within major airshed regions within the United States. The methods of analysis included random effects models to describe variation in mortality rates among cities, regional adjustment models reflecting differences among the seven regions considered, and spatial filtering models in which broad geographic patterns in air pollution levels or mortality rates (or both) were filtered out before the application of traditional methods of analysis for uncorrelated data (Krewski *et al.* 2002).

Subsequently, Burnett *et al.*, (2001) developed a regression survival model in which the residual community health responses are characterized by community-based stochastic variables called "random effects", after controlling for individual and community-level risk factors. The variance of the random effects represents the residual variation in mortality between communities. Broader spatial trends in

residual mortality are modelled by non-parametric smoothers of location in the deterministic component of the model. The complexity of the location surface is selected such that no apparent spatial autocorrelation is evident in the residuals.

Figure 3 illustrates the comparison between estimates of the air pollution-mortality association, given by the percentage increase in mortality associated with a 4.2 $\mu g/m^3$ increase in sulfate particulate matter, the interquartile range of the exposure distribution, based on a regression model assuming a subject's survival time was statistically independent, a model incorporating random community effects with no spatial modeling and a model incorporating both random community effects and spatial autocorrelation. The uncertainty in these estimates is given by $\pm 1.96 \times$ standard errors.

Although the air pollution association with mortality was similar for models assuming independent observations and that incorporating community random effects with no spatial modeling, the uncertainty in the estimate was much greater for the latter model specification, suggesting there was more variability in community mortality rates than could be explained by the simpler regression model. Further incorporation of a location surface into the random effects model reduced the air pollution effect, indicating correspondence at a broad spatial level between community mortality and sulfate pollution (see Figure 4). This result suggests that there are unexplained mortality risk factors that may be correlated with the spatial pattern of pollution.

We conclude that the simpler statistical model specification overestimated the air pollution association with mortality and underestimated its uncertainty.

VI. SUMMARY AND CONCLUSIONS

This article indicates that a range of empirical and biologically based models, ranging from simple to complex, have been used to describe the relationship

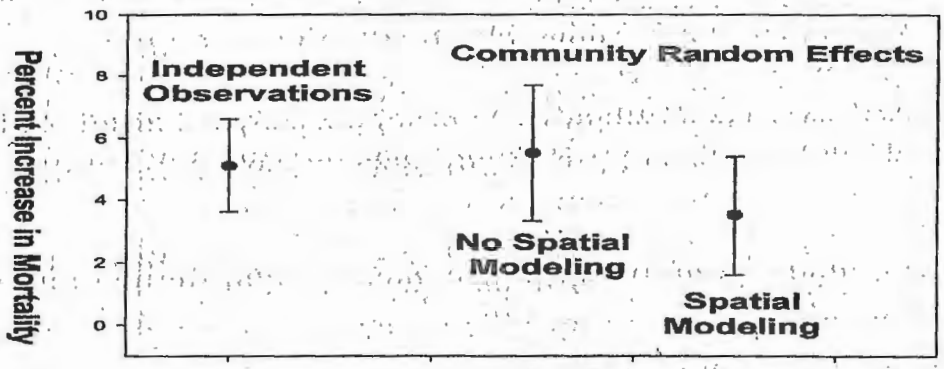
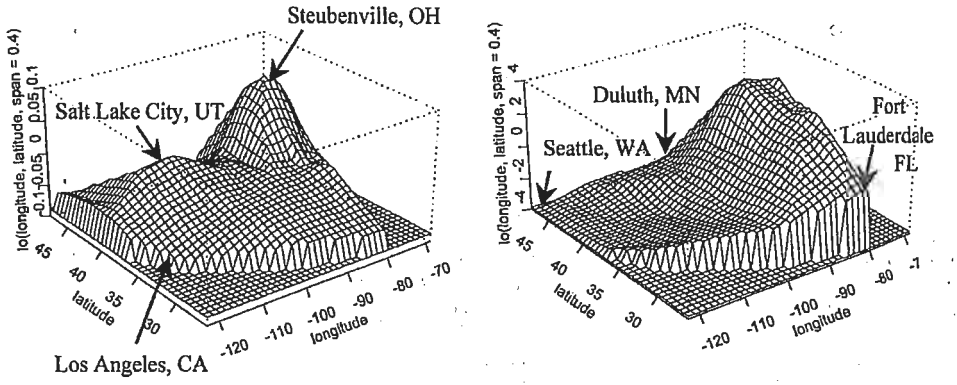


Figure 3. **Percentage increase in mortality associated with a 4.2 $\mu g/m^3$ increase in sulfate particulate matter, the interquartile range of the exposure distribution, based on a regression model assuming a subject's survival time was statistically independent (Independent Observations), a model incorporating community random effects with no spatial modeling (Community Random Effects — No Spatial Modeling) and a model incorporating both community random effects and spatial autocorrelation (Spatial Modeling).** The uncertainty in these estimates is given by +/-1.96 × standard errors as represented by error bars.



a) Spatial representation of community-specific mortality relative risk adjusted for individual risk factors

b) Spatial representation of particulate sulfate concentrations

Figure 4. **Nonparametric smoothed surface of mortality by latitude and longitude, adjusted for individual level covariates in American Cancer Society Study (panel a). Nonparametric smoothed surface of particulate sulfate concentrations by latitude and longitude (panel b).** Note, z-axis represents residuals from spatial random effects survival regression model.

between exposure or dose and health outcomes. Biologically based risk models are based on knowledge of the mechanisms by which toxic effects are induced. Since most toxicological processes involve an element of complexity, biologically based models tend to involve an element of complexity. The development of a biologically based risk model often involves revision of mechanistic hypotheses as poorly fitting models are refined in accordance with toxicological and epidemiological data, ultimately leading to a model that is consonant with a plausible mechanistic hypothesis. A mechanistic model (that has stood the test of several 'validation' efforts) has the advantages of enjoying biological plausibility, direct biological interpretation of key model parameters, and a degree of confidence in extrapolation beyond the range of the data.

Biologically based risk models for carcinogenesis and mutagenesis have received considerable attention by risk modellers, and have led to useful insights that would be hard to achieve otherwise. For example the two-stage clonal expansion model of carcinogenesis, which recognized the important roles of both mutation and cell proliferation in neoplastic transformation, demonstrates important differences in how agents that increase either mutation or cell proliferation rates could affect cancer risk in a dose-dependent manner. This model also offers valuable insight into the types of synergistic effects that can be expected as a consequence of joint exposure to two or more carcinogens.

Empirical models are more appropriate (indeed necessary) when a reasonable understanding of the fundamental biological processes underlying the induction of adverse health outcomes is lacking. Such models often provide a good fit to the data,

and can be reasonably expected to provide accurate estimates of risk within the range of exposures or doses studied. Empirical models can range from simple linear models used to extrapolate risks from higher to lower doses, to more complex models needed to adequately characterize intricate patterns in the available data. Given a compelling biological basis support for the assumption of low dose linearity, as is often argued for genotoxic effects, simple linear extrapolation may provide adequately accurate estimates of risk at low doses. In other cases, such as recently developed models used to describe covariation between spatial patterns of exposure and health risk, empirical models can be much more complex. Such complexity arises not because of the (not yet understood) complexity of toxic mechanisms leading to adverse health outcomes, but rather because of complex patterns in the observed data that need to be described.

Although empirical risk models do not offer much insight into the biological processes underlying the induction of toxic effects, they can provide valuable information on the factors that affect or modify risk. The exposure-age-duration and exposure-age-concentration models used by the BEIR VI Committee to describe the association between radon and lung cancer clearly demonstrate changes with risk in attained age and duration or concentration of exposure to radon gas. Unlike the two-stage clonal expansion model, however, such empirical models cannot distinguish between genotoxic and nongenotoxic effects of carcinogenic agents.

The ability of spatial risk models used to relate geographic variation in risk with concomitant variation in environmental exposures demonstrates the importance of accounting for spatial autocorrelation in such data. Failure to acknowledge spatial patterns that result in spatial autocorrelation can lead to biased estimates of risk, and overstatement of the level of precision of associated with risk estimates. For such analyses, the use of complex spatial models cannot be avoided if appropriate inferences about risk are to be drawn.

Not all risk models need be thought of as either empirical or biologically based. The dose-response used to describe developmental risks in rodents acknowledges the presence of intra-litter correlation, a biological characteristic of such data due to genetic similarity and a common intra-uterine environment. Such semi-empirical models, which also permit a multivariate treatment of fetal mortality (prenatal death) and morbidity (fetal malformations), can provide some biological insight into certain aspects of the data, without a full understanding of the biological pathways leading to developmental anomalies.

Preference for a simple or complex risk model depends on the risk assessment objectives, and the context in which risk modelling is undertaken. If an estimate of mutagenic or carcinogenic potency is required with an essentially linear dose response curve, a simple linear model may suffice, providing an estimate that is virtually indistinguishable from that obtained from a more sophisticated biologically motivated model. On the other hand, extrapolation of a curvilinear dose-response relationship to low doses may be more accurately accomplished using a model that incorporates the key biological elements involved in malignant transformation, particularly when cell proliferation, which may demonstrate a nonlinear association with dose, is an important element. Even when no attempt is made to develop a biologically based risk model, complex data structures may necessitate the use of complex empirical models.

We conclude that both simple and complex risk models can play useful roles in risk assessment. Complexity can arise not only because of attempts to describe complex biological phenomena, but also because of inherently complex patterns in the data. Risk modellers need to be aware of empirical and biologically based models of varying degrees of complexity, and select an analytic strategy capable of addressing the risk assessment objectives motivating the development of an appropriate risk projection model.

ACKNOWLEDGMENTS

This paper was originally presented at the Workshop on Future Research for Improving Risk Assessment Methods held in Aspen, Colorado, August 15 to 17, 2000. This work was supported in part by the Natural Sciences and Engineering Council of Canada, the Toxic Substances Research Initiative of Health Canada, and the U.S. Health Effects Institute. D. Krewski is the NSERC/SSHRC McLaughlin Chair in Population Health Risk Assessment at the University of Ottawa.

REFERENCES

Abbey DE, Nishino N, McDonnell WF, *et al.* 1999. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. Am J Respir Crit Care Med 159:373-82

Anderson EL and The Carcinogenic Assessment Group of the US-EPA. 1983. Quantitative approaches in use to assess cancer risk. Risk Anal. 1(4):277-95

Bernstein L, Lee FD, McCann J, *et al.* 1982. An empirical approach to the statistical analysis of mutagenesis data from the *Salmonella* test. Mutation Res 97:267-81

Burnett R, Ma R, Jerrett M, *et al.* 2001. The spatial association between community air pollution and mortality: A new method of analyzing correlated geographical cohort data. Environ Health Perspect (*in press*)

Catalano PJ and Ryan LM. 1994. Statistical methods in developmental toxicology. In: Patil GP and Rao CR (eds), Handbook of Statistics, vol 12, pp 507-32. North-Holland, Amsterdam, NY, USA

Denes J and Krewski D. 1995. An exact representation of the generating function for the Moolgavkar-Venzon-Knudson two-stage model of carcinogenesis with stochastic stem cell growth. Mathematical Biosciences 131:185-204

Dewanji A, Goddard MJ, Krewski D, *et al.* 1999. Two stage model for carcinogenesis: number and size distributions of premalignant clones in longitudinal studies. Mathematical Biosciences 155:1-12

Dockery DW, Pope III CA, Xu X, *et al.* 1993. An association between air pollution and mortality in six U.S. cities. New Engl J Med 329:1753-9

Fung KY, Marro L, and Krewski D. 1998. A comparison of methods for estimating the benchmark dose based on overdispersed data from developmental toxicity studies. Risk Anal 18:329-42

Gaylor DW and Gold LS. 1998. Regulatory cancer risk assessment based on a quick estimate of a benchmark dose derived from the maximum tolerated dose. Reg Tox Pharm 28:222-5

Gaylor DW, Ryan L, Krewski D, *et al.* 1998. Procedures for calculating benchmark doses for health risk assessment. Reg Tox Pharm 28:150-64

Goddard MJ and Krewski D. 1995. The future of mechanistic research in risk assessment: Where are we going and can we get there from here? Toxicology 102:53-70

Goddard MJ, Murdoch DJ, and Krewski D. 1995. Temporal aspects of risk characterization. Inhalation Toxicol 7:1005-18

IARC (International Agency for Research on Cancer). 1988. Man made Mineral Fibers and Radon. IARC Monographs on the Evaluation of Carcinogenic Risk to Humans, p 1 :43. Lyon, France

Kodell RL, Krewski D, and Zielinski JM. 1991. Additive and multiplicative relative risk in the two-stage clonal expansion model of carcinogenesis. Risk Anal 11:483-90

Krewski D and Leroux BG. 1993. AMESFIT: A microcomputer program for fitting linear-exponential dose-response models in the Ames Salmonella essay. Environ .Molecular Mutagenesis 22:78-84

Krewski D and Zhu Y. 1995. A simple data transformation for estimating benchmark doses in developmental toxicity experiments. Risk Anal 15:29-39

Krewski D, Gaylor D, and Szyszkowicz M. 1991. A model-free approach to low-dose extrapolation. Environ Health Perspect 90:279-85

Krewski D, Leroux BG, Creason J, *et al.* 1992. Sources of variation in the mutagenic potency of complex chemical mixtures based on the Salmonella/Microsome assay. Mutation Res 276:33-59

Krewski D, Leroux BG, Bleuer SR, *et al.* 1993. Modeling the Ames Salmonella/Microsome Assay. Biometrics 49(2):499-510

Krewski D, Gaylor DW, and Lutz, WK. 1995. Additivity to background and linear extrapolation. Low Dose Extrapolation of Cancer Risks: Issues and Perspectives (ed. by Olin S, Farland W, Park C, *et al.*), pp 105-121. ILSI Press, Washington, DC, USA

Krewski D, Zhu Y, and Fung K. 1999a. Benchmark doses for developmental toxicants. Inhalation Toxicol 11:579-91

Krewski D, Rai SN, Zielinski JM, *et al.* 1999b. Characterization of uncertainty and variability in residential radon cancer risks. Ann-N-Y-Acad-Sci. 895:245-72.

Krewski D, Burnett RT, Goldberg MS, *et al.* 2002. Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality, Part I & II. 129-240. Cambridge, Massachusetts, Health Effects Institute. Reanalysis of the Harvard Six Cities Study and the ACS Study of Particulate Air Pollution and Mortality: A Special Report of the Institute's Particle Epidemiology Reanalysis Project

Krewski D, Burnett R, Goldberg MS, *et al.* 2001. Overview of the re-analysis of the Harvard Six-cities study and American Cancer Society study of particulate air pollution and mortality. J Toxicol Environ Health (*in press*)

Leroux BG and Krewski D. 1993. Components of variation in mutagenic potency values based on the Ames Salmonella test. Canadian J Statistics 21:448-59

Lewtas J, Claxton LD, Rosenkranz HS, *et al.* 1992. Design and implementation of a collaborative study on the mutagenicity of complex mixtures in Salmonella typhimurium. Mutation Res 276:3-9

Lubin JH and Boice Jr JD. 1997. Lung cancer risk from residential radon: Meta-Analysis of eight epidemiologic studies. J Natl Cancer Inst 89:49-57

Lubin JH, Boice Jr JD, Edling C, *et al.* 1994. Radon and Lung Cancer Risk: A Joint Analysis of 11 Underground Miners Studies. National Institute of Health, National Cancer Institute. NIH Publication No. 94-3644. U.S. Department of Health and Human Services, Washington, DC, USA

Lubin JH, Boice Jr JD, Edling C, *et al.* 1995. Lung cancer in radon-exposed miners and estimation of risk from indoor exposure. J Natl Cancer Inst 87:817-27

Lubin JH, Boice Jr JD, Edling C, *et al.* 1997. Estimating lung cancer mortality from residential radon using data for low exposures in miners. Radiat Res 147(2):126-34

Margolin BH, Kaplan NL, and Zeiger E. 1981. Statistical analysis of the Ames Salmonella/Microsome test. Proc Nat Acad Sci 78:3779-83

Miron J. 1984. Spatial autocorrelation in regression analysis: a beginner's guide. In: Gaile GL and Willmott CJ (eds), Spatial Statistics and Models. Reidel Publishing Company, Boston, MA, USA

Moolgavkar SH and Luebeck G. 1990. Two-event model for carcinogenesis: Biological, mathematical and statistical considerations. Risk Anal 10:323-41

Moolgavkar SH, Luebeck EG, Krewski D, *et al.* 1993. Radon, cigarette smoke, and lung cancer: A re-analysis of the Colorado uranium miners' data. Epidemiology 4:204-17

Moolgavkar SH, Krewski D, Zeise L, *et al.* 1999. Quantitative Estimation and Prediction of Human Cancer Risks. IARC Scientific Publication No. 131. International Agency for Research on Cancer, Lyon, France

NRC (National Research Council). 1980. The Effects on Populations of Exposure to Low Levels of Ionizing Radiation (BEIR III). National Academy Press, Washington, DC, USA

NRC (National Research Council). 1983. Risk Assessment in the Federal Government: Managing the process. National Academy Press, Washington, DC, USA

NRC (National Research Council). 1988. Health Risks of Radon and other Internally Deposited Alpha-emitters: (BEIR IV). National Academy Press, Washington, DC, USA

NRC (National Research Council). 1994. Science and judgement in risk assessment. National Academy Press, Washington, DC, USA

NRC (National Research Council). 1999. Health Effects of Exposure to Radon (BEIR VI). National Academy Press, Washington, DC, USA

Olin S, Farland W, Park C, *et al.* 1995. Low-dose extrapolation of cancer risks: Issues and perspectives. Chapters 1-3, ILSI Press, Washington, DC, USA

Pope CA, Thun MJ, Namboodiri MM, *et al.* 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. Am J Respir Crit Care Med 151:669-74

Rai SN and Krewski D. 1998. Uncertainty and variability analysis in multiplicative risk models. Risk Anal 18:37-45

SAS PROC PHREG, SAS/STAT Software: Changes and Enhancements through Release 6.12. 1997. ISBN 1-55544-873-9. SAS Institute Inc., Cary, NC, USA

Seilken RL. 1987. Cancer dose-response extrapolations. Environ Sci Tech 21:15-21

Tan WY. 1991. Stochastic Models of Carcinogenesis. Marcel Dekker, NY, NY, USA

Thun MJ, Day-Lally CA, Calle EE, *et al.* 1995. Excess mortality among cigarette smokers: Changes in a 20-year interval. Am J Public Health 85:1223-30

U.S. Presidential/Congressional Commission on Risk Assessment and Risk Management. 1997. Framework for Environmental Health Risk Management (Volumes 1&2) Presidential/Congressional Commision, Washington, D.C.

Ware JH and Stram DO. 1988. Statistical issues in epidemiologic studies of the health effects of ambient air pollution. Can J Statist 16:5-13

Zeise L, Wilson R, and Crouch AC. 1987. Dose-response relationships for carcinogenesis: A review. Environ Health Perspect 73:259-308

Zheng Q, Lutz WK, and Gaylor DW. 1997. A carcinogenesis model describing mutational events at the DNA adduct level. Mathematical Biosciences 144:23-44

Zielinski JM, Krewski D, and Kodell RL. (2001) Interactions between two carcinogens in the two-stage clonal expansion model of carcinogensis. J Epidemiol Biostat. 6(2):219-28

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1375-1395 (2002)

Toxicokinetic Models: Where We've Been and Where We Need to Go!

Melvin E. Andersen* and James E. Dennison
Department of Environmental Health, Colorado State University, Ft. Collins, CO 80523–1680

ABSTRACT

Toxicokinetic (TK) models have many uses, some of which are now regarded as almost routine, in areas related to pharmaceutics, toxicology, and chemical risk assessment. These TK models span a range from simple empirical curve-fitting analyses of blood/tissue time courses to physiologically based toxicokinetic (PBTK) models that incorporate anatomical, physiological, and biochemical properties of laboratory animals and humans. While the PBTK models require more effort to develop and validate than do data-based compartmental models, the biological detail in these descriptions permits extrapolation to different doses, different exposure conditions, and different species, including humans. Efforts to develop PBTK models are frequently rewarded with reduced work on subsequent compounds, since the physiologic structure, once developed for a particular life stage and class of compounds, is not expected to change for other compounds in the class. A review of the literature shows that TK models have had many uses in occupational health and industrial hygiene; however, they have not been widely or systematically employed in these disciplines. This overview discusses the history of uses of TK models in occupational health areas and suggests future possibilities for these models. Notably, TK models and especially PBTK models could play much more important roles in establishing occupational exposure limits such as the U.S. Occupational Safety and Health Administration's Permissible Exposure Limits based on either animal or human studies; in assessing the range of susceptibility of diverse human populations based on individual variability; in interpreting epidemiological and biomarker studies for various exposure situations; in developing common methods to assess risks for exposures to both the general population and to worker populations; and in assessing exposures to chemical mixtures.

Key Words: pharmacokinetic, risk assessment, toxicokinetic (TK) modeling, PBPK, PBTK.

* Corresponding author: Center for Environmental Toxicology and Technology & International Center for Risk Assessment, Colorado State University, Ft. Collins, CO 80523–1680; Tel(voice):970-491-8253, Tel(fax):970-491-8304; Melvin.Andersen@colostate.edu

1080-7039/02/$.50

INTRODUCTION

Pharmacokinetics (PK) has traditionally described the absorption, distribution, metabolism, and elimination (ADME) of drugs and chemicals from the body in various animal species and in humans. If studies are done with chemicals that are not intended for use as drugs and have the potential for adverse responses, the field of study is sometimes referred to as "toxicokinetics." In this paper, we use the terminology "toxicokinetic models," although the terms pharmacokinetic and toxicokinetic are frequently used interchangeably with many toxic compounds. TK models apply a set of equations to capture the time course relationships of the chemical in blood and tissues in the body. One goal of much TK modeling is to elucidate the biological and chemical factors that lead to the observed time course relationships. Structures of these TK models vary depending on the level of detail included in describing the animal or human and the detail in describing the biological interaction of the chemical and its metabolites within the body. With continuing improvements in computational resources over the last five decades, TK modeling has evolved from simple empirical fitting of time course curves to development of more biologically structured models that are becoming widely used in assessing risks from exposure of the general population to various chemicals.

The main question in developing TK models is why bother to do it at all, *i.e.*, what is the purpose of developing a PK model for a specific application. To paraphrase this question, we ask "What do we expect to achieve through use of these models that we cannot derive from other kinds of studies or by casual inspection of time course curves from concentrations of occupational chemical or metabolites in blood or urine?" TK models have not been widely used in the occupational health/industrial toxicology fields, although they have found some specific areas of application over the last 30 years. One of the main goals of this paper is to ask what roles TK models might play in occupational health and what changes need to occur in occupational health research to encourage application of these models.

APPLICATIONS OF PK AND TK MODELING

Pharmaceutics

The major developer of PK data and PK models has been the pharmaceutical industry. In this industry, the rationale for PK model development is that efficacy of drug action depends on the concentrations of drugs in the body at target sites. Tablet dose is not necessarily directly related to active site concentrations, so studies are conducted with every drug and drug formulation to understand the factors that regulate the time course of active concentrations of drugs in the blood or plasma. These factors include rates of absorption, rates of distribution throughout the body to target tissues and to drug storage sites, rates of metabolism, and rates of elimination from the body. These time course data are organized within PK models that contain relatively few parameters (Wagner 1981) that allow predictions of time course for different dosing situations and aid in developing dosing strategies to optimize drug therapy. In general, PK models used in the pharmaceutical industry are multicompartmental models where the specific compartments have no direct anatomical correspondence with specific organs or groups of organs. With one

particular class of drugs, the antineoplastic drugs, a great deal of work has been conducted to develop more biologically structured models for the distribution and cytotoxic action of these highly toxic drugs within the body (Himmelstein and Lutz 1979).

Toxicology

Toxicologists frequently face the decidedly difficult task of inferring possible risks to human populations from toxicity studies in which laboratory animals are exposed to chemicals at high daily doses, for their entire lifetime, by routes of exposure different from normal exposure routes in humans. A significant concern among toxicologists is whether the much higher doses in the animal studies might lead to differences in kinetics compared to kinetic behavior at lower doses. Pharmacologists in the 1960s found that elimination of certain drugs, notably aspirin and ethanol, changed markedly as the administered dose increased (Welling 1986). As with most compounds, the elimination of these two compounds required metabolism, catalyzed by enzymes. The body contains a limited amount of these enzymes. Metabolic capacities become saturated if the concentrations of the drugs get too high. After saturation, the kinetics of the compound change and the drug concentrations in the body increase rapidly with small increases in dose. With these two drugs, these rapid increases in blood concentrations can lead to unwanted side effects — inebriation with ethanol and ototoxicity with aspirin, among other effects.

In the 1970s, industry scientists at Dow Chemical Company began to introduce TK studies routinely in the evaluation of the toxicity of commodity chemicals — such as vinyl chloride and vinylidene chloride, two important monomers in the plastics industry. The TK models used initially with various industrial compounds also relied on compartments whose structure was determined by curve fitting of models to the time course data. Unlike ethanol and aspirin, metabolism increases the toxicity of these compounds. Despite these differences in modes of toxic action, the common issues that led to PK/TK model implementation in both pharmaceutics and toxicology was the recognition of the need to equate responses — either beneficial, therapeutic responses or adverse, toxicological responses — with measures of dose within the body. The measure of tissue dose chosen in most cases is the blood or plasma concentration of the chemical under the assumption that tissue concentrations are generally proportional to blood concentrations.

Risk Assessment

Beginning in the 1980s, TK modeling applications in toxicology exploded as the methods were increasingly brought to bear in supporting new mechanistically based approaches to chemical risk assessment. These risk assessments required methods to assist in extrapolating from the conditions used in toxicity studies to various untested situations. These extrapolations are from high doses to low doses, from one dose route to other doses routes, and from the laboratory animals to humans (Clewell and Andersen 1985). Strategies have been developed more recently to make these extrapolations based on target tissue dose rather than administered dose. In these extrapolations, measures of toxicity such as the Benchmark Dose (BMD) would be expressed in relation to tissue dose occurring under these expo-

sure conditions in the test animals. Then, human TK models would be used to assess exposure conditions that would lead to similar tissue doses of toxic compounds in humans exposed by relevant routes.

The PK models used for pharmaceutical materials are generally suited to interpolation because of the ill-defined nature of the compartments in the model. Extrapolation to untested conditions required a more explicit structural formulation provided by PBTK models with biologically realistic compartments and metabolic parameters for pathways of metabolism. PBTK models have become widespread in toxicology (Leung 1991) and their importance has grown with the new emphasis on mode of action and dosimetry as centerpieces of the U.S. Environmental Protection Agency's (USEPA's) approaches to chemical risk assessment (USEPA 1994; USEPA 1996). After their introduction into the risk assessment community in the mid-1980s, PBTK modeling approaches underwent a period of intense scrutiny, including examination of appropriate model structures, assessment of parameter sensitivity, and development of methods to assess the influence of population variability on dose predictions with these models. The value of the models was not really at issue: it was widely agreed that tissue dose is the preferred measure on which to index the toxic responses. The debate was about the correct manner in which to implement these dosimetry concepts from the TK and PBTK models into an existing methodology that was based on much simpler empirical analyses of the relationship between response and administered doses. During this period of evaluation and critical review of these modeling approaches from 1985 to 1995, methods were applied to assess distributions of tissue doses expected for particular populations. It also became apparent that factors related to differences in dose caused by individual differences in metabolism or clearances, including polymorphisms, disease conditions, and multicompound exposure, could be easily accommodated by these PTPK models to better assess the range of tissue doses expected in a diverse human population (Clewell 1995).

Occupational Health/Industrial Hygiene

In contrast to what is common in the pharmaceutics, toxicology and risk assessment communities, TK models, especially PBTK models, are not as commonly applied in industrial hygiene/occupational health studies. Why is this? Is there less need to evaluate the relationship of expected adverse worker responses based on tissue dose than in these other disciplines? Or, are the extrapolative capabilities of PBTK modeling unnecessary because, in industrial toxicology, we usually study defined human populations directly? Either of these points may be valid in particular situations. However, an alternative explanation, one favored by the authors here, is that the potential of TK modeling approaches simply has not been actively pursued in occupational health and industrial hygiene. These models offer promise in several areas — establishing Occupational Exposure Levels, evaluating risk factors for susceptible individuals, and achieving more uniform methods for risk assessments for workers and for general population, among other things. How might the promise of these models become realized? To answer this question, we first note the different types of TK models used in various studies and then note the areas where further development and application of the models appear promising for occupational health/industrial hygiene activities.

TK MODELING APPROACHES

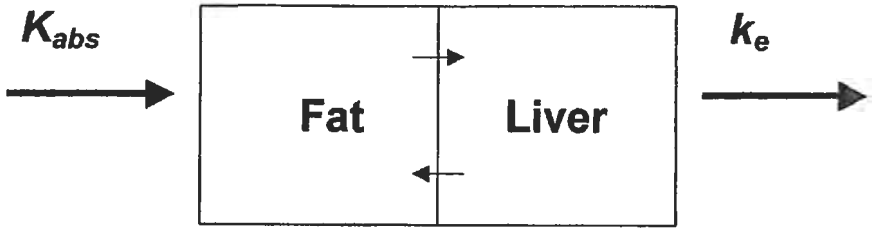
Empirical TK Models: Just the Facts!

For the purposes of simplification, there seem to be three general types of TK models that are distinguished by their anatomical and physiological detail. The simplest models rely on a straightforward mathematical analysis of the time course data in populations of interest. Many of the kinetic data developed in humans and animals prior to 1960 were analyzed in this way. A more recent example was with inhalation exposures to carbon monoxide (CO). Time course blood carboxyhemoglobin (HbCO) concentrations after controlled human exposures to CO were analyzed to estimate the maximum HbCO at the end of exposure and to determine whether the elimination time course curve was exponential or polyexponential (Peterson and Stewart 1975). The curves provide direct estimates of rates of elimination and elimination half-lives, but do not reveal the biological factors that determine these time courses. The challenge in relying on the data alone arises when attempting to extrapolate these relationships to a broader set of situations. What would happen for longer or shorter exposure durations? What role does health status, breathing rate, or oxygen concentrations play in regulating the blood HbCO concentrations after CO exposures? The empirical analysis remains silent on these questions.

Another example of a more empirically based TK model was the description of the dose-dependence of accumulation of dioxin-like compounds in liver in humans and in laboratory animals (Carrier *et al.* 1995a). The model structure (Figure 1) for sequestration included terms for the maximum proportion of body burden found in the liver f_h^{max}, the minimum proportion found in the liver f_h^{min}, and a "binding constant", K, that determined the body burden at which there was a half-maximal increase in the proportional body burden found in liver. This empirical model structure reproduced the dose-dependence, *i.e.*, as the body burden increased, a larger proportion of the body burden of the dioxin-like compounds was found in the liver. For both laboratory animals and humans, the fitted relationships permit interpolation within the dose ranges studied. However, the model structure does not speak to the biological characteristics that determine f_h^{max}, f_h^{min}, and K. If the biological determinants of these parameters were established, it would be possible to extrapolate to a broader set of exposure situations. These same authors developed a model for the distribution of these dioxin-like compounds that was more physiologically realistic (Carrier *et al.* 1995b). Another simple PBTK model was developed by Salvan and referred to as a Minimal Physiological Toxicokinetic (MPTK) model. This model examined the effect of uncertainty in model parameterization in the absence of protein binding (Salvan *et al.* 1999).

Data-Based Compartmental Models - Creating a Discipline

The models used today in the pharmaceutical industry and most pharmaceutical research have been defined as data-based compartmental models. These models trace their ancestry back to pioneering work by Teorell in the 1930s (Teorell 1937a; Teorell 1937b) and are now well described in various pharmacokinetic textbooks (Gibaldi and Perrier 1982). As with the empirical models, constants in the data-based compartmental models are fit to the time course data. These data-based



$$f_h(C_b) = f_h^{\min} + \frac{(f_h^{\max} - f_h^{\min}) * C_b}{K + C_b}$$

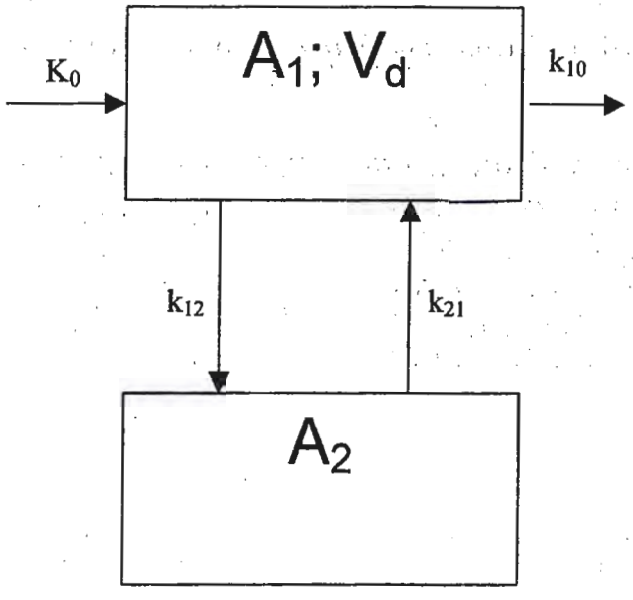
Where K_{abs} = absorption rate constant
k_e = elimination rate constant
$f_h(C_b)$ = fraction of body burden in liver
f_h^{min} = minimum fraction of body burden in liver
f_h^{max} = maximum fraction of body burden in liver
K = equilibrium constant
C_b = concentration in body

From Carrier (1995a).

Figure 1. An empirical model for dose-dependent dioxin sequestration in the liver.

compartmental models are based on a specific representation of the body (Figure 2) in relation to a central compartment, where sampling usually occurs, and peripheral storage compartments. These models took advantage of properties of the solution of the set of differential equations to estimate model parameters, such as volumes of distribution, intercompartmental transfer rate constants, elimination rate constants, *etc.* from the experimental data. The development of these compartmental models brought structure and rigorous mathematical and statistical methods to the analysis of these time course data. These models were brought to toxicology by several scientists in the 1970s including Gehring and colleagues at Dow (McKenna *et al.* 1977; Watanabe and Gehring 1976, and O'Flaherty 1981).

A good example of the use of compartmental models with an organic solvent was the analysis of the time course of styrene in rats after exposure at inhaled concentrations of 80 to 1200 ppm (Ramsey and Young 1978). One of the most elegant applications of this kind of modeling in toxicology was the development of two-



A_1: amount in compartment 1
A_2: amount in compartment 2
V_d: volume of distribution
K_0: uptake rate
k_{10}: rate constant for elimination
k_{12}: intercompartmental transfer rate constant from compartment 1 to 2
k_{21}: intercompartmental transfer rate constant from compartment 2 to 1

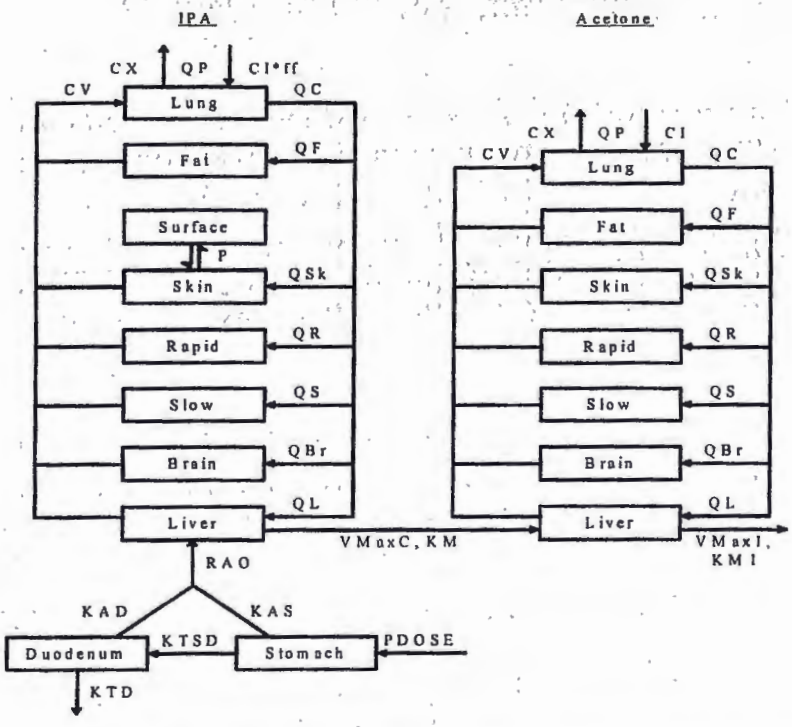
Figure 2. **A Two-Compartment PK model with intake and elimination from the central compartment.**

compartment models for metabolism of inhaled gases in rats in so-called "closed chamber" studies (Filser and Bolt 1981; Filser and Bolt 1979). These studies provided unprecedented accuracy in assessing the total distribution of chemicals into a living animal and assessing the kinetic characteristics of metabolism. In the closed chamber compartmental model, the sampling is done on the air phase and metabolism and storage occurs in the animal phase of the experimental system. In TK parlance, this is a two-compartment model with metabolism occurring in the peripheral, or deep, compartment. A recent example of insights derived from these data-based compartmental models in support of risk assessments was with hydrofluorocarbons (Emmen *et al.* 2000).

PBTK Models — Increasing Model Complexity for Broader Utility

A distinct advantage of the empirical and data-based compartmental models is simplicity and the relative ease of conducting statistical analysis for formal estimation of model parameters. A major disadvantage is the lack of a direct correspondence between model parameters and anatomical, physiological, or biochemical characteristics of the human or animal. PBTK models are the reverse. Their advantages are derived from the direct correspondence of model parameters with specific biological processes. Their disadvantage is that providing this level of detail requires models with significantly more parameters that are not easily evaluated with conventional statistical methods. Despite the increased complexity, these PBTK models are increasingly used because they permit many extrapolations and because the basic structure, once developed, is often chemical independent. The physiological and anatomical detail required to develop a PBTK model for a particular lifestage and type of compound may require a good deal of work and acute attention to validation. By validation, we mean that the model adequately predicts, as determined through visualization or by an appropriate statistical technique, data collected in experiments that were not used to develop model parameters originally. In practice, after application to these validation data sets, all the data should be evaluated with the model structure to get a refined parameter set. Once in place, the basic structure can be applied to any number of other compounds within a broad class (Andersen *et al.* 1999).

In PBTK models, compartment volumes correspond to specific organs or groupings of organs within the animal. Blood flows to each compartment, anatomical relationships and other physiological parameters necessary to build the model are based, to the extent possible, on measured values. However, these models are all simplifications of a more complex biology. The skill in developing any model, including these PBTK models, is purposeful simplification — ensuring that the model captures important aspects of the system that determines delivery of compounds to target tissues. Often, groups of organs that serve as storage sites are lumped together based on the ratio of perfusion rates divided by their storage capacities, defined by their volume multiplied by their partition coefficient. Usually, the major compartments necessary in the TK models are target tissues for toxicity, major storage tissues, organs of metabolism or elimination, and routes of uptake into the body or contact sites for more direct acting compounds. For example, the PBTK model (Figure 3) recently developed for isopropanol (IPA) includes compart-



CX: concentration in mixed-exhaled air	P: absorption from skin
CI: concentration in inhaled air	PDOSE: dose (by oral gavage)
QC: cardiac output	KAS: absorption rate from stomach
QSk: blood flow to skin	KAD: absorption rate from duodenum
QS: blood flow to slowly perfused tissues	KTD: elimination rate from duodenum
QL: blood flow to liver	KTSD: mass transfer rate between stomach and duodenum
CV: concentration in mixed venous blood	
QP: pulmonary ventilation	RAO: combined absorption rate from stomach and duodenum
QF: blood flow to fat tissue	
QR: blood flow to richly perfused tissues	VmaxC(1): Vmax for IPA or for acetone
QBr: blood flow to brain	KM(1): affinity constant for IPA or for acetone

Adapted from Clewell *et al.* (2001)

Figure 3. PBTK model of isopropanol and its metabolite, acetone.

ments for IPA distribution, metabolism to a primary metabolite, acetone, and distribution of acetone throughout the body (Clewell *et al.* 2001).

In addition to anatomical and physiological parameters, most PBTK models require various chemical and physical data about the potentially toxic chemicals. These parameters include the relative solubility of the compound in tissues versus blood (*i.e.*, the partition coefficients or non- specific binding) and metabolic constants for pathways of metabolism that create and clear toxic compounds. Metabolism, like the overall model itself, is usually simplified in accordance with the hypothesis testing being performed or other purposes of the model. At times, simple metabolic descriptions suffice (Ramsey and Andersen 1984) and at other times more complex formulations are required to capture detail of enzymatic mechanisms and juxtaposition of metabolic pathways in localized tissue regions (Johanson and Filser 1993). Kinetic constants for metabolism can sometimes be acquired from *in vitro* experiments or may be estimated by fitting the PBTK model to *in vivo* time course data. When estimating kinetic constants for metabolism from *in vivo* time course data, model sensitivity should also be examined to ensure that the time course behavior is actually sensitive to the parameters that are being varied in the fitting routine. Methods for estimating metabolic constants include *in vitro* studies with organelles, homogenates, cells, tissue slices, isolated perfused organs, and *in vivo* studies evaluating clearance of compounds from blood.

Human PBTK Models

Extrapolating the PBTK model to humans requires knowledge of the important anatomical, physiological, and metabolic parameters for individual humans and for the heterogeneous population that might be exposed to or work with potentially toxic compounds. Anatomical and physiological parameters for humans are generally fairly well established in the literature. Metabolic parameters are usually more difficult to predict. For well-characterized families of enzymes, these constants — the maximum velocity of metabolism (Vmax) and the affinity of the enzyme for substrate (Km) — have sometimes been estimated by using Km values for different species and by allometric scaling of Vmax values in relation to body weight $^{0.74}$. The pharmaceutical literature contains many studies on metabolic parameters involved in clearance of drugs in humans. In general, the enzymes most commonly active in oxidative drug metabolism are not the most common oxidative enzymes involved in metabolism of occupationally relevant compounds. Optimal approaches for assessing constants for metabolic pathways utilize human tissues or tissue preparations to assess these constants directly (Reitz *et al.* 1989). Extrapolations of metabolic parameters remain a challenge for interspecies scaling of PBTK models, but are a critical need in assuring confidence in the human models.

Contrasting Empirical and PBTK Models

Over the last 75 years, various investigators have developed PBTK/TK models for many drugs and occupational/environmental compounds. The earliest physiologically based description for chemical disposition within the body was provided for an inhaled vapor, diethyl ether, in 1924 (Haggard 1924a; Haggard 1924b). Further developments in modeling anesthetic gases occurred with work by Kety (1951),

Mapleson (1963), and Fiserova-Bergerova *et al.* (1974). The latter author first introduced pathways of metabolism in the PBTK models for occupational chemicals. Chemical engineers provided a series of PBTK models for cancer chemotheraputic drugs, starting with work on methotrexate (Bischoff *et al.* 1971). Short reviews of physiologically based models have appeared (Gerlowski and Jain 1983; Himmelstein and Lutz 1979; Leung 1991). Before embarking on a discussion of more current uses of these PBTK models, it is useful to contrast PBTK models developed for CO and tetrachlorodibenzo-*p*-dioxin (TCDD) with the more empirical modeling approaches noted earlier.

A physiological approach for assessing blood HbCO arising from metabolism of heme to CO provided an equation describing the physiological factors that regulate production, binding, and exhalation of the CO (Coburn *et al.* 1965). The Coburn-Forster-Kane relationship was subsequently used in a model for CO inhalation (Andersen *et al.* 1991) that predicted HbCO concentrations for rats or humans in both short-term and chronic inhalation exposures. The PBTK model also predicted the HbCO concentrations expected from inhalation of methylene chloride due to metabolism of this solvent by oxidation in the body. The ability to extrapolate across species, duration of exposure, and compound (from CO directly to CO generated by metabolic processes with methylene chloride) are all directly attributable to the physiological structure embedded within the PBTK model.

In the past 10 years, several PBTK models for TCDD have appeared (Andersen *et al.* 1997a; Andersen *et al.* 1997b; Andersen *et al.* 1993; Kohn *et al.* 1993). These descriptions include the disposition of the TCDD and its ability to induce specific TCDD-binding proteins, now identified as cytochrome P450 1A2, within the liver (Leung *et al.* 1990). In a way, these models are both toxicokinetic and toxicodynamic models because of the need to include protein induction. These models successfully account for the dose-dependent sequestration in liver at high-doses that formed the basis for the empirical TK model developed by Carrier (Carrier *et al.* 1995a; Carrier *et al.* 1995b). In the PBTK models of kinetics and induction, the induction of hepatic binding protein occurs due to the interaction of TCDD with the aryl hydrocarbon (Ah) receptor, binding of the Ah- TCDD complex with putative binding sites on DNA, and transcriptional activation of the CYP 1A2 gene. In the physiological model, there is no term for maximum sequestration, the f_h^{max} in the empirical model, or for K, the body burden at which there is half-maximal increase in liver sequestration. Nonetheless, the PBTK model is capable of estimating these composite parameters and, through the use of sensitivity analysis, to determine the biological processes within the body that determine the magnitude of these two parameters.

In what turned out to be slightly counterintuitive, Evans (Evans and Andersen 2000) showed recently that this composite K value was relatively insensitive to the dissociation constant for binding of TCDD to the TCDD binding protein. Instead, the "K" of the empirical model was mainly determined by the affinity of TCDD for the Ah receptor and the affinity of the Ah receptor-TCDD complex with sites on DNA. The f_h^{max} parameter was largely dependent on the maximum increase in CYP 1A2 and inversely related to the solubility of TCDD in fat and the dissociation constant between TCDD and the CYP1A2 binding protein. In addition, the f_h^{max} in a PBTK model was noted to occur where there was maximal induction of the

binding protein with a low degree of saturation, not when the binding protein was fully occupied. PBTK models for TCDD are still under development by several investigators. They have served as important tools for determining the importance of biological parameters through sensitivity analysis and for prioritizing mechanistic studies to aid in improving fidelity of the model with a range of biological observations. These PBTK models are extensively discussed in the dose-response assessment chapter of USEPA's re-assessment of the risks of TCDD and related compounds (USEPA 2000).

USES OF PBTK MODELS IN RISK ASSESSMENT

The Role of Dosimetry Models in Contemporary Risk Assessments in the United States

The new directions in risk assessment are increasingly emphasizing two concepts — mode of action and dose metrics for toxicity. The mode of action is defined as the series of key (necessary, but not sufficient) events and processes starting with the interaction of an agent with a biological target, through the functional and anatomical changes eventually resulting in adverse responses in the organism. The dose metric is that measure of target tissue dose that is most closely related to the ensuing adverse responses. Dose metrics include concentration of the parent compound or specific metabolites in blood or tissues, area under blood/tissue concentration curves, or measures of early responses, such as enzyme induction patterns with TCDD. The dose metric must be defined based on the mode of action. Optimally, the mode of action statement should contain information about the biological processes disrupted by the toxic chemical and a statement of the form of chemical involved in the deleterious interactions. PBTK models fit in here because they can be used to calculate the expected values of the dose metrics under a wide variety of conditions in laboratory animals, in individuals, and in human populations. These models have been most attractive in those situations where human risks are derived from animal toxicity studies and the need to extrapolate from laboratory species to a human population requires a more detailed mechanistic understanding of the kinetic behavior. In a nutshell, mode of action, in and of itself, is used to defend methods of low dose extrapolation — linear, nonlinear, or threshold. Dosimetry models then form the basis of the approaches for quantitative extrapolation to low doses and to humans.

The first human health risk assessment that employed a PBTK model was proposed for methylene chloride (Andersen *et al.* 1987). This PBTK model was developed to examine the dose dependence of liver and lung tumors in mice. To accomplish this task, the model included information on metabolism of methylene chloride by oxidation and by glutathione conjugation, concentration of metabolites, including CO, interspecies differences in rates of metabolism, and multiple exposure routes. The dose metric most closely correlated with tumor formation was metabolism by glutathione transferase, a conclusion that has been amply corroborated by subsequent mechanistic studies (Green 1997). The risk assessment based on tissue doses of the glutathione reaction products indicated that humans would be at considerably lower risk than were the mice for equivalent exposure concentra-

tions. Because the use of these methods represented a novel departure from the risk assessment defaults in the mid-1980's, the National Academy of Sciences sponsored a workshop and constituted a workgroup to determine if these modeling approaches were mature enough for more routine use in chemical risk assessments. The deliberations of this group, including the group of papers from the public meeting, appeared as a volume in the Drinking Water and Health Series (NRC 1987).

The U.S. Occupational Safety and Health Administration (OSHA) used similar PBTK models to perform a risk assessment for occupational exposure to methylene chloride that supported their decision to promulgate a Permissible Exposure Limit (PEL) of 25 ppm (OSHA 1997). However, dosimetry models are much more commonly employed for assessments with environmental exposures to the general population. The USEPA actively encourages the use of PBTK modeling in the risk assessment process (USEPA 1996). Some examples where PBTK models have been used are in the Integrated Risk Information System's (IRIS) Pilot Project with vinyl chloride (USEPA 1999a) and with ethylene glycol monobutylether (EGBE) (USEPA 1999b). The dose metric for vinyl chloride was the amount of vinyl chloride metabolized to epoxides within target tissues. Using the PBTK model to calculate this dose metric, the USEPA developed a Reference Dose (RfD), Reference Concentration (RfC), cancer slope factors, and inhalation unit risks. The reassessment of dioxin risks also included use of PBTK models (USEPA 2000). The proposed revisions to US EPA's cancer risk assessment guidelines promote application of mode of action and dosimetry concepts in risk assessments. Several groups have conducted risk assessments following these proposed guidelines to assess the issues raised by this new procedure. The compounds evaluated include chloroform (ILSI 1997) and formaldehyde (CIIT 1999). Dosimetry models have also been developed for a set of inhaled compounds — vinyl acetate (Bogdanffy *et al.* 1999), acrylic acid (Frederick *et al.* 1998), and methylmethacrylate (Andersen *et al.* 1999) — that damage the nasal olfactory epithelium in exposed rodents. This latter group of models, along with a dosimetry model for formaldehyde (CIIT 1999), offers considerable promise for applications in assessing PELs with a wider variety of compounds and for others that have direct effects on epithelial tissues throughout the respiratory tract. Dosimetry modeling is an important part of the methodology for establishing RfCs and is highlighted in the RfC documentation (USEPA 1994).

Interpretive Applications of PBTK Models

Biological Exposure Indices (BEIs) and other biomarkers of exposure are often developed by relating the exposure of a group of workers to concentrations of chemicals or metabolites in blood, urine, breath or hair. The standard analysis of this type of data usually relies on empirical relationships to analyze findings and trends. When a valid PBTK model can be developed for the chemical and its biomarker, the relationship between exposure and the biomarker can be explained mechanistically and extrapolated, if necessary, to other exposure situations and other routes of exposure. Examples of PBTK models in support of biomarker research includes hemoglobin adducts (Fennell *et al.* 1992), metal accumulation in hair (Clewell *et al.* 1999), and urinary/exhaled breath measures following exposure (Leung 1992; Perbellini *et al.* 1990; Thomas *et al.* 1996). Another application of

PBTK models is in evaluating tissue dosimetry in connection with epidemiological studies (Smith 1991). Models addressing absorption through the skin have also been developed (McDougal *et al.* 1990; McKone 1993). Incorporation of these PBTK models into epidemiological studies to assist exposure assessment would enhance the ability to relate biomarkers to tissue doses across different routes and differing temporal patterns of exposure.

RISK ASSESSMENT APPLICATIONS IN THE OCCUPATIONAL SETTING

The usual approach in setting PELs (*i.e.*, risk assessment for occupational exposures) relies heavily on available human data relating workplace exposure levels and specific symptoms or complaints. However, even for common chemicals such as xylene and other solvents, there is considerable uncertainty in establishing PELs with human data due to inconsistency in results of the human studies (for an example see OEHHA 1997). Usually, the uncertainties in extrapolating from animal studies results in a decided bias toward use of human studies, whatever their limitations. Typically, these human studies evaluate the data from controlled volunteer studies or from workers at their job sites, using statistical analysis (*e.g.*, Ong *et al.* 1991), empirical relationships, or compartmental models. Benignus provided an interesting example of the use of PBTK modeling as an adjunct to the human studies (Benignus *et al.* 1998). Here, a PBTK model estimated tissue dose metrics for toluene exposure for a variety of diverse exposure conditions. By relying on a measure of tissue dose within the body, a common dose-response curve described results from all the studies. Johanson and colleagues have developed PBTK models for a variety of compounds such as ethyl *t*-butyl ether (Nihlen and Johanson 1999), trimethylbenzene (Jarnberg and Johanson 1999) and acetone and toluene as influenced by ingestion of clorzoxazone (Ernstgard *et al.* 1999) based on controlled human exposures. In the future, there will likely be emphasis in providing a more quantitative evaluation of animal studies as part of the process of setting PELs. Analysis of animal studies in the context of human relevance will encourage more routine development of interspecies PBTK models for compounds of occupational interest.

A concern in the workplace is differential responses of susceptible individuals. Molecular tools have become available to characterize differences in metabolizing enzymes and differences in genetic endowment within a population. Many factors, such as gender, age, ethnicity, genetic makeup, environmental or lifestyle exposure, may alter susceptibility to toxicity (Hirvonen 1999; Portier and Bell 1998; Smith 1999). Environmental/lifestyle factors include the use of tobacco, non-prescription and prescription drugs, ethanol consumption, nutrition/diet, and disease states. Polymorphisms in activating and detoxifying enzymes exist for most biotransformation enzymes (Klaassen 1995; Stephens *et al.* 1994). The effects of these factors on metabolism vary. These polymorphisms may decrease or increase enzyme activities and clearance of toxic compounds from the body. Existing risk assessments typically offer only a qualitative evaluation of at-risk groups/individuals (see, *e.g.*, USEPA 1999a). To the degree that these individual differences are related to differences in dosimetry among individuals, the risk for any given population/individual could be estimated by incorporating the physiological or metabolic characteristics of the

population into the PBTK dosimetry model. For populations that exhibit a distribution in the range of critical parameters, the distribution of expected risks can be characterized by use of cumulative probability type analyses (Bois *et al.* 1996; Clewell and Andersen 1996; Portier and Kaplan 1989). Researchers with interest in TK modeling in the occupational environment will likely continue to explore individual variability and its impact on risks for both the general population and for workers.

These methods generally rely on Monte Carlo simulation techniques. The Monte Carlo approach addresses variability by permitting the model input parameters to vary in accordance with some known or assumed statistical distribution, *i.e.*, normal, lognormal, *etc.* The model is run iteratively, with the input parameters randomly sampled for each run. Model output is presented as a distribution of values, and the distributions can be analyzed statistically. This approach has been used to estimate the percentage of a worker population that would be protected by the BEIs for six organic chemicals (Thomas *et al.* 1996). Similarly, OSHA used Monte Carlo approaches to estimate the excess cancer risk from exposure to methylene chloride at the proposed new PEL during rulemaking (OSHA 1997). These techniques provide a means of dealing with the larger inherent variability in humans as compared to laboratory test animals and to allow estimation of upper bound dose metrics as required in protecting the majority of workers. Yet, there are large areas of uncertainty in how Monte Carlo analysis should be performed, including the determination of relevant parameter distributions in different populations and subpopulations of humans, in issues pertaining to autocorrelation of model parameters, and in the interpretation of the resulting dose metric distributions.

The toxicity of many compounds in the work environment occurs due to their metabolism to reactive, toxic compounds. Carbon tetrachloride is metabolized to a free radical in the liver to cause hepatotoxicity. Chloroform is oxidized in liver and kidney to phosgene. Vinyl chloride is converted to an epoxide. Sequential oxidation processes produce the neurotoxic metabolite, 2,5- hexanedione, from hexane. Other compounds are removed by metabolism, including toluene and xylenes. The metabolism of many of these low molecular weight volatile compounds occurs via a common enzyme, cytochrome P450 2E1 (Guengerich and Shimada 1991). Some cases of susceptibility are likely to be associated with alterations in the activity of this enzyme among individuals because of lifestyle, health, or genetic differences. The activity of this enzyme is affected by co-exposures to ketones, persistent use of alcohol, drug therapies that block or enhance the activity, diabetes, and altered liver function. A useful tool for assessing susceptibility to these compounds bioactivated or cleared by CYP2E1 will be further development of mechanistic models for CYP 2E1 activity under different situations in various populations. Chien (Chien *et al.* 1997) developed a mechanistic model for the induction of cytochrome P450 2E1 enzymes by ethanol. This model, in some ways similar to the induction models for TCDD, is both a toxicokinetic and toxicodynamic model for the effects of ethanol. Another example of a pharmacodynamic response in relation to susceptibility would be CYP 2E1 inhibition due to the use of pharmaceuticals, such as disulfiram (Emery *et al.* 1999). The influence of the concomitant CYP 2E1 inhibition and induction by acetone and other chemicals, the effects of nutrition, diet, disease, genetics, and other aspects of individual susceptibility could be addressed in similar fashion.

The standard approach to developing PBTK models for risk assessment purposes involves several steps that vary due to the nature of the chemical, the exposures, and the anticipated risks. Initially, the PBTK model is developed with available information regarding the kinetics of the chemical in rodents and/or humans. This phase includes analysis, as appropriate, of time course data, enzymology, and other critical effects such as respiratory depression, glutathione depletion, or metabolite interactions. Most often, initial model development will uncover specific data needs that are filled by conducting additional experiments in animal, human or *in vitro* models. A working model is then used for performing dose-response assessments of animal toxicity or occupational epidemiology data, such as was performed with the methylene chloride cancer bioassays. Once the models are sufficiently developed, the sensitivity of the model to various input parameters is assessed using sensitivity analysis techniques. These sensitivity analyses provide information regarding the importance of various parameters with respect to the model dose metrics. These analyses also serve as priority setting tools. The more sensitive the model dose metric is to a given process (*i.e.*, parameter) the more important it is to accurately know the parameter. Sensitivity analyses have been performed recently with PBPK models used to support risk assessments for environmental and occupational exposures to vinyl acetate and acrylic acid (Plowchalk *et al.* 1997; Andersen *et al.* 2000).

In refinement steps, the model structure can be changed or key parameters may be estimated by direct experiments. Eventually, the model will be scaled up to humans by substituting human values for applicable parameters. These values again may be estimated from similar chemicals, scaled allometrically, or measured *in vitro*. *In vitro* experiments with human tissues can be important in developing estimates for various parameters, particularly biochemical parameters including tissue binding, enzyme activity, enzyme induction, and similar phenomena. If human kinetic data are available, they can then be used to validate the human version of the PBTK model. The model is then used to predict values for dose metrics that can be used in the dose-response phase of a human health risk assessment, to extrapolate to pertinent exposure regimens, or to assist in designing experiments to collect human kinetic data for further validation of the model. Tardif (Tardif *et al.* 1997; Tardif *et al.* 1993) used this approach to develop a human PBTK model for toluene, ethylbenzene, and xylene mixtures. Initially, the model was developed in rats, using data on the time course of chemical in venous blood to assist in estimating kinetic parameters at several exposure levels. The model was then scaled up to humans using allometric relationships. Data collected in controlled human studies at exposure levels below the Threshold Limit Value for the chemicals were then used to show the validity of the model after scaling. The model was then available for various purposes including extrapolation to other exposure levels and evaluation of risks posed by mixtures.

Many of these issues (Table 1) surrounding chemical risk assessment in the workplace can be addressed by developing PBTK models for tissue dose metrics that take advantage of a number of informational and experimental resources.

Table 1. Some possible uses of TK/PBTK models in occupational health/ industrial hygiene.

Incorporation of mechanistic data from in vitro studies
Identification and use of appropriate dose metrics for risk assessment
Adjustment of PELs for varying exposure scenarios
Evaluations of exposure by non-inhalation routes
Evaluation of kinetic interactions between chemicals in mixtures
Extrapolation outside the range of data for dose, dose route, and species
Estimation of population variability in response to toxicants
Assessment of differences in risk for susceptible populations
Biologically-based standard setting

CONCLUSIONS

TK models in general have matured over the last 15 years into tools that possess a clear role in pharmacokinetic research and, especially, in human health based chemical risk assessment for the general population and for workers. PBTK models are generally more useful for toxicity- related assessments than simpler TK models because they are more firmly rooted in the underlying biology of the animal. The biological underpinnings that lead to varying sensitivity of different human populations is readily incorporated in PBTK models that include specific biochemical pathways for activation and detoxification processes. Variability and uncertainty analyses, paramount in new approaches to health risk assessment, are readily addressed within the structure of PBTK models. Future occupational health chemical risk assessment will also likely depend more on PBTK modeling where these structured models become more useful in linking information on blood levels/tissue dose and toxic responses. As the use of mechanistic models, such as that for ethanol or TCDD induction of liver enzymes, coupled with toxicokinetic models becomes more commonplace, greater insight should be developed into how sensitive individuals are affected by chemical exposure and how we may go about protecting all workers.

REFERENCES

Andersen ME, Clewell HJ, Gargas ML, *et al.* 1987. Physiologically based pharmacokinetics and the risk assessment process for methylene chloride. Toxicol Appl Pharmacol 87:185-205

Andersen ME, Clewell HJ, Gargas ML, *et al.* 1991. Physiologically based pharmacokinetic modeling with dichloromethane, its metabolite, carbon monoxide, and blood carboxyhemoglobin in rats and humans. Toxicol Appl Pharmacol 108:14-27

Andersen ME, Mills JJ, Gargas ML, *et al.* 1993. Modeling receptor-mediated processes with dioxin: Implications for pharmacokinetics and risk assessment. Risk Anal 13:25-36

Andersen ME, Birnbaum LS, Barton HA, *et al.* 1997a. Regional hepatic CYP1A1 and CYP1A2 induction with 2,3,7,8- tetrachlorodibenzo-p-dioxin evaluated with a multicompartment geometric model of hepatic zonation. Toxicol Appl Pharmacol 144:145-55.

Andersen ME, Eklund CR, Mills JJ, *et al.* 1997b. A multicompartment geometric model of the liver in relation to regional induction of cytochrome P450s. Toxicol Appl Pharmacol 144:135-44

Andersen ME, Sarangapani R, Frederick CB, *et al.* 1999. Dosimetric adjustment factors for methylmethacrylate derived from a steady-state analysis of a physiologically-based clearance-extraction model. Inhal Toxicol 11:899-926

Andersen M, Sarangapani R, Gentry R, *et al.* 2000. Application of a hybrid CFD-PBPK nasal dosimetry model in an inhalation risk assessment: An example with acrylic acid. Toxicol Sci 57:312-25

Benignus VA, Boyes WK, and Bushnell PJ. 1998. A dosimetric analysis of behavioral effects of acute toluene exposure in rats and humans. Toxicol Sci 43:186-95

Bischoff KB, Dedrick RL, Zaharko DS, *et al.* 1971. Methotrexate pharmacokinetics. J Pharm Sci 60:1128-33

Bogdanffy MS, Sarangapani R, Plowchalk DR, *et al.* 1999. A biologically-based risk assessment for vinyl acetate-induced cancer and non-cancer inhalation toxicity. Toxicol Sci 51:19-35

Bois FY, Jackson ET, Pekari K, *et al.* 1996. Population toxicokinetics of benzene. Environ Health Perspect 104(suppl 6):1405-11

Carrier G, Brunet RC, and Brodeur J. 1995a. Modeling of the toxicokinetics of polychlorinated dibenzo-p-dioxins and dibenzofurans in mammalians, including humans. I. Nonlinear distribution of PCDD/PCDF body burden between liver and adipose tissues. Toxicol Appl Pharmacol 131:253-66

Carrier G, Brunet RC, and Brodeur J. 1995b. Modeling of the toxicokinetics of polychlorinated dibenzo-p-dioxins and dibenzofurans in mammalians, including humans. II. Kinetics of absorption and disposition of PCDDs/PCDFs. Toxicol Appl Pharmacol 131:267-76

Chien JY, Thummel KE, and Slattery JT. 1997. Pharmacokinetic consequences of induction of CYP2E1 by ligand stabilization. Drug Metab Dispos 25:1165-75

CIIT (Chemical Industry Institute of Toxicology). 1999. Formaldehyde: Hazard Characterization and Dose-Response Assessment for Carcinogenicity by the Route of Inhalation. Research Triangle Park, NC, USA

Clewell HJ, 3rd. 1995. The application of physiologically based pharmacokinetic modeling in human health risk assessment of hazardous substances. Toxicol Lett 79:207-17

Clewell HJ and Andersen ME 1985. Risk assessment extrapolations and physiological modeling. Toxicol Ind Health 1:111-31

Clewell HJ, 3rd, and Andersen ME. 1996. Use of physiologically based pharmacokinetic modeling to investigate individual versus population risk. Toxicology 111:315-29

Clewell HJ, Gearhart JM, Gentry PR, *et al.* 1999. Evaluation of the uncertainty in an oral reference dose for methylmercury due to interindividual variability in pharmacokinetics. Risk Anal 19:547-58

Clewell HJ, III, Gentry PR, Gearhart JM, *et al.* 2001. Development of a physiologically based pharmacokinetic model for isopropanol and acetone. Toxicol Sci (*in press*).

Coburn RF, Forster RE and Kane PB. 1965. Considerations of the physiological variables that determine the blood carboxyhemoglobin concentration in man. J Clin Invest 44:1899-910

Emery MG, Jubert C, Thummel KE, *et al.* 1999. Duration of cytochrome P-450 2E1 (CYP2E1) inhibition and estimation of functional CYP2E1 enzyme half-life after single-dose disulfiram administration in humans. J Pharmacol Exp Ther 291:213-9

Emmen HH, Hoogendijk EM, Klopping-Ketelaars WA, *et al.* 2000. Human safety and pharmacokinetics of the CFC alternative propellants HFC 134a (1,1,1,2-tetrafluoroethane) and

HFC 227 (1,1,1,2,3,3, 3- heptafluoropropane) following whole-body exposure. Regul Toxicol Pharmacol 32:22-35

Ernstgard L, Gullstrand E, Johanson G, *et al.* 1999. Toxicokinetic interactions between orally ingested chlorzoxazone and inhaled acetone or toluene in male volunteers. Toxicol Sci 48:189-96

Evans MV and Andersen ME. 2000. Sensitivity analysis of a physiological model for 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD): assessing the impact of specific model parameters on sequestration in liver and fat in the rat. Toxicol Sci 54:71-80

Fennell TR, Sumner SC, and Walker VE. 1992. A model for the formation and removal of hemoglobin adducts. Cancer Epidemiol Biomarkers Prev 1:213-9

Filser JG and Bolt HM. 1979. Pharmacokinetics of halogenated ethylenes in rats. Arch Toxicol 42:123-36

Filser JG and Bolt HM. 1981. Inhalation pharmacokinetics based on gas uptake studies. I. Improvement of kinetic models. Arch Toxicol 47:279-92

Fiserova-Bergerova V, Vlach J, and Singhal K. 1974. Simulation and prediciton of uptake, distribution, and exhalation of organic solvents. Br J Ind Med 31:45-52

Frederick CB, Bush ML, Lomax LG, *et al.* 1998. Application of a hybrid computational fluid dynamics and physiologically based inhalation model for interspecies dosimetry extrapolation of acidic vapors in the upper airways. Toxicol Appl Pharmacol 152:211-31

Gerlowski LE and Jain RK. 1983. Physiologically based pharmacokinetic modeling: Principles and applications. J Pharm Sci 72:1103-27

Gibaldi M and Perrier D. 1982. Pharmacokinetics. Marcel Dekker, NY, NY, USA

Green T. 1997. Methylene chloride induced mouse liver and lung tumours: An overview of the role of mechanistic studies in human safety assessment. Hum Exp Toxicol 16:3-13

Guengerich FP and Shimada T. 1991. Oxidation of toxic and carcinogenic chemicals by human cytochrome P-450 enzymes. Chem Res Toxicol 4:391-407

Haggard HW. 1924a. The absorption, disribution, and elimination of ethyl ether. II. Analysis of the mechanism of absorption and elimination of such a gas or vapor as ethyl ether. J Biol Chem LIX:753-70

Haggard HW. 1924b. The absorption, disribution, and elimination of ethyl ether. III. The relation of the concentration of ether, or any similar volatile substance, in the central nervous system to the concentration in the arterial blood, and the buffer action of the body. J Biol Chem LIX:771-81

Himmelstein KJ and Lutz RJ. 1979. A review of the applications of physiologically based pharmacokinetic modeling. J Pharmacokinet Biopharm 7:127-44

Hirvonen A. 1999. Polymorphisms of xenobiotic-metabolizing enzymes and susceptibility to cancer. Environ Health Perspect 107 Suppl 1:37-47

ILSI (International Life Sciences Institute). 1997. An Evaluation of EPA's Proposed Guidelines for Carcinogen Risk Assessment Using Chloroform and Dichloroacetate as Case Studies: Report of an Expert Panel. Washington, DC, USA

Jarnberg J and Johanson G. 1999. Physiologically based modeling of 1,2,4-trimethylbenzene inhalation toxicokinetics. Toxicol Appl Pharmacol 155:203-14

Johanson G and Filser JG. 1993. A physiologically based pharmacokinetic model for butadiene and its metabolite butadiene monoxide in rat and mouse and its significance for risk extrapolation. Arch Toxicol 67:151-63

Kety SS. 1951. The theory and applications of the exchange of inert gas at the lungs. Pharmacol Rev 3:1-41

Klaassen CD (ed). 1995. Casarett and Doull's Toxicology: The Basic Science of Poisons. McGraw Hill, NY, NY, USA

Kohn MC, Lucier GW, Clark GC, *et al.* 1993. A mechanistic model of effects of dioxin on gene expression in the rat liver. Toxicol Appl Pharmacol 120:138-54

Leung HW. 1991. Development and utilization of physiologically based pharmacokinetic models for toxicological applications. J Toxicol Environ Health 32:247-67

Leung HW. 1992. Use of physiologically based pharmacokinetic models to establish biological exposure indexes. Am Ind Hyg Assoc J 53:369-74

Leung HW, Paustenbach DJ, Murray FJ, *et al.* 1990. A physiological pharmacokinetic description of the tissue distribution and enzyme-inducing properties of 2,3,7,8-tetrachlorodibenzo-p- dioxin in the rat. Toxicol Appl Pharmacol 103:399-410

Mapleson WW. 1963. An electric analog for uptake and exchange of inert gases and other agents. J Appl Physiol 18:197-204

McDougal JN, Jepson GW, Clewell HJ, *et al.* 1990. Dermal absorption of organic chemical vapors in rats and humans. Fundam Appl Toxicol 14:299-308

McKenna MJ, Watanabe PG, and Gehring PJ. 1977. Pharmacokinetics of vinylidene chloride in the rat. Environ Health Perspect 21:99-105

McKone TE. 1993. Linking a PBPK model for chloroform with measured breath concentrations in showers: Implications for dermal exposure models. J Expo Anal Environ Epidemiol 3:339- 65

Nihlen A and Johanson G. 1999. Physiologically based toxicokinetic modeling of inhaled ethyl tertiary- butyl ether in humans. Toxicol Sci 51:184-94

NRC (National Research Council). 1987. Pharmacokinetics in Risk Assessment. National Academy Press, Washington, DC, USA

OEHHA (Office of Environmental Health Hazard Assessment). 1997. Public Health Goal for Xylene In Drinking Water. Pesticide and Environmental Toxicology Section, California Environmental Protection Agency, Sacramento, CA, USA

O'Flaherty EJ. 1981. Toxicants and Drugs: Kinetics and Dynamics. Wiley, NY, NY, USA

Ong CN, Sia GL, Ong HY, *et al.* 1991. Biological monitoring of occupational exposure to methyl ethyl ketone. Int Arch Occup Environ Health 63:319-24

OSHA (Occupational Safety and Health Administration). 1997. Methylene Chloride VI. Quantitative Risk Assessment. U.S. Department of Labor, Washington, DC, USA

Perbellini L, Mozzo P, Olivato D, *et al.* 1990. "Dynamic" biological exposure indexes for n-hexane and 2,5-hexanedione, suggested by a physiologically based pharmacokinetic model. Am Ind Hyg Assoc J 51:356-62

Peterson JE and Stewart RD. 1975. Predicting the carboxyhemoglobin levels resulting from carbon monoxide exposures. J Appl Physiol 39:633-8

Plowchalk DR, Andersen ME, and Bogdanffy MS. 1997. Physiologically based modeling of vinyl acetate uptake, metabolism, and intracellular pH changes in the rat nasal cavity. Toxicol Appl Pharmacol 142:386-400

Portier CJ and Bell DA. 1998. Genetic susceptibility: significance in risk assessment. Toxicol Lett 102-103:185-9

Portier CJ and Kaplan NL. 1989. Variability of safe dose estimates when using complicated models of the carcinogenic process. A case study: Methylene chloride. Fundam Appl Toxicol 13:533-44

Ramsey JC and Andersen ME. 1984. A physiologically based description of the inhalation pharmacokinetics of styrene in rats and humans. Toxicol Appl Pharmacol 73:159-75

Ramsey JC and Young JD. 1978. Pharmacokinetics of inhaled styrene in rats and humans. Scand J Work Environ Health 4:84-91

Reitz RH, Mendrala AL, and Guengerich FP. 1989. In vitro metabolism of methylene chloride in human and animal tissues: Use in physiologically based pharmacokinetic models. Toxicol Appl Pharmacol 97:230-46

Salvan A, Thomaseth K, Bortot P, *et al.* 1999. Uncertainty in estimating exposure using a toxicokinetic model. The example of 2,3,7,8-tetrachlorodibenzo-p-dioxin. Ann N Y Acad Sci 895:125-40

Smith MT. 1999. Benzene, NQO1, and genetic susceptibility to cancer. Proc Natl Acad Sci USA 96:7624-6

Smith TJ. 1991. Pharmacokinetic models in the development of exposure indicators in epidemiology. Ann Occup Hyg 35:543-60

Stephens EA, Taylor JA, Kaplan N, *et al.* 1994. Ethnic variation in the CYP2E1 gene: Polymorphism analysis of 695 African-Americans, European-Americans and Taiwanese. Pharmacogenetics 4:185-92

Tardif R, Lapare S, Krishnan K, *et al.* 1993. Physiologically based modeling of the toxicokinetic interaction between toluene and m-xylene in the rat. Toxicol Appl Pharmacol 120:266-73

Tardif R, Charest-Tardif G, Brodeur J, *et al.* 1997. Physiologically based pharmacokinetic modeling of a ternary mixture of alkyl benzenes in rats and humans. Toxicol Appl Pharmacol 144:120- 34

Teorell T. 1937a. Kinetics of distribution of substances administered to the body. I. The extravascular modes of administration. Arch Int Pharmacodyn 57:205-25

Teorell T. 1937b. Kinetics of distribution of substances administered to the body. II. The intravascular mode of administration. Arch Int Pharmacodyn 57:226-40

Thomas RS, Bigelow PL, Keefe TJ, *et al.* 1996. Variability in biological exposure indices using physiologically based pharmacokinetic modeling and Monte Carlo simulation. Am Ind Hyg Assoc J 57:23-32

USEPA (U.S. Environmental Protection Agency). 1994. Methods for Derivation of Inhalation Reference Concentrations and Application of Nasal Dosimetry. EPA-600/8-90-066F. Office of Research and Development. National Center for Environmental Assessment, Research Triangle Park, NC, USA

USEPA (U.S. Environmental Protection Agency). 1996. Proposed Guidelines for Carcinogen Risk Assessment. EPA-600/P-92-003C. Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1999a. Toxicological Review of Vinyl Chloride. NCEA-S-0619. External review draft. Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1999b. Toxicological Review of Ethylene Glycol Monobutyl Ether (EGBE). EPA-635/R-00-002. Integrated Risk Information System. National Center for Environmental Assessment, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 2000. Integrated Summary and Risk Characterization for 2,3,7,8-Tetrachlorodibenzo-p-Dioxin (TCDD) and Related Compounds. Part II: Health Assessment for 2,3,7,8-Tetrachlorodibenzo-p-Dioxin (TCDD) and Related Compounds. Chapter 8. Dose-Response Modeling for 2,3,7,8-TCDD. NCEA-I-0835. SAB Review Draft. National Center for Environmental Assessment. Office of Research and Development. Washington, DC, USA

Wagner JG. 1981. History of pharmacokinetics. Pharmacol Ther 12:537-62

Watanabe PG and Gehring PJ. 1976. Dose-dependent fate of vinyl chloride and its possible relationship to oncogenicity in rats. Environ Health Perspect 17:145-52

Welling PG. 1986. Pharmacokinetics: Processes and Mathematics. ACS Monograph 185. American Chemical Society, Washington, DC, USA

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1397-1404 (2002)

Improving Risk Assessment: Priorities for Epidemiologic Research*

Herman J. Gibb,[1,] Harvey Checkoway,[2] and Leslie Stayner[3]**
[1]National Center for Environmental Assessment (8601D), U.S. Environmental Protection Agency, Washington, DC 20460; Tel(voice):202- 564-3334, Tel(fax):202-565-0059; gibb.herman@epa.gov. [2]Department of Environmental Health, University of Washington, Box 357234, Seattle, WA 98195; Tel(voice):206-543-2052, Tel(fax):206-685-3990. [3]Risk Evaluation Branch, National Institute for Occupational Safety and Health, Cincinnati, OH 45226; Tel(voice):513-533-8365, Tel(fax):513-533-8224

ABSTRACT

The Epidemiology Work Group at the Workshop on Future Research for Improving Risk Assessment Methods, *Of Mice, Men, and Models*, held August 16 to 18, 2000, at Snowmass Village, Aspen, Colorado, concluded that in order to improve the utility of epidemiologic studies for risk assessment, methodologic research is needed in the following areas: (1) aspects of epidemiologic study designs that affect dose-response estimation; (2) alternative methods for estimating dose in human studies; and (3) refined methods for dose-response modeling for epidemiologic data. Needed research in aspects of epidemiologic study design includes recognition and control of study biases, identification of susceptible subpopulations, choice of exposure

* Corresponding author.

** Epidemiology Work Group: Michael Attfield, National Institute for Occupational Safety and Health; Paul Brandt-Rauf, Columbia University; Harvey Checkoway, University of Washington at Seattle; Herman J. Gibb, U.S. Environmental Protection Agency; Roy Fleming, National Institute for Occupational Safety and Health; Irva Hertz-Picciotto, University of North Carolina at Chapel Hill; David Kriebel, University of Massachusetts at Amherst; Dana Loomis, University of North Carolina at Chapel Hill; John Morawetz, Center for Worker Health and Safety Education; Steve Rappaport, University of North Carolina at Chapel Hill; Anne Sassaman, National Institute of Environmental Health Sciences; Rob Schnatter, ExxonMobil Biomedical Sciences, Inc.; Joel Schwartz, Harvard School of Public Health; Allan Smith, University of California at Berkeley; Tom Smith, Harvard School of Public Health; Leslie Stayner, National Institute for Occupational Safety and Health; Kyle Steenland, National Institute for Occupational Safety and Health; Jane Teta, Union Carbide, Inc.; Duncan Thomas, University of Southern California; Dan Wartenberg, Environmental and Occupaitonal Health Services Institute

1080-7039/02/$.50

metrics, and choice of epidemiologic risk parameters. Much of this research can be done with existing data. Research needed to improve determinants of dose in human studies includes additional individual-level data (*e.g.*, diet, co-morbidity), development of more extensive human data for physiologically based pharmacokinetic (PBPK) dose modeling, tissue registries to increase the availability of tissue for studies of exposure/dose and susceptibility biomarkers, and biomarker data to assess exposures in humans and animals. Research needed on dose-response modeling of human studies includes more widespread application of flexible statistical methods (*e.g.*, general additive models), development of methods to compensate for epidemiologic bias in dose-response models, improved biological models using human data, and evaluation of the benchmark dose using human data.

There was consensus among the Work Group that, whereas most prior risk assessments have focused on cancer, there is a growing need for applications to other health outcomes. Developmental and reproductive effects, injuries, respiratory disease, and cardiovascular disease were identified as especially high priorities for research. It was also a consensus view that epidemiologists, industrial hygienists, and other scientists focusing on human data need to play a stronger role throughout the risk assessment process. Finally, the group agreed that there was a need to improve risk communication, particularly on uncertainty inherent in risk assessments that use epidemiologic data.

Key Words: risk assessment, epidemiology, statistical models, dose modeling, dose-response.

BACKGROUND

In the past, risk assessment has been largely performed by toxicologists and statisticians, and research methods in this field have mainly emphasized toxicologic and statistical issues. Epidemiologists have generally been under represented in the risk assessment process because most risk assessments have been based on toxicologic rather than epidemiologic data. This situation appears to be changing with more epidemiologists expressing an interest an participating in risk assessment (Samet *et al.* 1998; Hertz-Piccioto 1995; Stayner *et al.* 1995), primarily due to an increase in the availability of high quality epidemiologic studies that are available for quantitative risk assessment, and lingering questions about appropriateness of animal models for predicting human risk (Ames and Gold 1990; Huff 1999).

EPIDEMIOLOGIC RESEARCH NEEDS

The Workshop on Future Research for Improving Risk Assessment Methods, *Of Mice, Men, and Models*, held August 16 to 18, 2000, at Snowmass Village, Aspen, Colorado, included an Epidemiology Work Group. Prior to the meeting in Snowmass, members of the group were invited to develop descriptions of research areas that they believed would enhance the utility of epidemiologic studies for risk assessment purposes. After presentation of recommendations by approximately 20 members of the group, the priority areas were organized into four broad categories: (1) aspects of epidemiologic study designs that affect the validity of dose-response estimation; (2) alternative methods for estimating dose in human studies; (3) refined methods for dose-response modeling for epidemiologic data; and, (4) the range of health

outcomes for consideration in future risk assessments. Additional comments were made by members of the group with respect the importance of greater involvement of epidemiologists throughout risk assessment and communication.

Aspects of Epidemiologic Studies Affecting Dose-Response Estimation

Study Biases

All epidemiologic studies are vulnerable to biases, including nonrepresentative selection of study subjects, misclassification of exposure or health outcome, and confounding. The extent to which biases are recognized and controlled determines study validity. Nonetheless, bias is never fully eliminated and may cause distortion of findings that ultimately are incorporated into risk assessments. For example, occupational cohort studies are often prone to a selection bias known as the "healthy worker effect," which is due to selection of relatively healthy workers for employment and inappropriate comparisons with national or regional population disease rates (McMichael 1976). For risk assessment, a particularly serious aspect of this bias is related to the fact that workers must be healthy to continue working, which is sometimes referred to as the "Healthy Worker Survivor Effect" (HWSE) (Arrighi and Hertz-Picciotto 1996; Robbins 1987). It has been shown empirically that HWSE may lead to dampening of dose-response relationships, which may even appear negative in occupational cohorts (Steenland and Stayner 1991).

Incomplete or erroneous exposure assessment is a pervasive shortcoming of epidemiologic research that can lead to biased dose-response estimates (Armstrong *et al.* 1998; Thomas *et al.* 1993). Deficiencies in exposure assessment can be avoided in future research provided adequate resources are allocated. For existing epidemiologic studies on which risk assessments might be based, systematic efforts to address the direction and magnitude of bias can be attempted with sensitivity analyses. The Work Group strongly endorsed both improving exposure assessments for planned future studies and exploiting statistical methods to examine measurement error bias in available datasets.

Susceptibility

Little is known about the quantitative effect of various human factors (*e.g.*, genetics, gender, age, diet) on responses to toxic environmental agents. The results of epidemiologic studies generally reflect "average" responses of the population, but may poorly reflect risks to susceptible subgroups. Epidemiologists commonly perform stratified analyses and related methods to detect effect modification of toxic exposures by host factors. This practice and the reporting of important subgroup-specific findings are encouraged; advances in statistical methods to quantify interactive effects will be beneficial in this regard. Identification of susceptible subgroups and quantification of subgroup risks will undoubtedly increase in importance in the foreseeable future as new information emerges from the Human Genome Project. There will be a definite need for the development of improved statistical methods such as those of Greenland and Poole (1994) to handle the extremely large amount of information that will emerge soon from molecular genetic research. There will also be a need for research on statistical techniques to incorporate this information into the development of risk assessment models.

Choice of Exposure Metrics

Cumulative Exposure. A convenient exposure metric for many epidemiologic studies has been cumulative exposure: the product of exposure concentration (c) and time (t). The convention is simple, recognizes that both c and t are important, and is a reasonable default when peak exposure episodes, exposure intermittence, and other exposure patterns cannot be estimated. In the future, epidemiologic research should take greater advantage of advances in environmental exposure assessment techniques that enable measurement of time-varying exposure levels. Research that quantifies the relative effects of cumulative, peak, and average exposures should shed light on exposure/disease mechanisms. This would logically lead to risk assessments based on the most relevant exposure patterns, not limited necessarily to cumulative exposure effects.

Exposure Distributions. Exposure estimates are frequently summarized (*e.g.*, average exposure) for a job, time period, or geographical location. This practice is prone to misclassification in situations where there is considerable heterogeneity of exposure within classification units. For example, in the workplace, exposures for a given job type can vary greatly depending on specific tasks performed, use of protective equipment, and local ventilation. Modern computational resources allow the incorporation of these factors to generate exposure distributions that are more valid personal indicators than are group averages. Research on methods for incorporation of exposure distributions in epidemiologic dose-response analyses should therefore be encouraged.

Exposure Windows. Related to the subject of exposure distributions is the timing of exposure which can be critical to the effect. For example, for many cancers with long latency intervals, exposures immediately preceding diagnosis or death may not be etiologically relevant (Thomas 1988). In other situations, recent exposures may be especially important if there are acute or late-stage effects. The timing of exposure is critical, sometimes even to a few days, as in the case of exposure to the developing fetus. Epidemiologic research on a variety of health outcomes is needed that more fully explores relevant exposure time windows in order to reduce exposure misclassification and hence uncertainty in risk assessments. Experience to date with these analyses is limited.

Choice of Epidemiologic Risk Parameters

Relative risk (*e.g.*, rate ratio, odds ratio) is the most common measure used in epidemiologic studies, particularly for studies of chronic diseases. Epidemiologic studies seldom report measures of risk difference (*i.e.*, excess risk) that might be used to estimate attributable risk and effect modification other than on a multiplicative scale. Other measures of risk such as years of life lost or lifetime excess risk may be more useful for informing risk managers. The ramifications of using various risk parameters in dose-response assessments deserves further study.

Dose-Response Modeling in Epidemiologic Studies

Application of "New" Statistical Methods

A variety of "new" flexible statistical models are available for application to dose-response assessment (Thomas 1998). For example, general additive models are

increasingly being applied to model complex dose-response data (Hastie and Tibshirani 1990). Hierarchial models (Greenland and Poole 1994) are also being increasingly used to combine information from different studies. Further application of such models should facilitate some of the more complex aspects of dose-response estimation, such as determination of threshold doses and estimation of nonlinear effects from peak exposures.

Benchmark Dose

The use of the benchmark dose in risk assessment has gained wide acceptance, but it has generally been employed with animal rather than human data. How human data are utilized in a benchmark dose presents its own set of issues, especially ambiguities in estimating points of departure and taking into account dose uncertainties.

Determinants of Dose in Human Studies

Development of Better Human Data for PBPK Models

Risk quantification by PBPK modeling generally has been limited by scarce human data on parameters, including tissue volumes, blood flows, tissue partition coefficients, and metabolism of common environmental contaminants. Human exposure experiments using a broad range of population subgroups are needed to define the parameter distributions associated with differences in age, sex, race, and genetic factors. Thus, it is recommended that a panel of relatively common toxicants be selected for testing that includes: (1) lipid- and water-soluble gases and vapors; and (2) particles in a range of sizes to determine deposition, uptake, and distribution. Simple field testing approaches to assess human exposures need to be developed, such as automated methods to monitor gas/vapor breath levels, measure ambient particle exposures, and measure exhaled breath for particles. In conjunction with these approaches, biomonitoring of toxicants and their metabolites in accessible tissues may be beneficial. Such data should be gathered in a variety of community and workplace settings to obtain representative samples for the population.

Biomarkers

Risk assessment for carcinogens, and other endpoints, is often hampered by extrapolations between animals and humans due to differences in toxicokinetic and toxicodynamic processes. Incorporating biomarkers of exposure, effect, and susceptibility into epidemiologic studies of occupationally and environmentally exposed cohorts has great potential to enhance the application of the studies in risk assessment. The application of biomarkers in epidemiologic studies is dependent on the availability of appropriate biological samples from exposed populations. Centralized repositories for archiving such samples would greatly facilitate the incorporation of biomarkers into epidemiologic studies. Sample repositories might be developed at appropriate Federal agencies or other large research institutes.

Biological Markers of Disease Susceptibility

Risk assessment for cancer has traditionally been built on stochastic models of the carcinogenesis process with parameters fitted to experimental or epidemiologic

data. It is generally accepted that cancer is a genetic disease, involving somatic mutations or other changes to DNA that can be induced by environmental exposures to carcinogens. It is also well established that some germ line mutations can produce a hereditary predisposition to cancer or an unusual sensitivity to environmental carcinogens. There is ongoing, widespread research in molecular genetics to identify polymorphisms involved in the metabolic activation of precarcinogens to their active form or their de-activation. Information derived from this research has great potential, ultimately, for characterizing especially susceptible subgroups within populations. Consequently, as new genetic marker information emerges, there will be an increased need for refined risk assessment methods that permit estimation of subgroup-specific risks.

Health Outcomes Needing Further Study

Historically, risk assessments based on epidemiologic data have focused mainly on cancer. The Work Group voiced a need for other endpoints to be studied as well. These include the following:

Reproductive and Developmental Outcomes. Reproductive and developmental effects have characteristics that make risk assessment for these endpoints more complex than for many other outcomes. For example, there is a wide range of specific endpoints, ranging from gonadal dysfunction, endocrine disturbances, and impaired reproductive performance to effects observed early in life, such as pregnancy wastage in subclinically or clinically detected conceptions, infant death, structural malformations, intrauterine growth retardation, deficits in development of structure or function, and transplacental carcinogenesis. Furthermore, the occurrence of an outcome may preclude another outcome or influence the shape of its dose-response. The role of repair and the timing of exposure are also key to understanding and quantifying risks from reproductive/developmental toxins.

Injury. Little research has been done on risk assessment for injuries, yet injuries are one of the leading causes of death and lost work time in the United States. Difficulties in studying injuries include problems of defining correct population denominators, estimation of appropriate doses, and poorly identified risk-modifying factors. Nonetheless, these should not be insurmountable problems for future epidemiologic research.

Cardiovascular and Respiratory Diseases. Cardiovascular disease is the leading cause of death in the United States, and there is evidence that exposure to various environmental agents, such as arsenic and fine particulate air pollution, may increase the risk of the disease. Respiratory diseases are currently of major public health concern. The increasing rates of asthma in the population offers a vivid example. Consequently, there is a clear need for including these diseases in future risk assessments.

OTHER CONSIDERATIONS

The Work Group discussed the need for increased representation of epidemiologists and industrial hygienists in the risk assessment community. It was noted that health risk assessment committees frequently do not include epidemiologists or exposure assessment scientists. Limited use of epidemiologic data and inadequate

training in risk assessment methods among epidemiologists and exposure assessors are the main reasons for this precedent.

It was felt that there was a need to develop guidelines for the skill combinations needed to adequately evaluate all studies when conducting risk assessment. Also needed are guidelines to be used in evaluating human studies in risk assessment projects. Increased training in risk assessment for epidemiologists is needed as well as additional support for epidemiologic research. There is a National Center for Toxicological Research; a National Center for Epidemiologic Research that emphasizes training in risk assessment might also be created.

Much of the research needed to improve risk assessments that use epidemiologic data can be done with existing data. The group decided that it was not necessary to conduct new epidemiologic studies to address the questions raised in the first area of research described above, "Aspects of epidemiologic studies affecting dose-response estimation". Re-analysis of existing data sets would be a logical starting point. Eventually, pooling of shared datasets to address low-level risks will be desirable. The Work Group endorsed this idea, but cautioned that appropriate attention will need to be paid to data confidentiality and protection of study subjects.

Finally, the communication of risk assessments to the general public in a manner that they can understand is a difficult challenge. Characterizing uncertainty is particularly difficult. Methods to characterize and explain uncertainty quantitatively and qualitatively are needed.

REFERENCES

Ames BN and Gold LS. 1990. Too many rodent carcinogens: Mitogenesis increases mutagenesis. Science 249:970-1

Armstrong BG. 1998. Effect of measurement error on epidemiological studies of environmental and occupational exposures. Occup Env Med 55:651-6

Arrighi HM and Hertz-Piccioto I. 1996. Controlling the healthy worker effect: an example of arsenic exposure and respiratory cancer. Occ Env Med 53:455-62

Greenland S and Poole C. 1994. Empirical-Bayes and semi-bayes approaches to occupational and environmental hazard surveillance. Arch of Env Health 49(1):9-16

Hastie TJ and Tibshirani RJ. 1990. Generalized additive models. Chapman and Hall, London, UK

Hertz-Picciotto I. 1995. Epidemiology and quantitative risk assessment: A bridge from science to policy. Am J Pub Health 85:484-91

Huff J. 1999. Long-term chemical carcinogenesis bioassays predict human cancer hazards. Issues, controversies, and uncertainties. Ann NY Acad Sci 895:56-79

McMichael AJ. 1976. Standardized mortality ratios and the "Healthy Worker Effect": Scratching beneath the surface. J Occ Med 28(3):165-8

Robbins J. 1987. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. J Chron Dis 40(suppl 2):139S-61S

Samet J, Schnatter R, and Gibb H. 1998. Invited commentary: Epidemiology and risk assessment. Am J Epi 148(10):929-36

Stayner L, Smith R, Bailer J, *et al.* 1995. Modeling epidemiologic studies of occupational cohorts for the quantitative assessment of carcinogenic hazards. Am J Ind Med 27:155-70

Steenland K and Stayner L. 1991. The importance of employment status in cohort mortality studies. Epidemiology 2(6):418-23

Thomas DC. 1988. Exposure-time-response relationships with applications to cancer epidemiology. Annual Review of Public Health 9:451-82

Thomas DC. 1998. New approaches to the analysis of cohort studies. Epidemiologic Reviews 14:122-34

Thomas DC, Stram D, and Dwyer J. 1993. Exposure measurement error: Influence on exposure- disease relationships and methods of correction. Annual Reviews of Public Health 14:69- 93

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1405-1419 (2002)

Improving Risk Assessment: Toxicological Research Needs[1]

Mark Toraason,[2]* Mel Andersen,[3] Matthew S. Bogdanffy,[4] David Dankovic,[5] Elaine Faustman,[6] Paul Foster,[7] Clay Frederick,[8] Lynne Haber,[9] Carole A. Kimmel,[10] Steven Lewis,[11] Roger McClellan,[12] Ronald Melnick,[13] Frank Mirer,[14] Kevin Morgan,[15] Val Schaeffer,[16] Ellen Silbergeld,[17] William Slikker,[18] James Swenberg,[19] and Harri Vainio[20]

ABSTRACT

A workshop convened to define research needs in toxicology identified several deficiencies in data and methods currently applied in risk assessment. The workshop panel noted that improving the link between chemical exposure and toxicological response requires a better understanding of the biological basis for inter- and intra-human variability and susceptibility. This understanding will not be complete unless all life stages are taken into consideration. Because animal studies serve as a foundation for toxicological assessment, proper accounting for cross-species extrapolation is essential. To achieve this, adjustments for dose-rate effects must be

* Corresponding author: National Institute for Occupational Safety and Health, C23, 4676 Columbia Parkway, Cincinnati OH 45226; Tel(voice): 513-533-8207, Tel(fax): 513-533-8138; mtoraason@cdc.gov

1 Report of the "Toxicology and Risk Assessment" Breakout Group session of the Workshop "Future Research for Improving Risk Assessment Methods: Of Mice, Men, and Models, August 16-18, 2000, Snowmass, Colorado. Invited participants of the breakout group are the authors of this paper. [2]National Institute for Occupational Safety and Health, MSc23, 4676 Columbia Parkway, Cincinnati OH 45226. [3]Ft. Collins, CO. [4]E.I. Dupont, Newark, DE. [5]National Institute for Occupational Safety and Health, Cincinnati, OH. [6]University of Washington, Seattle, WA. [7]Chemical Industry Institute of Toxicology, Research Triangle Park, NC. [8]Rohm and Haas Company, Spring House, PA. [9]Toxicology Excellence for Risk Assessment, Cincinnati, OH. [10]U.S. Environmental Protection Agency, Washington, DC. [11]Exxon Mobil, Annandale, NJ. [12]Albuquerque, NM. [13]National Institute of Environmental Health Sciences, Research Triangle Park, NC. [14]International Union, U.A.W. Detroit, MI. [15]Glaxo Wellcome, Inc. Research Triangle Park, NC. [16]U.S, Occupational Safety and Health Administration, Washington, DC. [17]University of Maryland, Baltimore, MD. [18]U.S. Food and Drug Administration, Jefferson, AR. [19]University of North Carolina, Chapel Hill, NC. [20]International Agency for Research on Cancer, Lyon, France.

1080-7039/02/$.50

improved, which will aid in extrapolating toxicological responses to low doses and from short-term exposures. Success depends on greater use of validated biologically based dose-response models that include pharmacokinetic and pharmacodynamic data. Research in these areas will help define uncertainty factors and reduce reliance on underlying default assumptions. Throughout the workshop the panel recognized that biomedical science and toxicology in particular is on the verge of a revolution because of advances in genomics and proteomics. Data from these high-output technologies are anticipated to greatly improve risk assessment by enabling scientists to better define and model the elements of the relationship between exposure to biological hazards and health risks in populations with differing susceptibilities.

Key Words: susceptibility, life stages, pharmacokinetics, dose response, exposure assessment, genomics.

INTRODUCTION

Considerable progress has been made over the last 20 years in defining discrete components that affect relationships between exposure circumstances and biological effects. However, toxicology must move toward better characterization and understanding of the key cellular and molecular alterations that are responsible for adverse effects observed in experimental animals and humans. To achieve this goal, research is needed to address several deficiencies in data and methods currently applied in risk assessment. To improve the link between chemical exposure and toxicological response, the following issues must be considered: (1) increased understanding of inter- and intra- individual variability in susceptibility with special attention to susceptibility during all life stages, (2) accounting for factors that affect cross species extrapolation, (3) adjusting for dose rate effects, (4) defining toxicological responses at low doses, (5) making use of continuous as well as quantal data from toxicological responses, (6) developing better response data from short-term exposures, (7) addressing exposures to chemical mixtures and by multiple exposure routes, and (8) refining uncertainty factors with reliable experimental data. Improvement in most of these areas requires identification and quantification of molecular and cellular biomarkers along the critical pathway between exposure to an agent and clinical or functional expression of toxicity. Improvement also depends on developing biologically based dose response (BBDR) models that can link exposure and biological response in a physiologically realistic framework. The framework must account for the time- and dose-dependent delivery of the toxic form of the agent to its biological target (kinetics) and the time- and dose-dependent changes in the biological system that lead to an adverse response (dynamics).

INTER- AND INTRA-INDIVIDUAL VARIABILITY IN SUSCEPTIBILITY

Traditionally, chemical risk assessors and their methods have focused on defining qualitative and increasingly quantitative relationships between exposure to toxicants and adverse effects. A risk assessor's ability to establish this connection has been hampered by inter-individual variability (Hattis 1996). Although much uncertainty in risk assessment is attributable to absent or inconsistent data, this source of

uncertainty theoretically can be addressed by filling identified data gaps. In contrast, inter-individual variability can only be addressed by understanding the underlying basis for the variability and then applying new methods that enable the incorporation of this information into risk assessments. A major source of uncertainty in risk assessment is how responses to chemicals and physical agents vary not only among individuals but within an individual under changing circumstances. Gender, race, ethnicity, lifestyle, genetic predisposition, and age (conception to senescence), are factors that must be considered in risk assessment, as they can contribute to variation among individuals in disease outcome resulting from environmental insult (Perera 2000). Age, life-style changes, reproductive status, drug use, and previous exposures, among other factors, can also contribute to response variation within an individual over time.

While experimental toxicologists generally design studies to control for inter-individual variability, epidemiologists routinely include corrections for confounders or effect modifiers such as smoking, alcohol use, diet, gender, race, and age. When appropriate, this information is included in risk assessments with the ultimate goal of reducing uncertainty. Although gender, age, and diet have been addressed in experimental animal studies, race, ethnicity, and lifestyle factors are not easily addressed. As a consequence, risk assessments must rely solely on human data to assess the contribution of these factors.

Increasingly, epidemiological studies have attempted to identify genetic polymorphisms that may explain in part inter-individual variations. Genetic polymorphisms that appear to predispose individuals to cancer have received the most attention. The most extensively studied susceptibility factors are the genetic variations in Phase I and Phase II xenobiotic-biotransformtion enzymes (Perera 2000). Investigation of genetic variants in cytochrome P450 (Ishibe *et al.* 1997; Mollerup *et al.* 1999), glutathione *S*-transferase, and *N*-acetyltransferase (Trizna *et al.* 1998) have contributed greatly to an understanding of the source of variation among individuals in terms of their response to chemical carcinogens. Altered expression of these enzymes because of inherited polymorphisms or differences in the levels of enzyme induction can lead to different abilities to activate or detoxify xenobiotics and thereby alter a person's risk to disease.

Biotransformation and metabolism are not the only concerns. Also important are inherited variations that predispose a person to cancer. Receiving considerable attention in the lay press are the breast and ovarian cancer susceptibility genes BRCA1 and BRCA2 (Brody and Biesecker 1998). These genes are believed to be important for tumor suppression. Furthermore, BRCA1 has been linked to a DNA repair protein (Gowen *et al.* 1998). Variation in DNA repair can affect a person's risk from agents that directly or indirectly damage DNA. Increased understanding of these sources of variation is needed as they have the potential to identify persons at increased risk from exposure to a toxicant.

Completion of the sequencing of the human genome and new technologies to evaluate gene expression, have given rise to greater opportunity to use this molecular information in risk assessment. However at present, the understanding is inadequate to determine when and how incorporation of data on genetic polymorphisms may affect a risk assessment and by implication, influence risk-based policy. To determine this, it will be necessary to investigate the nature and magnitude of

the impact of incorporating this information into specific risk assessments. Specifically, research must investigate the extent to which a single or combination of genetic polymorphisms affects the toxicity of environmental and occupational exposures. Physiological and molecular methods can be used to establish the phenotypic significance of genetic polymorphisms in order to estimate the impact of genotype on response to toxicant exposure. This may provide a means of calculating the significance, for risk assessment, of genetic variances that encode measurable phenotypic differences.

SUSCEPTIBILITY DURING LIFE STAGES

Sufficient, valid scientific data now exist to assert that both pre- and postnatal exposures to a variety of toxic substances can deleteriously affect the health and development of neonates and young children. Indeed, there is basis for concern that such prenatal exposures can have life-lasting effects and can manifest impacts on later life stage function and behaviors. Such toxic substances include lead, methylmercury, PCBs, ethanol, and carbon monoxide, among others. Depending on the dose received by a fetus and the specific toxicant, health consequences can range from subtle toxicological changes in animal models, including neurobehavioral effects, to death following high exposures. One area of concern regarding adverse effects of exposure on reproductive health is occupational exposures to solvents (Taskinen *et al.* 1999; Plenge-Bonig and Karmus 1999). Several recent studies have reported on relationships between parental occupational exposures and risks of childhood cancer (Colt and Blair 1998).Critical for an evaluation of potential toxicological impacts throughout development is both knowledge of key developmental pathways and their potential susceptibilities to toxicants. This necessitates the availability of test methods to evaluate such potential impacts across development as well as evaluations with a sufficiently large database of test chemicals to validate such systems.

Testing approaches for early (prenatal and early postnatal) developmental toxicity have been available for some time, although gaps still exist in evaluation of certain endpoints after early exposures, *e.g.*, immunotoxicity, respiratory, cardiovascular, renal, and liver function, and cancer. Recent efforts to expand the exposure period for prenatal assessments reflect the knowledge that organ systems continue to develop beyond organogenesis. Further testing at later stages in development arises from concern that effects on development may be manifested much later in adulthood (Selevan *et al.* 2000). Effects of exposures in the periadolescent period have not been studied sufficiently. Yet many teenagers are in the work force and may be exposed to toxic chemicals, particularly in agricultural settings (Golub 2000).

At the other end of the spectrum, exposures in older age groups are not well-evaluated in current toxicity testing approaches. Only the 2-year chronic/carcinogenicity testing protocol includes exposures into later ages, and the effects of agent exposures may be masked or exacerbated by *ad libitum* diet, resulting in obesity, and its consequences. Diet restriction in rodents has clearly been shown to increase life span and reduce disease. Thus, testing of rodents at later ages in diet-restricted and unrestricted situations is needed to identify factors in the aging population that are important in the toxicity of various exposures. With the current trend toward an

aging workforce; this information may be useful in setting more appropriate exposure limits to protect the health of this segment of the population.

Recent studies have shown that isoenzymes of xenobiotic metabolizing systems are not expressed uniformly in developing humans (de Wildt *et al.* 1999). For example, human CYP3A7 is found prominently in human fetal liver but not in adult liver (Katida *et al.* 1985; Wrighton *et al.* 1988) whereas CYPs 1A2, 2B6, and 2C8 are expressed highly in children over 1 year old (Tateishi *et al.* 1997). Similar observations of variable expression of glutathione *S*-transferase have been reported (Tee *et al.* 1992). These results suggest that activation and detoxification capacity could be age- dependent. Thus, susceptibility of neonates to chemical exposure is likely to depend on the prevalence of activation/detoxification enzymes at the time of exposure.

Due to a lack of data on the exposure effects of human pregnant mothers to environmental or occupational agents and even less on their developing children, better use must be made of available animal data. Current regulatory study designs for examining potential effects on the developing fetus or neonate induced by environmental or occupational agents do not require any information on internal dose to the mother or fetal/neonatal dosimetry. Dosimetry information in the exposed mother, fetus, and neonate would certainly improve dose-response analysis. Most of the common laboratory species on which studies of developmental toxicity are conducted during these critical windows have a markedly different rate of development and timing of developmental stages compared with humans. Essentially, most rodents are born "premature" and many critical periods of development (e.g., brain and sexual differentiation) take place postnatally in rodents that occur *in utero* in humans. Thus, the effects of lactation on *in utero* exposure can be extremely different between humans and the animal species used for testing. Animal studies must be designed to capture the comparable critical windows of human development. To meet the need in risk assessment for linking temporal exposure information with temporally sensitive developmental processes, united pharmacokinetic and dynamic models must be developed. Such models are needed to provide more accurate evaluation of dose-response relationships for the dynamic processes occurring during critical periods in development (Faustman *et al.* 2000).

IMPACT OF LIFESTYLE FACTORS ON SUSCEPTIBILITY

The toxicity of many chemicals is regulated in part by enzymatic metabolism, whose activity is determined by a variety of intrinsic and extrinsic factors. The risk of toxicity depends on changes in these enzyme activities over time which may be influenced by both genetic and lifestyle factors. Lifestyle factors include such things as pharmaceutical use, alcohol intake patterns, and health status. Several drugs are known to be enzyme inducers, dietary factors and alcohol alter the activity of an important oxidative enzyme, CYP 2E1, which is involved in the metabolism of many environmental and occupational chemicals (Chien *et al.* 1997). The complex, time-dependent interactions of multiple lifestyle factors as determinants of chemical toxicity are often overlooked in normal standard toxicity testing venues. The pattern of chemical interaction is further complicated because many of the inducing compounds may also serve as inhibitors of degradative metabolism of workplace chemi-

cals. Therefore, research is needed to determine the conditions under which these interactions are likely to enhance toxicity, which members of the population are most at risk for these interactions, and which activities and lifestyle factors lead to higher risks from chemical exposures.

Research is needed that combines and integrates experimental studies in animals and studies with human tissues and/or human volunteers to create mechanistic models of the influence of lifestyle factors, enzyme induction, and temporally disparate exposures on metabolism and expected toxicity of environmental and occupational compounds in diverse populations. It will be essential to convey the method by which qualitative and quantitative inferences drawn from mechanistic animal studies can be extended to human populations and the method by which ancillary data from human tissues and/or human volunteer studies would support inferences from animal research. It will also be necessary to demonstrate the method by which these integrated studies provide improved quantitative characterization of the variability expected in a diverse human population in response to chemicals and to alterations in enzyme activities by lifestyle factors.

PHYSIOLOGICALLY BASED PHARMACOKINETIC (PBPK) MODELING

PBPK modeling is a powerful tool for extrapolating dosimetry across species, from high doses to low doses, and across various exposure durations. Recently, PBPK models have been developed that simulate the induction of various proteins over time (Santostefano *et al.* 1998; Andersen *et al.* 1993). Extending this technology to the development of the xenobiotic metabolizing system in the neonatal liver seems reasonable. PBPK models of lactational transfer have also recently been developed for several volatile organic chemicals, and these models have been used to predict exposure of infants as a result of occupational exposure of mothers to toxic chemicals (Fisher *et al.* 1997; Byczkowski *et al.* 1994). Data must be developed to model fetal/neonatal exposure to chemicals, through lactational transfer, where neonates may be more or less susceptible to chemicals that may or may not be bioactivated or detoxified by the maternal system. Experimental dosimetry data in adult and neonatal animals must be collected for use in the development of PBPK models to describe chemical kinetics in the adult and neonate for extrapolation of the results to the humans. Model systems should focus on key xenobiotic biotransformation enzymes that are expressed differently in neonates and adults.

BIOLOGICALLY BASED DOSE-RESPONSE (BBDR)MODELS

Biological models that address mechanistic steps linking exposure to adverse effects offer an objective, data-based approach to test biologically based hypotheses and to generate alternative hypotheses for laboratory testing (Leroux *et al.* 1996; Shuey *et al.* 1994). If mechanistic hypotheses are not adequately tested in an appropriate dose-response framework, then approaches to estimate occupational risks that rely on such hypotheses may simply be substituting one set of assumptions for another, and the latter set may not provide adequate health protection. Properly validated models (*i.e.*, those most consistent with the experimental data) are needed to accurately predict measured biomarkers of exposure and biomarkers of effect.

The sequence of events between exposure and response must be linked so that BBDR models can provide mechanistic insights on the origin of biological changes that occur at the cellular and molecular levels. These models can help identify biomarkers that are appropriate measures of exposure, effect, and susceptibility. Validated BBDR models can provide a sound scientific basis for extrapolating dose-response relationships across species and outside the range of experimental observation and thus reduce uncertainties in estimating human risk.

CROSS-SPECIES EXTRAPOLATION

Species-specific information at the cellular and molecular levels is critical for developing models that can be used to quantify relationships between time-dependent target tissue dose and tissue response as a function of exposure to hazardous agents. BBDR models combine toxicokinetic data on the absorption, distribution, metabolism, and elimination of agents at different levels of exposure with mechanistic data of time-dependent tissue response (*e.g.*, mutagenicity, altered gene expression). Species-specific mechanistic data, including parameters that are measurable in humans, are critical for developing these models. Experimental data are needed to estimate relevant parameter values (*e.g.*, tissue partition coefficients, enzymatic activities, binding constants) and to resolve uncertainties in the accuracy of parameter estimates, interdependence of parameters, validity of scaling methods, variability of parameters among individuals, and effects of co-exposure to other agents that may alter any of the critical biological processes. These models should evaluate similarities and differences in animal and human response as a function of the time-dependent tissue dose, whether the correct dose metric(s) have been specified for extrapolations, and whether responses in animals reflect the range of responses that might occur in exposed workers.

A major limitation to predicting the susceptibility or resistance of human neonates to chemical exposure is a lack of similar information about the development of xenobiotic metabolizing enzymes in rodent models of human toxicity. Differences in the extent of expression of xenobiotic metabolizing enzymes between rodents and humans at a given period of development complicates the interpretation of neonatal chemical exposure studies where rodents are used as models for humans. Thus, a systematic characterization of the ontogenetic development of key xenobiotic biotransformation enzymes and repair enzymes is needed in common laboratory animal models of developmental and reproductive toxicity (*e.g.*, the rat and rabbit) compared with that of humans. Such an analysis needs to be carried out using protein expression methods (including functional analyses) rather than just through mRNA expression. The objective of this work should be a map showing the degree of expression of key isoenzymes of xenobiotic metabolism over time in laboratory animal models and in humans.

DOSE RATE EFFECTS

In many situations, human risk assessment relies on toxicity data from studies conducted in laboratory animals under standard testing protocols. Compounds are administered at constant levels over regular intervals (*e.g.*, daily 6-hour inhalation

exposure) for defined periods of time (*e.g.*, 13 weeks; 2 years). On the other hand, human exposures rarely conform to these prescriptive dosing regimes. As a matter of practical consequence, a number of default assumptions with respect to dose rate and exposure duration have become implemented in risk assessment. Doses averaged over a work shift (in most occupational scenarios) or even a lifetime (in cancer risk assessments) are generally assumed to result in equivalent risk regardless of exposure pattern. Adverse response is often assumed to be linearly related to the product of exposure level times duration (Haber's Law). For example, 1 hour exposure to 80 ppm is equivalent to 8 hours exposure to 10 ppm (Andersen *et al.* 1987). For some endpoints (*e.g.*, irritation; some developmental effects), it is commonly assumed that exposure level dominates and duration has almost no influence on risk. Most of these default assumptions have not been rigorously supported by scientific research (Jarabek 1995). Recent studies by Weller et al. (1999) indicate that the developmental effects of ethylene oxide exposure depend on both exposure level and duration, but do not conform exactly to Haber's law. Similarly, some of the carcinogenic effects of 1,3-butadiene are more dependent on exposure level than exposure duration (Melnick *et al.* 1990)

Recent improvement in our understanding of the underlying determinants of toxicity and the ability to make quantitative predictions of tissue dosimetry should facilitate more focused research in the area of dose-rate effects. PBPK models are now able to relate the time course of a wide variety of internal dose metrics from a vast number of external exposure level and duration combinations. Time- and concentration-dependent processes such as metabolism must be accounted for, so that these models will be useful in selecting experimental conditions and interpreting results. The critical biochemical determinants of dose-rate effects (*e.g.*, interaction with molecular components, repair of cellular damage) must be incorporated into the framework of risk assessment methods. Finally, early cellular biomarkers of tissue dose and toxicity are needed for investigations of the temporal relationships at lower and more relevant exposure levels. These advances are needed to permit the development of more scientifically sound approaches in accounting for exposure pattern and duration in estimations of human risk.

TOXICOLOGICAL RESPONSE IN THE LOW-DOSE REGION

Quantitative risk assessments typically involve establishing a dose-response relationship; however, it is common that the exposures of interest for environmental risk assessment purposes are below the region where a response may be observed in experimental studies. For occupational chemicals, experimental studies at times include exposures in the range that has been encountered by workers. Toxicological investigations have traditionally required the observation of overt, quantifiable response in a relatively small samples of animals. This necessity is commonly addressed by using high doses in toxicological studies, compared to the typical region of interest for humans. The term "low dose region" refers to the range of exposures encountered by humans. For industrial chemicals, workers may be exposed to levels that are several orders of magnitude higher than those found in the general environment. High doses that cause generalized toxicity may lead to altered patterns of metabolism and elimination,

compared with those that prevail at lower doses. Mechanistic studies that are conducted only at these high doses may be misleading relative to mechanisms operating at lower doses. In these cases, extrapolation of the dose-response relationship to the low-dose region below the range of experimental data may be affected by the mechanism underlying the toxicity. Well conducted toxicity and mechanistic studies include multiple exposure concentrations that extend into the region where toxicity does not alter metabolism or elimination. The risk estimates derived from high-dose extrapolations depend critically on the estimated shape of the dose-response curve in the low-dose region. Greater understanding is needed of the shapes of typical dose-response relationships for cancer and non-cancer endpoints. This will lead to reduced uncertainty in quantitative risk assessment, and an increased confidence in the resulting risk estimates. Biomarker studies focusing on the mechanistic events that ultimately lead to an overt toxicological response hold the promise of extending the range of observable response into the low-dose range, which is more relevant to human exposures. Mechanistic biomarker studies are needed to better distinguish between linear and nonlinear responses. The ultimate goal of these studies should be to provide appropriate data for low- dose extrapolations, for both cancer and non-cancer endpoints.

USE OF CONTINUOUS DATA FROM TOXICOLOGICAL RESPONSES

Although considerable research has been reported concerning the use of quantal data in dose- response modeling, far less progress has been reported on the use of continuous data (Gaylor *et al.* 1998). Continuous data are often generated in the case of noncancer endpoint studies including those of reproductive toxicity, immunotoxicity, and neurotoxicity. Useful endpoints including body weight, enzyme activities, protein and neurotransmitter concentrations, cell counts, and neuronal cell death are usually reported as continuous data. Although continuous data can be converted to quantal values in some instances, substantial precision may be lost during this process (Gaylor 1996). Therefore, procedures must be established for using continuous data in dose-response assessment.

The most controversial aspect of the using continuous data for dose-response assessments is determining the "cut-off" value for defining an adverse effect. Defining this adverse level of change from controls is a critical decision and should be grounded on sound biological and toxicological principles. An ideal method should be based on the available data, apply to most continuous data sets, and minimize arbitrary decisions. Several approaches currently available include "amount of change" considered to be adverse by experts, use of an historically-based cut off for a particular continuous data endpoint, or amount of change in the experimental mean value based on the mean and standard deviation of the control data set. Data can then be modeled as continuous data, or be converted to quantal values. Although these and other approaches have been investigated to a limited extent (Gaylor and Slikker 1990; Crump 1995; Glowa and MacPhail 1995; Kavlock *et al.* 1995; Kodell *et al.* 1995; Slikker *et al.* 1996, 1998; Bosch *et al.* 1996; Chen *et al.* 1996; Gaylor *et al.* 1998, Haber *et al.* 1998), a systematic comparison of these methods is needed to develop a valid approach for using continuous data in risk assessment.

DEVELOPMENT OF TOXICOLOGICAL RESPONSE DATA FROM ACUTE AND SHORT-TERM EXPOSURES

Current testing approaches for acute and short-term toxicity tend to be limited or nonexistent. Often, the only acute data available are from studies that are designed to determine an LD_{50} or some form of severe toxicity and done for the purpose of dose-setting for longer-term (*e.g.*, 2-week or 90 day) studies. Some data from other studies are available and are used to derive an acute reference value; *e.g.*, clinical observations in the first few days of the subchronic study may be helpful in setting standards. Also, developmental toxicity data are often used for setting acute and short-term reference values even though the exposure periods may be as long as 10 days to several weeks. This is because it is presumed that most, if not all, developmental effects are possible to induce with single exposures. However, no acute or short-term data currently are developed on the aging population. Obviously, having pharmacokinetic information and understanding the mechanism of action of the effects induced would provide more information about whether they are appropriate for acute or short-term standard setting. Thus, testing protocols are needed that can be used for setting no-observable-adverse-effect levels (NOAELs) or benchmark doses (BMDs) for acute and short-term exposures, and for determining how to use data from other studies (*e.g.*, developmental toxicity, data in the aging population, other organ studies, and longer-term studies) in addition to appropriate adult toxicity studies. In addition, useful mechanistic and pharmacokinetic data are needed to aid in understanding the best approach for testing as well as using these data in risk assessment.

EXPOSURE TO COMPLEX MIXTURES AND MULTIPLE EXPOSURE ROUTES

Most toxicological testing conducted in experimental animals relies on administration of a single compound by a single route. On the other hand, humans are often exposed to mixtures of chemicals through multiple routes. In many occupational and environmental situations, it is increasingly recognized that risk of illness or injury may be the result of combined inhalation and dermal exposure to the same chemical source. To complicate matters, the chemical source may actually be a complex mixture of several substances, all of which may contribute to the risk. Exposure to chemical mixtures may also cause chemical interactions that could either potentiate or inhibit the expression of adverse response. Recent laws, such as the Food Quality Protection Act of 1996, contain provisions that require risk assessment to address aggregate exposures from multiple routes and cumulative risk from exposure to multiple chemicals with a common mode of action. New toxicological test protocols and approaches are needed to generate data on exposures to complex mixtures and multiple routes of exposure and to integrate that information into risk assessments. If appropriately constructed and validated with experimental data, BBDR models can be used to estimate the amount of internal dose (*e.g.*, blood or tissue level) from multiple routes (*e.g.*, inhalation and dermal contact) or predict interactions at molecular targets (*e.g.*, receptor binding) from exposure to two or more compounds.

UNCERTAINTY FACTORS

Default uncertainty factors of tenfold have been used traditionally for extrapolation from animals to humans, and to account for the variability among humans including sensitive sub- populations. Human health risk assessment can be improved by improving the choice of uncertainty factors and moving away from defaults, when supported by scientific evidence. A first step away from defaults is the use of categorical defaults based on characteristics of the substance or species differences. Uncertainty factors based on categorical defaults are used in the case of animal to human extrapolation for reference concentrations when dosimetric adjustments are done (USEPA 1998; Jarabek 1995), or in the use of surface area and metabolism adjustment for oral dosing. Renwick and colleagues (Renwick 1991; Renwick and Lazarus 1998) expanded the categorical defaults into a data- derived approach, in which the interspecies and intraspecies uncertainty factors are broken into toxicokinetic and toxicodynamic components, based on relative contributions of these components for a number of chemicals. The data-derived adjustment factor approach is being enhanced to make further use of the data, and allow the incorporation of chemical-specific data without requiring the detailed level of toxicokinetic information required to build a PBPK model. In addition, data-derived factors can address intra-species and inter-species toxicodynamic variability differences, but these data are more difficult to develop. Together, this hierarchy of approaches using increasing amounts of chemical-specific information allows the replacement of defaults with chemical- and species-specific information to improve the accuracy of the assessment. Research into a number of issues is necessary before factors based on chemical-specific data can be broadly used. For example, criteria need to be developed on how to evaluate whether the critical determining factor has been identified. Similarly, it will be necessary to clarify how information about human variability is used in kinetic models. Evaluation of PBPK models for chemicals acting via selected modes of action are needed to elucidate whether the distribution of certain parameters adequately describes human variability.

TOXICOGENOMICS

Genomics and proteomics have been widely hailed as fundamental technological breakthroughs in the evaluation of both biological response and biological susceptibility (Lovett 2000; Waring and Urlich 2000). The toxicological application of this technology is referred to as 'toxicogenomics.' Strengths of this methodology are the speed of screening a large number of genes and their responses to exposures, and the potential linking of a response to its underlying mechanism. This technology has potential usefulness for risk assessment, but may only represent the first step in a process that currently must include epidemiology and animal toxicology evaluations. Linking toxicogenomics to disease outcomes is needed before it can be routinely used as a risk assessment tool.

Of particular importance is the increasing availability of information about expression patterns of tens of thousands of genes, the link between them and protein production, and the translation to eventual adverse outcome. Availability of such information requires redefining biological responses to toxicants. Such techniques offer the potential to follow biological responses with time as occupational

diseases progress. They may offer new hope in identifying early biomarkers of toxicity as well as better assessments for cross-species extrapolation of data on biomarkers of effect and susceptibility. To effectively use such information, biomarker research needs to be expanded to enable such data to be put into effective dose-response and temporal contexts. In addition, guidelines are needed for collecting and interpreting toxicogenomic information for human health risk assessments. In particular, guidelines are needed to establish criteria for acceptable levels of sensitivity, specificity, accuracy, and predictiveness for gene expression as biomarkers of disease. Research is also needed to ensure that information obtained by these technologies is highly quantitative, include evaluations of time-dependent changes consequent to specific exposures, and adequately account for the effects of mixed exposures.

With advances in genomics and proteomics, identifying the complex gene environment interaction has become increasingly possible. Genetic testing, including all the elements of gene expression to protein production, promises a possible future presymptomatic determination. Current uncertainties regarding interpretation of the results from testing raise new risk management problems. Several complex ethical, legal, and social issues (though not discussed here) will arise with the advent of this new information. Therefore, research is needed regarding the most effective use of this genetic information and appropriate management strategies must be established (Fasouliotis and Schenker 2000).

CONCLUSIONS

In the coming decade, the application of experimental data to chemical hazard identification and characterization will require risk assessors to simultaneously address toxicological issues on three fronts. First, long standing and in many cases unresolved issues need to be addressed to improve traditional toxicological testing (*e.g.*, addressing exposures to complex mixtures and accounting for multiple routes of exposure) for expanded use in risk assessment and setting of regulatory standards. Second, new types of data including biomarkers of effect and susceptibility with corresponding data in both animals and humans are needed for improved species extrapolations and dose-response assessments. Lastly, toxicology will need to develop methods to properly use data from new developments in genomics and proteomics. The enormous quantities of data expected from these high through put technologies may require a revolution in the way data can be used in risk assessment for protecting public health. Priority issues that need to be addressed on these three fronts include inter- and intra-human variability and susceptibility with special emphasis on toxicological risks through all life stages (conception though senescence). To accomplish this, improved extrapolation is needed of experimental data to environmental and occupational human exposure situations. An essential component of this will be the linking of exposures to toxicological response, including exposure-rate and dose-response relationships. The development of biologically based dose-response models offers a mechanism-based approach to summarize all available data, identify data gaps, extrapolate dose- response relationships across species and outside the range of experimental observation, and account for factors influencing inter-individual differences in susceptibility.

ACKNOWLEDGMENTS

Special thanks for participation in workshop discussions go to the following: Mariann Anticoli, James Blanchard, David Clarke, Paul Dugard, Wing Fung, Hector Garcia, Donald Gardner, Sandra Hacon, Jean Hampton, Rudolph Jaeger, Barry Johnson, Eileen Kuempel, James McDougal, Anita Meyer, Michael Pelekis, Jessica Sandler, Russell Savage, Robert Tardiff, Elizabeth Ward, Ainsely Weston, and Edna Yokoo.

REFERENCES

Andersen ME, MacNaughton MG, Clewell HJ, *et al.* 1987. Adjusting exposure limits for long and short exposure periods using a physiological pharmacokinetic model. Am Ind Hyg Assoc J 48: 335-43

Andersen ME, Mills JJ, Gargas ML, *et al.* 1993. Modeling receptor-mediated processes with dioxin: implications for pharmacokinetics and risk assessment. Risk Anal 13:25-36

Bosch RJ, Wypij D, and Ryan LM. 1996. A semiparametric approach to risk assessment for quantitative outcomes. Risk Anal 16:657-65

Brody LC and Beisecker BB. 1998. Breast cancer susceptablity genes. BRCA1 and BRCA2. Medicine 77:208-26

Byczkowski JZ, Gearhart JM, and Fisher JW. 1994. "Occupational" exposure of infants to toxic chemicals via breast milk. Nutrition 10:43-8

Chen JJ, Kodell RL, and Gaylor DW. 1996. Risk assessment for nonquantal toxic effects. In: Fan AM and Chang LW (eds), Toxicology and Risk Assessment, pp 503-513. Marcel Dekker, Inc., New York

Chien JY, Thummel KE, and Slattery JT. 1997. Pharmacokinetic consequences of induction of CYP2E1 by ligand stabilization. Drug Metab Dispos 25:1165-75

Colt JS and Balir A. 1998. Parental occupational exposures and risks of childhood cancer. Environ Health Perspec 106 (Suppl. 3):909-25

Crump KS. 1995. Calculation of benchmark doses from continuous data. Risk Anal 15:79-89

de Wildt SN, Kearns GL, Leeder JS, *et al.* 1999. Glucuronidation in humans. Pharmacogenetic and developmental aspects. Clin Pharmacokinet 36:439-52

Fasouliotis SJ and Schenker JG. 2000. BRCA1 and BRCA2 gene mutations: Decision-making dilemmas concerning testing and management. Obstet Gynecol Surv 55:373-84

Faustman EM, Silbernagel SM, Fenske RA, *et al.* 2000. Mechanisms underlying children's susceptibility to environmental toxicants. Environ Health Perspec 108 (Suppl 1):13-21

Fisher J, Mahle D, Bankston L, *et al.* 1997. Lactational transfer of volatile chemicals in breast milk. Am Ind Hyg Assoc J 58:425-31

Gaylor DW. 1996. Quantalization of continuous data for benchmark dose estimation. Toxicol Pharmacol 24:246-50

Gaylor DW, Ryan L, Krewski D, *et al.* 1998. Procedures for calculating benchmark doses for health risk assessment. Toxicol Pharmacol 28:150-64

Gaylor DW and Slikker Jr W. 1990. Risk assessment for neurotoxic effects. Neurol Toxicol 11:211-8

Glowa JR and MacPhail RC. 1995. Quantitative approaches to risk assessment in neurotoxicology. In: Chang LW and Slikker Jr W (eds), Neurotoxicology: Approaches and Methods, pp 777- 787. Academic Press, San Diego

Golub MS. 2000. Adolescent health and the environment. Environ Health Perspect 108, (Suppl 3): 355-62

Gowen LC, Avrutskaya AV, Latour AM, *et al.* 1998. BRCA1 required for transcription-coupled repair of oxidative DNA damage. Science 281:1009-12

Haber LT, Allen BC, and Kimmel CA. 1998. Non-cancer risk assessment for nickel compounds: issues associated with dose-response modeling of inhalation and oral exposures. Toxicol Sci 43:213-29

Hattis D. 1996. Human interindividual variability in susceptibility to toxic effects: From annoying detail to central determinant of risk. Toxicology 111:5-14

Ishibe N, Wiencke JK, Zuo ZF, *et al.* 1997. Susceptibility to lung cancer in light smokers associated with CYP1A1 polymorphisms in Mexican- and African-Americans. Cancer Epidemiol Biomarkers Prev 6:1075-80

Jarabek A. 1995. Consideration of temporal toxicity challenges current default assumptions. Inhalation Toxicology 7:927-46

Katida M, Kamataki K, Itahashi K, *et al.* 1985. Purification and properties of cytochrome P-450 from homogenates of human fetal livers. Arch Biochem Biophys 241:275-80

Kavlock RJ, Allen BC, Kimmel CA, *et al.* 1995. Dose-response assessment for developmental toxicity: IV. Benchmark doses for fetal weight changes. Fund Appl Toxicol 26:211-22

Kodell RL, Chen JJ, and Gaylor DW. 1995. Neurotoxicity modeling for risk assessment. Toxicology and Pharmacology 22:24-9

Leroux BG, Leisenring WM, Moolgavkar SH, *et al.* 1996. A biologicallybased dose-response model for developmental toxicology. Risk Anal 4:449-58

Lovett RA. 2000. Toxicologists brace for genomics revolution. Science 289:536

Melnick RL, Huff J, Chou BJ, *et al.* 1990. Carcinogenicity of 1,3-butadiene in C57BL/6 x C3HF1 mice at low exposure concentrations. Cancer Res 50:6592-9

Mollerup S, Ryberg D, Hewer A, *et al.* 1999. Sex differences in lung CYP1A1 expression and DNA adduct levels among lung cancer patients. Cancer Res 59:3317-20

Perera FP. 2000. Molecular epidemiology: On the path to prevention? J Nat Cancer Inst 92:602-12

Plenge-Bonig A and Karmaus W. 1999. Exposure to toluene in the printing industry is associated with subfecundity in women but not in men. Occup Environ Med 56:443-8

Renwick AG and Lazarus NR. 1998. Huamn variability and noncancer risk assessment-an analysis of the default uncertainty factor. Regul Toxicol Pharmacol 27:3-20

Renwick RG. 1991. Safety factors and establishment of acceptable daily limits. Food Addit Contam 8:135-50

Santostefano MJ, Wang X, Richardson VM, *et al.* 1998. A pharmacodynamic analysis of TCDD-induced cytochrome P450 expression in multiple tissues: Dose- and time-dependent effects. Toxicol Appl Pharmacol 151:294-310

Selevan SG, Kimmel CA, and Mendola P. 2000. Identifying critical windows of exposure for children's health. Environ Health Perspect 108 (Suppl 3):451-5

Shuey DL, Lau C, Logsdon TR, *et al.* 1994. Biologically based dose-response modeling in developmental toxicology: Biochemical and cellular sequelae of 5-fluorouracil exposure in the developing rat. Toxicol Appl Pharmacol 126:129-44

Slikker Jr W, Crump KS, Andersen ME, *et al.* 1996. Biologically based, quantitative risk assessment of neurotoxicants. Fund Appl Toxicol 29:18-30

Slikker Jr W, Scallet AC, and Gaylor DW. 1998. Biologically-Based dose-response mode for neurotoxicity risk assessment. Toxicol Letters 102-103:429-33

Taskinen HK, Kyyronen P, Sallmen M, *et al.* 1999. Reduced fertility among female wood workers exposed to formaldehyde. Am J Ind Med 36:206-12

Tateishi T, Nakura H, Asoh M, *et al.* 1997. A comparison of hepatic cytochrome P450 protein expression between infancy and postinfancy. Life Sci 61:2567-74

Tee LB, Gilmore KS, Meyer D, *et al.* 1992. Expression of glutathione S-transferase during rat liver development. Biochem J 282:209-18

Trizna Z, de Andrade M, Kyritsis AP, *et al.* 1998. Genetic polymorphisms in glutathione *S*-transferase and , N-acetyltransferase, and *CYP1A1* and risk of gliomas. Cancer Epidemiol Biomark Prev 7:553-5

USEPA (U.S. Environmental Protection Agency). 1998. Methods for derivation of inhalation reference concentrations and application of inhalation dosimetry. EPA-600/8-90-066. Health Committee of the Science Advisory Board, Washington, DC, USA

Waring JF and Urlich RG. 2000. The impact of genomics-based technologies on drug safety evaluation. Ann Rev Toxicol 40:335-52

Weller E, Long N, Smith A, *et al.* 1999. Dose-rate effects of ethylene oxide exposure on developmental toxicity. Toxicol Sci 50:259-70

Wrighton SA, Molowa DT, and Guzelian PS. 1988. Identification of a cytochrome P-450 in human fetal liver related to glucocorticoid-inducible cytochrome P-450HLp in the adult. Biochem Pharmacol 37:3053-5

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1421-1444 (2002)

Improving Risk Assessment: Research Opportunities in Dose Response Modeling to Improve Risk Assessment[1]

Lauren Zeise,[2] Dale Hattis,[3] Mel Andersen,[4] A. John Bailer,[5] Steve Bayard,[6] Chao Chen,[7] Harvey Clewell,[8] Rory Conolly,[9] Kenny Crump,[8] David Dunson,[10] Adam Finkel,[11] Lynne Haber,[12] Annie M. Jarabek,[13] Ralph Kodell,[14] Daniel Krewski,[15] Duncan Thomas,[16] Todd Thorslund,[6] and James T. Wassell[17]

ABSTRACT

Substantial improvements in dose response modeling for risk assessment may result from recent and continuing advances in biological research, biochemical techniques, biostatistical/mathematical methods and computational power. This

[1] Report of the "Dose Response Models" Breakout Group, of the Workshop "Future Research for Improving Risk Assessment Methods: Of Mice, Men, and Models," August 16-18, 2000, Snowmass, Colorado, sponsored by the National Institute for Occupational Safety and Health, National Institute of Environmental Health Sciences, US Environmental Protection Agency, Chemical Manufacturers Association (currently American Chemistry Council), and United Auto Workers. Invited participants of the workgroup are the authors of this paper. Additional participants and contributers to the discussion were: Ronald Brown, James Deddens, Brian Miller, Robert Park, Raghupathy Ramanathan, Jan Sassaman, Edward Slavin, Randall Smith, Robert Stenner, Jana Sorvari, and David Utterback.

[2]California Environmental Protection Agency, Office of Environmental Health Hazard Assessment, Oakland, California; breakout group chair. [3]Research Professor, Marsh Institute, Clark University, 950 Main Street, Worcester, Massachusetts 01610, (p) (508)751-4603, (f) (508)751-4600, dhattis@aol.com; breakout group rapporteur. [4]Colorado State University, Fort Collins, Colorado. [5]Miami University, Oxford, Ohio. [6]US Occupational Safety and Health Administration; Washington, DC, submitted research proposal but was unable to attend workshop. [7]US Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, Washington, DC. [8]K.S. Crump Group/ICF Consulting, Ruston, Louisiana. [9]CIIT Centers for Health Research (previously Chemical Industry Institute of Toxicology), Research Triangle Park, North Carolina. [10]National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina. [11]US Occupational Safety and Health Administration; Washington, DC. [12]Toxicology Excellence for Risk Assessment, Cincinnati, Ohio. [13]US Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, Research Triangle Park, North Carolina. [14]National Center for Toxicological Research, Jefferson, Arkansas. [15]University of Ottawa, Ottawa, Ontario. [16]University of Southern California, Los Angeles, CA. [17]National Institute for Occupational Safety and Health, Morgantown, WV.

1080-7039/02/$.50

report provides a ranked set of recommendations for proposed research to advance the state of the art in dose response modeling. The report is the result of a meeting of invited workgroup participants charged with identifying five areas of research in dose response modeling that could be incorporated in a national agenda to improve risk assessment methods. Leading topics of emphasis are interindividual variability, injury risk assessment modeling, and procedures to incorporate distributional methods and mechanistic considerations into now-standard methods of deriving a reference dose (RfD), reference concentration (RfC), minimum risk level (MRL) or similar dose-response parameter estimates.

Key Words: interindividual variability, injury risk, Bayesian analysis, latent variables, Markov Chain Monte Carlo analysis.

INTRODUCTION

The recent advances in biological research, biochemical techniques, molecular epidemiology, biostatistical/mathematical methods, and computational power provide opportunities for considerable improvement in the assessment of dose response relationships for risk assessment. This paper describes the discussion and results of a one-day meeting of thirteen invited participants (listed as authors) of a dose response workgroup and several additional attending experts in dose response modeling. The workgroup meeting was part of a three-day workshop exploring research possibilities to improve risk assessments done to support occupational and environmental health policies and standards. The charge to the participants of the workgroup was to identify five specific research areas that would significantly improve dose response modeling for occupational and environmental risk assessment. Each invited participant was asked to prepare a brief description of research programs that could be the subject of a request for proposals. Table 1 shows titles and authors of each initial written proposal. This paper draws from the proposal write-ups, often verbatim, to describe the areas of research the workgroup found most promising.

The dose-response workgroup meeting proceeded in three stages. First proposals were sorted into broad topic areas, and then each proposal was presented and discussed. Additional topic areas and ideas for proposals were then solicited from the group. Next there was an extended period during which the individual proposals were further refined, extended, and consolidated, into ten groups. Finally, each participant was asked to rank the proposal groups giving five points to the top choice in terms of desirability for funding, four for the second choice, three for the third, two for the fourth, one for the fifth and zero for the remaining five proposal groups. In all, thirteen people participated in this ranking exercise.

PROPOSED LINES OF DOSE RESPONSE RESEARCH

The ten areas the group chose to explore, in ranked order are given in Table 2. Some proposals benefited by the multiplicity of ideas attached within the various groups—and the process of grouping is the principal reason why there is not a one-to-one mapping of the initial proposals listed in Table 1 to the final ranked items in Table 2. The groupings may well have been a disadvantage to some of the

Table 1. Initial research ideas proposed to improve dose response modeling for risk assessment.

Participant	Research Proposal
Mel Andersen	Mechanistic Models for the Impact of Lifestyle Factors on Metabolism, Pharmacokinetics, and Toxicity of Workplace Chemicals in Diverse Human Populations
John Bailer	Developing Strategies for Quantitative Risk Estimation for Hazards of Occupational Injury
Chao Chen	Advances in Molecular Research and Risk Assessment
Harvey Clewell	Accounting for Pharmacokinetic Uncertainty and Variability in Risk Assessment
Rory Conolly	Risk Assessment for What?
Kenny Crump	Investigation of Prevalence of Hormesis
David Dunson	Risk Assessment Based on Multiple Reproductive and Developmental Endpoints
Lynne Haber	1. Use of Data-Derived Uncertainty Factors in Human Risk Assessment 2. Characterizing Uncertainty in RfDs and RfCs
Annie Jarabek	1. Characterizing Variability to Decrease Uncertainty 2. Extending the Range of Observation
Ralph Kodell	Development of Mechanistic Biomathematical Models that Predict Less-than-Background Risk of Cancer at Low Doses of Toxicants
Duncan Thomas	Incorporating Genetics into Risk Assessment
Todd Thorslund Steve Bayard James Wassell	Proposed Research to Strengthen the Rationale for Extrapolating Animal Cancer Bioassay Results to Humans Based on Relative Risk Occupational Injury Risk Assessment

proposals that might have benefited from being joined with others, but were presented in relative isolation from other proposals. In any event, the differences between the second through the seventh ranked proposal groups are relatively small, and the eighth proposal also received significant support. This paper discusses the first eight ranked groups, rather than the five top rated proposals originally requested by the organizers.

CHARACTERIZING INTERINDIVIDUAL AND INTERSPECIES VARIABILITY IN SUSCEPTIBILITY

An important consideration in regulators' decisions to mitigate risks is that people differ in their responses to environmental and occupational exposures. During the last 15 years, innovations in both observational study design and risk analysis methodology have led to improved descriptions of interindividual differences in risk (see, *e.g.*, Bogen and Spear 1987; Bois *et al.* 1996 and 1999; Hattis and Burmaster 1994; Cullen and Frey 1999) and in the mechanistic bases of pharmacokinetic and pharmacodynamic determinants of exposure-dose-response across species used in laboratory testing (USEPA 1994; Jarabek 1995a,b; USEPA 1996; Schlosser and Bogdanffy 1996). Yet, mathematical tools to reach policy decisions that adequately address characterization of exposure-dose-response in a mechanistic fash-

Table 2. Summary ranking of ideas for the dose response modeling workgroup.

1. Characterizing Interindividual and Interspecies Variability in Susceptibility
2. Models for Injury Risk Assessment – Including Development and Characterization of Risk Estimates and Exposure Metrics
3. Adaptations and Modifications to Existing Standard Procedures (*e.g.*, to Derive RfDs)
4. Acceptability Criteria for Mechanistic Hypotheses and Data
5. Modeling Analyses for Multiple Endpoint Data (Especially for Different Endpoints on Different Scales)
6. New Mechanistic Models of Carcinogenesis
7. Combining Data of Different Types in Risk Analyses
8. Exploring Evidence and Models for Complex Dose Response Relationships in the Context of Homeostasis
9. Ideals for Risk Analysis/ Relationship with Societal Decision Processes
10. Interspecies Extrapolation Based on Relative Risk

ion or that address heterogeneity have not been adopted by regulatory agencies for routine use. The standard practice is to base standards on the population risk, sometimes in combination with ad hoc procedures thought to be conservative (*e.g.*, use of most sensitive species) to estimate an upper bound on population risk. It is hoped that decisions based on an upper bound so calculated will provide protection for sensitive individuals. To make such estimates, simple dose-time-response or stochastic models of the carcinogenesis process are fit to experimental or epidemiological data. Noncancer methods rely on designation of a sentinel adverse effect in a putatively sensitive species, and then apply interspecies and intrahuman "uncertainty factors" that have been traditionally based on empirical, not mechanistic, motivation (Jarabek 1995a).

Break-out group participants identified several lines of research that would lead to improved methods to assess interindividual variability in dose response. These are outlined below.

A. Methods to Incorporate Genetic Determinants of Heterogeneity to Model Variability in Cancer Risk

It is generally accepted that cancer is a genetic disease, involving somatic mutations or other changes to DNA that can be induced by environmental exposures to carcinogens. The mathematical models used in risk assessment such as the Armitage-

Doll multistage model and the Moolgavkar-Knudson two-mutation clonal expansion model are based on fundamental concepts about mutation and heterogeneity. However, neither model accounts for an individual's specific genetic make-up, or population heterogeneity in unmeasured genetic or environmentally-induced factors. Some germline mutations can also produce a hereditary predisposition to cancer or an unusual sensitivity to environmental carcinogens. For example, Ataxia-Telangiectasia patients (homozygous for the *ATM* mutation) are exquisitely sensitive to ionizing radiation, and *ATM* heterozygotes are at elevated risk. Other examples include predisposing genes such *BCRA1* and *BCRA2* for familial breast and ovarian cancer, *APC* for familial polyposis coli, and the mismatch repair genes *MLH1* and *MSH2* for hereditary nonpolyposis colorectal cancer. Polymorphisms in genes involved in the metabolic activation of pre-carcinogens to their active form or their deactivation, also may confer a larger population attributable risk (e.g., *NAT*1 and *NAT2* activation of aromatic amines). A number of these polymorphisms affect appreciable fractions of the population. Finally genomic instability is an important mechanism in certain kinds of cancer, such as hereditary nonpolyposis colon cancer, in which mutations in one gene produces extensive loss of DNA replication fidelity, leading to high rates of somatic mutation at other loci involved in the carcinogenic process. Absent specific information about an individual's genetic makeup, substantial heterogeneity between individuals in their baseline risk and sensitivity to environmental carcinogens is expected.

Duncan Thomas proposed development of mathematical methods to model heterogeneity that would take advantage of the powerful molecular tools currently available for genetic testing (*e.g.*, including whole genomic scans and gene-expression arrays). Priorities noted for possible development included:

1. Methods for incorporating variation in identifiable genetic factors and for estimating residual heterogeneity

2. Stochastic models of carcinogenesis incorporating genomic instability. The key idea is to allow for the possibility that an early event in carcinogenesis could lead to a somatic mutation in a mismatch repair or other regulatory gene, inducing a complex cascade of subsequent events. The data from microdissected tumors on clonal variation in molecular markers within a single tumor could potentially be exploited for this purpose.

3. Methods for incorporating information on metabolic genes into physiologically-based pharmacokinetic models to describe complex pathways (elaborated further below).

Expanding on these ideas, the group recognized the value of methods exploiting data for specific metabolic gene polymorphisms, induction of activating and detoxifying enzymes, DNA repair, and major gene changes on defined germ-line mutations in genes along known molecular pathological pathways in cancer. The quantitative work of Hattis and Barlow (1996) based on variability in phenotype rather than genotype observations (activities of metabolic activation, inactivation, and DNA repair) could be extended through greater use of in vivo observations of relevant enzyme activities, and procedures to separate underlying variability from

measurement errors (see, *e.g.*, Hattis and Silver 1994). Methods to measure variability in both susceptibility and relevant exposures by examining the pattern of age-specific cancer incidence could be further developed. For example, Finkel (1987, 1995) and others (Manton and Stallard 1979; Manton *et al.* 1986) using techniques of heterogeneity dynamics have extracted estimates of variability from age specific incidence data. Heterogeneity dynamics provide methods for describing changing characteristics of a heterogeneous population as its members age.

B. Characterizing Interindividual Variability in Effective Doses and Risk within the Framework of Physiologically Based Pharmacokinetic Models

Occupational risks from chemicals vary among individuals in a workplace due to varying exposure patterns and differences among workers' rates for enzymatic activation and detoxification of chemicals, health status, and other factors. Enzyme activity is determined by various intrinsic and extrinsic factors. These include genetic makeup, lifestyle, pharmaceutical usage, alcohol intake, health status, and age. For example, drugs serve as potential enzyme inducers. Dietary factors, health status (*e.g.*, ketotic states associated with diabetes), and alcohol alter the activity of an important oxidative enzyme, P450-2E1, involved in metabolism of some workplace chemicals. The impacts of complicated, time-dependent interactions of multiple factors on metabolism, and the complex exposure patterns in the workplace, are often overlooked in the setting of occupational standards.

Melvin Andersen proposed research to investigate conditions under which interactions are likely to enhance toxicity, and to identify members of the population most at risk. The research would explore the quantification of increased risk, and the identification of activities and lifestyle factors that lead to higher workplace risks. How such characterizations could influence PEL or TLV standard setting activities for occupational exposures would also be examined.

Past assessments of interactions have mostly involved observation of altered pharmacokinetic behavior of compounds in healthy human volunteers under controlled exposure conditions. Conclusions were typically drawn from observed alterations in the kinetic properties, for example, in the presence of chemical mixtures, with intake of small to moderate amounts of alcohol, or with varying levels of exercise. It is proposed that methods now be developed utilizing physiologically based toxicokinetic (PBTK) and toxicodynamic (TD) models. PBTK models describe the disposition, metabolism and transport of chemicals and metabolites in various tissues of the body, and certain TD models characterize alteration of enzyme levels, receptor levels, and effects due to exogenous compounds or dietary factors (see, *e.g.*; Chien *et al.* 1997). The PBTK/TDs modeling framework provides the means for integrating critical data of different types into the assessment. This includes data from mechanistic studies of pharmacokinetics and enzyme induction in animals, and limited studies in human volunteers or with human tissue. The overall goal of the modeling research would be to define the temporal and lifestyle factors that lead to variability in responses to chemical exposure in diverse worker populations, and to quantify that variability.

Variability in risk due to interindividual differences in pharmacokinetics has been addressed in recent risk assessments, including that by the Occupational Safety

and Health Administration for methylene chloride, the US Environmental Protection Agency's draft assessment for trichloroethylene, and California's draft assessment for tetrachloroethylene (Cal/EPA OEHHA, 2000, based on Bois *et al.* 1996). Recent uses of PBTK models have also evaluated the adequacy of current standard setting approaches when variability is taken into account (Thomas *et al.* 1999).

In a related proposal, Harvey Clewell proposed research to investigate the hierarchical Bayesian approach to characterizing uncertainty and variability in pharmacokinetic models for cancer risk assessment, as developed by Frederic Bois and others (Bois *et al.* 1996; Gelman *et al.* 1996). The power of PBTK models is obtained at the expense of using a large number of parameters, some of which may vary significantly among individuals (*e.g.*, the pharmacokinetic constants) and few of which are known with precision. The impact of parameter uncertainty in PBTK models has often been evaluated using a Monte Carlo approach (Bois *et al.* 1990; Allen *et al.* 1996; Clewell and Andersen 1996; Bailer and Dankovic 1997; Dankovic and Bailer 1994; Hattis 1990; Portier *et al.* 1989), wherein specified probability distributions are randomly sampled for each model parameter and the PBTK model is run. The process is repeated numerous times to define a probability distribution for the desired PBPK dose metric. In typical applications of this approach, variability in individual toxicodynamic response susceptibility per unit of internal dose is not addressed, and if it is, it is not decoupled from uncertainty (*e.g.*, due to measurement error). Further, although model parameters are correlated, only limited correlation is typically assumed, or it is ignored altogether.

Hierarchical statistical models within a Bayesian framework have been applied to disentangle model uncertainty from variability, using the computational technique of Markov chain Monte Carlo simulation (Bois 1999; Bernillon and Bois 2000). Markov chain Monte Carlo (MCMC) procedures are widely useful for fitting of hierarchical models that incorporate individual-specific parameters. For example, MCMC algorithms have been used for estimating the population distribution of physiological parameters based on data for the pharmacokinetics of the chemical in different individuals (Bois *et al.* 1996; Jonsson and Johanson 2000). The approach adopts a hierarchical population model to enable uncertainty and variability in an individual's response to be distinguished from the variability of individual responses within the population. The same pharmacokinetic model structure applies to all individuals, but model parameters vary among individuals. The Bayesian framework provides a formal structure for combining prior knowledge on parameters from the scientific literature with data from pharmacokinetic experiments, to generate posterior distributions for any given parameter value. Thus widely different types of data can be integrated, for example, from studies of distribution and elimination in human volunteers, in vitro and in vivo metabolic studies in experimental animals, and physiological measurements from various sources. The overall approach provides the statistical foundation to support PBPK model calibration that is lacking in most PBPK applications (Bernillon and Bois 2000; Kohn 1995 and 1997).

Some members of the group raised concerns over cases when the use of this approach results in parameter estimates that differ substantially from values expected based on a priori knowledge. This may occur when the "prior" distributions assumed are relatively broad, and parameter values are significantly influenced by distantly related parameters with extensive observations used in the Bayesian "up-

dating". The programs and analyses are complicated, and there is the concern regarding how possible subtle misspecification of model structure and possible underestimation of the uncertainty in available observational data might influence posterior estimates of parameter values. The proposed study is, using both actual and simulated data, to distinguish the conditions under which reestimation of model parameters with the approach produces relatively accurate results from those under which it may be misleading. This might be done, for example, by applying the approach to a well-characterized system, deleting information, and predicting the deleted data from the remaining information (a technique referred to as "cross validation").

Other proposed work in this area would focus on greatly needed research on model uncertainty and lack of identifiability in complex models. Both may be dealt with effectively using a Bayesian approach. For example, if a biologically realistic model is under identified, prior information can be brought in to enable Bayesian identifiability. In addition, if there are definable uncertainties in the model, one can utilize Bayesian model averaging techniques.

C. Human Variability in Baseline Values for Parameters as Predictors of Non-Cancer Susceptibility

One important determinant of the population distribution of susceptibility to non-cancer toxic insults is the baseline distribution in the human population of functions and functional reserve capacities for physiological process such as kidney, lung, or liver functions. We define functional reserve capacity as the amount of change in a physiological parameter needed to produce abnormal function or an adverse outcome. As an example from Hattis *et al.* (1999) the distribution of low density lipoprotein (LDL) is considered. LDL is thought to be an important physiological parameter likely to be on the causal pathway to cardiovascular disease. The variability of LDL in the population loosely reflects differing susceptibility to cardiovascular disease. The distribution of functional capacity among those not receiving intervention might be described as the distribution of differences between a standard cutoff for clinical intervention and measured values in the population. Baseline observation studies, such as surveying via NHANES LDL levels, have the advantage that they do not require deliberate administration of toxicants or drugs to humans.

From LDL and other parameters one can predict the risk of cardiovascular disease. Similarly, from study of other continuous variables related to serious health outcomes it may be possible to develop relationships for use in predicting non-cancer risks. Examples include, birth weight and infant mortality; sperm quality parameters and male fertility performance; forced expiratory volume in one second as a predictor for general cardiovascular mortality; iodine deficiency and thyroid pathology. Also by examining the distribution of indicators of functional capacity one can gauge the extent to which the certain risk assessment practices are protective (*e.g.*, assignment of certain uncertainty factors). Dale Hattis proposed the development of data to explore the use of baseline observations in quantitative non-cancer risk assessment procedures. For additional discussion of the potential for this type of study, see Hattis (1998) and Hattis *et al.* (1999).

D. Variability in Mechanistic Determinants of Chemical Disposition (*e.g.* Related to Age, Species, Sex, Disease State, for Oral, Inhalation, and Dermal Exposures)

As discussed above, dosimetry based on PBTK models has become a useful tool to adjust for differences in delivered and internal dose across species and within human populations. Examples of applications in risk assessments for gases include: formaldehyde (CIIT 1999), tetrachloroethylene (Bois *et al.* 1996), vinyl acetate (Bogdanffy *et al.* 1999), EGBE (IRIS 1999), and vinyl chloride (IRIS 2000). Dosimetry modeling was also a key aspect of the National Ambient Air Quality Standard for particulate matter in 1996a (USEPA 1996). Reduced mechanistic model structures and empirical models of mass transport have formed the basic default procedures for the calculation of human equivalent concentrations in the U.S. Environmental Protection Agency reference concentration methods (USEPA 1994; Jarabek 1995b).

Annie Jarabek proposed research to fill substantive gaps in the anatomical, physiological and mechanistic data needed to explicitly describe the major factors influencing chemical disposition defined to encompass the processes of deposition (*e.g.* primary deposition of an inhaled toxicant or particles on airway surfaces), uptake, distribution, metabolism and elimination as well as subsequent toxicant-target interactions, *i.e.*, mode of action (Jarabek 2000). Such data would allow a comprehensive characterization of the exposure-dose-duration-response continuum across species so that variability can be addressed by describing differences in the mechanistic factors that determine disposition and pathogenesis. She proposed the acquisition of such data for different ages and genders in experimental animals and humans and for different disease states (*e.g.*, COPD) in humans for inhalation, oral and dermal routes.

Parameters of interest include the anatomical parameters of airway lengths, portal of entry (respiratory tract, GI, dermis) tissue thickness, cell types and locations, and other physiological parameters such as ventilation rates, GI transport rates, dermal transport rates, metabolism, and the fraction of specific types of cells in various phases of the cell cycle. These data would support improved dosimetry modeling to increase the accuracy of descriptions of dose differences among species and within the human population. For example, the respiratory tract of children differs dramatically from that of the adult in anatomical structure and ventilatory pattern (*e.g.*, the oral-nasal switching point between nose versus mouth-breathing with exertion) and has not been well-described (Dietert *et al.* 2000). Simple scaling assumptions do not adequately address variability in uptake via inhalation due to these age/developmental differences. Further, the collected data would provide a basis for assessing confidence in the description, thereby informing the magnitude of the interspecies and intra-human uncertainty/adjustment factors used in risk assessment.

This research would inform efforts to assign uncertainty factors for pharmacokinetic and pharmacodynamic differences (see below) and complement a Federal interagency collaborative effort to develop a suite of dosimetry models for oral, inhalation and dermal exposures. Such models have potential applications for both cancer and non-cancer endpoints. Development of such models therefore contributes to a harmonized approach for cancer and non-cancer risk assessment, and is

consistent with USEPA's *Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry* (USEPA 1994) and proposed carcinogen guidelines (USEPA 1996b).

E. Comparative Studies of Variability in Susceptibility to Toxic and Other Effects in Animals and People

Traditional protocols for animal toxicology experiments usually go to considerable lengths to minimize the variability among tested animals. This is because, in general, the more the variability, the larger the sample size required to demonstrate differences between the effects of experimental and control exposures. Variability is therefore deliberately restricted in most animal experimental work by using genetically homogenous animals, a single age class of subjects ("young adults" usually) and sometimes a single sex of animals. Before initiation of treatment the animals usually have been subject to relatively uniform environmental stimuli, including uniform and unchanging diet. Efforts are also made to maintain healthy animal colonies, as free of infections as possible. Thus the laboratory animal is not exposed to the diversity of living conditions of wilder, outbred populations. And, of course, there are also no deliberate neuroactive drug exposures (*e.g.*, narcotics, alcohol, tobacco, caffeine) unless they are explicit subjects of experimental study (Hattis 1996).

Studies are needed to experimentally assess in animal systems how much some or several of these common practices actually reduce variability in the doses producing defined toxic responses. Studies are also needed to assess the general distribution of comparative degrees of variability in toxic response for the free-living human population, relative to the types of animal groups usually used for toxicological testing. That is, how often is there a large difference in animal/human variability?

F. Procedures to Utilize Interindividual Variability Information in Cancer Risk Assessment

This would be an extension of the efforts described above under genetic determinants of variability (A), and (D) variability of mechanistic determinants of chemical disposition. Efforts are needed to elucidate an appropriate set of operational procedures to incorporate available generic and chemical-specific information on human interindividual variability in susceptibility to carcinogenesis into risk analyses utilized for risk management under a range of regulatory authorities and risk management criteria. For an exploration of some different potential implications of variability *vs* uncertainty for risk management under different regulatory authorities see Hattis and Anderson (1999), Krewski *et al.* (1999), Hattis and Minkovitz (1996), and Bois *et al.* (1996).

MODELS FOR INJURY RISK ASSESSMENT INCLUDING DEVELOPMENT AND CHARACTERIZATION OF RISK ESTIMATES AND EXPOSURE METRICS

James T. Wassell and John Bailer proposed the development of methodology for assessing risk and exposure metrics for occupational injury. The application of risk

assessment methods to this problem provides many opportunities for innovative development, but more importantly has the potential to significantly contribute to public health improvements. The incidence of occupational and non-occupational injuries is substantial, with a relatively early average age of occurrence for most serious injuries. In comparison to cancers and manifestations of cardiovascular disease, social effect measures such as years-of-potential-life-lost (YPLL) (Gilbert *et al.* 1998) are potentially greater for injury. The field of injury risk is just beginning to define relevant endpoints for social policy evaluation such as working lifetime risk (Fosbroke *et al.* 1997; See and Bailer 1998).

Among the most interesting methodological challenges is the definition of relevant causal and potentially confounding exposures. The most commonly available occupational injury data are based on the number of hours worked in the workplace and the size of the workforce. The proportion of time actually at risk (for example while using a particular type of machine or while engaged in material handling) is not distinguished from the portion of time worked while not at risk or at risk in other ways. Wassell (1989) discussed probability considerations for commonly used methods of statistical analysis of injury data. Poisson regression and other models of injury occurrence that would be more descriptive and helpful for evaluating the efficacy of interventions would include prospective or blind retrospective assessments. Modifiable targets for preventive action include factors such as the time spent operating unguarded machines, or the number of maintenance operations that could lead to a violation of lock-out/tagout precautions. Methods should account for personal confounding factors (*e.g.* worker age, job tenure, training) to control for non-engineering contributions to injury risks.

Another important area of research is the development of exposure-response models for chronic repetitive motion injuries, analogous to cancer multistage or classical toxicological probit dose response models. Such models should be based on considerations of physiological factors and data on the variability in the frequency and intensity of repetitive motion stresses created by particular work tasks, as well as variability in physiological responses to repeated subclinical injury events. They ideally should also be based on a mechanistic theory for how irreversible or very slowly reversible injury events happen. Finally, work is needed on priority-setting for injury prevention efforts.

ADAPTATIONS AND MODIFICATIONS TO EXISTING STANDARD PROCEDURES (*e.g.*, FOR DERIVING RFD'S)

There were three proposals aimed at improving on existing default procedures for risk assessment:

- Further explorations of "data-derived" uncertainty/adjustment factors (UFs) through the development of
  - — data bases for existing uncertainty factors
  - — probabilistic distributions of UFs adapted to specific types of agents and effects

— probabilistic approaches for reference dose (RfD) derivation

— criteria for using data-derived factors

- Development of quantitative guidance for expected risks posed by an RfD or, alternatively, the probability that the population threshold is below calculated RfD levels.

- Approaches to recognize influences such as age, latency, and pattern of exposure in bioassays and exposed human populations

The first two were originally developed by Lynne Haber, and the third was generated in group discussion. The proposals are briefly discussed in the following.

A. Data-Derived UFs

Traditionally, default UFs of 10-fold have been used for extrapolation from animals to humans, and for accounting for variability among humans due to sensitive populations and individuals. Development and examination of data bases, for example on human and rodent chronic toxicity for certain classes of agents, may provide the rationale for a different default for application in specific circumstances. A factor so identified is termed a "data derived UF;" a few assessments have been conducted using such factors (*e.g.*, Bogdanffy and Jarabek 1995; Bogdanffy *et al.* 1999; Dourson *et al.* 1998; IPCS 1998). Renwick (1991) and Renwick and Lazarus (1998) suggest interspecies and intraspecies UFs can each be broken into toxicokinetic and toxicodynamic components, based on the relative contributions of these components for a number of chemicals examined (*e.g.*, for the interspecies factor, a factor of 3.3 for each component). Similarly, reduction of the interspecies UF from 10 to 3 is used to address uncertainty in laboratory animal to human extrapolation for Reference Concentrations (RfCs) when dosimetric adjustments are made for species differences in toxicokinetics (USEPA 1994; Jarabek 1995a). As a further step, as outlined in the second bulleted section below, the use of distributions, rather than single point estimates, has been proposed (Hasseblad and Jarabek 1996; Baird *et al.* 1996; Hattis *et al.* 1999).

The issue of criteria for replacement of default UFs has been raised (*e.g.*, Meek 2000). (Since the workshop, the term "chemical-specific adjustment factors" has replaced the term "data derived UFs." Guidance for the use of data in the development of chemical-specific adjustment factors for interspecies differences and human variability have been developed by the IPCS, and are available at http://www.who.int/pcs/pubs/pub_list.htm.) Whether the critical determining factor for toxicity has been identified in developing the factors can be questioned. To derive a factor for the interspecies toxicokinetic component, laboratory animal to human ratios of the values of some human toxicokinetic parameters have been compiled, but it is usually not clear that these ratios are adequate surrogates for the ratios of critical tissue doses in the two species. Similarly, it is currently unclear how to translate information on human variability in key metabolic parameters into an uncertainty factor for human variability in kinetics. Descriptions of human variability in the critical pharmacokinetic parameters through the PBPK modeling of well-

characterized model chemicals acting via selected modes of action may help to elucidate whether such variability is adequately described by the variability in certain metabolic parameters (*e.g.*, a key enzyme's V_{max} or V_{max}/Km ratio).

Risk assessment applications involving uncertainty factors less that the traditional defaults of 10-fold for inter- and intraspecies variation are beginning to appear. The National Research Council (2000) has used such factors to establish acute exposure guideline levels (AEGLs) for highly hazardous substances such as mono and dimethylhydrazine, when the available data are deemed sufficient to support such practice. Further details of the risk assessment methodology used in establishing AEGLs are given by the National Research Council (2001).

B. Characterizing Uncertainty in RfDs and RfCs

Traditionally, RfDs and RfCs have been derived by dividing "no observed adverse effect levels" (NOAELs) observed in toxicological experiments or epidemiology by fixed uncertainty factors (although recently there has been some use of "benchmark doses" instead of NOAELs). A number of investigators have recently sought to replace fixed uncertainty factors with a probability density function, or PDF, that characterizes the uncertainty in the size of a "true but unknown" scaling factor (*e.g.*, Baird *et al.* 1996; Slob and Pieters 1997; Price *et al.* 1997; Swartout *et al.* 1998). For the UF of interest, the PDF is derived from data for that factor (*e.g.*, interspecies difference) for a reasonably large sample of chemicals. In addition to developing a PDF for the uncertainty factor, the uncertainty in the identification of a no effect level from animal study observations can be addressed (*e.g.*, Leisenring and Ryan 1992). Some residual risk is expected at the NOAEL due to the limited numbers of animals studied and other limitations of experimental design. The distributions of the UFs and for the animal no observed effect level may be convolved to develop a probabilistic distribution for the RfD. A probabilistic RfD has been developed for methylmercury (Clewell *et al.* 1999), based on a distribution of the input variables to the PBPK model used to convert hair mercury concentrations to a chronic intake rate.

PDFs for the RfD would provide useful information for both risk assessment and characterization, including uncertainty disclosure, information on the degree the RfD may be protective, and uncertainty in the overall health protection process. However, much work is needed before such distributions can be reliably derived, particularly the development of a statistical framework for the analysis. The derivations of PDFs for UFs typically involve comparisons of NOAELs from experiments conducted using different numbers of animals and experimental protocols. Proposed research includes the development of a statistical framework, for example to account for the uncertainties in the derived ratios, the residual risk at the NOAEL, and the updating of distributions based on chemical specific information. Work to date has not considered in detail how to take into account chemical-specific information in the development of the PDFs. The impact of ad hoc changes in UFs based on chemical specific data considerations on the characteristics of the appropriate PDF could be explored.

Dale Hattis raised a distinct but related research need to develop data and methodology to examine the assumption of true population thresholds embodied

in the traditional RfD concept. The recent work of Hattis et al. (1999) is based on the idea that human population distributions of susceptibility may be described by continuous lognormal or other more complex distributions, implying finite and potentially estimable risks at various levels of exposure above and below traditional RfDs. Also, when the disease of concern occurs in the population by the same mechanism as that of the toxicant in question, a population threshold is unlikely and some risk from exposure is expected. In this framework, RfDs would be defined as the human dose or exposure levels expected to produce no more than a specific incidence of harm at a minimal severity with a defined degree of confidence.

C. Influences of Age, Latency, and Pattern of Exposure and Other Factors

This is related to proposal #E discussed under the first program group above on variability. The work here would be to explore whether the traditional uncertainty factors adequately capture heterogeneity given the necessary limitations in toxicological study design.

CRITERIA FOR ACCEPTABILITY OF MECHANISTIC HYPOTHESES AND DATA

Rory Conolly proposed the initiation of a process to develop consensus criteria for the use on mechanistic information in risk assessment and management. He noted that the movement away from routine use of default assumptions towards more data-based approaches to risk assessment is in step with the increasing sophistication of laboratory methods for the study of biochemical mechanisms and the identification of trace levels of contaminants. These developments are raising questions of how data-based assessments should be structured and of the selection of endpoints for assessment.

Sensitive analytical techniques and the new biochemical techniques such as genomics and proteomics have the potential to link contamination at widespread environmental levels with changes in the expression levels of genes or the concentrations of specific proteins. In some cases, as for TCDD, biologically plausible hypotheses can be developed, suggesting that such early biochemical effects are the precursors of downstream frankly toxic effects (USEPA 2000). Dose response information for the early biochemical effects might be interpreted as indicative of the expected shape of the dose response curve for the downstream frankly toxic effect. As a result, risk assessors and risk managers are faced with several issues, including:

1. Do biochemical changes, such as a change in the expression level of a gene or the amount of a protein in a cell, constitute "adverse" effects?

2. How often and under what circumstances does the shape of the dose response curve for an early effect, such as a change in gene expression, inform us about the expected shape of the dose response curve for a frankly toxic effect further "downstream" in the causal pathway of harm?

3. Is homeostasis a determinant of the shape of the dose response curve? If so, what are the implications of homeostasis for the shape of the dose response curve?

4. Does the additivity to background argument (Crump *et al.* 1976) apply if homeostasis is operative or if pharmacokinetic nonlinearities exist?

5. When the understanding of mechanism is incomplete (as it always is), what criteria should be used to judge the acceptability of hypothetical linkages between early biochemical and later, frankly toxic effects?

6. How should mechanistic information be used in risk assessment when available data do not allow discrimination between alternative, plausible mechanistic hypotheses? Should a spectrum of assessments be developed based on the alternative hypotheses?

The broader risk assessment community, consisting not only of risk assessors but also of researchers collecting the critical data, need to carefully consider these questions in order to work toward shared understanding and consensus that is helpful for social decision-making. The critical needs are for (a) research design and development of mechanistic risk assessment models to be intelligently coordinated with each other and, (b) development of consensus criteria for the use of mechanistic data in risk assessment and risk management under different types of regulatory mandates. Workshops on the development of consensus criteria would be helpful in promoting reasoned discussion of these issues.

In a related proposal, "Extending the Range of Observation: Quantitative Relationships Between Key Biological Events to Aid Designation of Adversity and Identify Health Effects," Annie Jarabek emphasized the importance placed by the USEPA on characterizing the mode of action, defined as a chemical's influence on molecular, cellular or physiological functions (USEPA 1996; Jarabek 2000) in risk assessments. This requires a conceptual model that evaluates key events along the exposure-dose response continuum. Biomarkers data based in a mode-of-action framework essentially provide precursor lesion data and can serve as a basis for a parallelogram approach to extrapolation and determination of human homology for the health effect of interest (U.S. EPA 1994; Jarabek 1999). Thus, the framework provides for the extension of the range of observation, *e.g.*, for identification of biochemical or cellular events as measures of response, provided that causal links can be established to health outcome. She proposed research to evaluate the quantitative relationships among key events (*e.g.*, liver and cellular proliferation linked to tumor outcome) — from internal dose, to biologically effective dose, to various early effect indicators, to various outcome measures. This would provide a platform for the integration of diverse data, for example epidemiological data on effects in the population, and toxicological and mechanistic data acquired at the target tissue, cellular and subcellular levels. This work would also provide the necessary data to begin development of criteria for designation of adversity for use in risk assessment, *e.g.*, a specified degree of perturbation in cellular event such as 10% increase in cellular proliferation might be designated as a NOAEL or LOAEL. This work is important to accurately defining a given biomarker (defined by the NAS for exposure, effect, and susceptibility or combinations thereof) and distinguishing adaptive versus adverse effects.

MODELING ANALYSES FOR MULTIPLE ENDPOINT DATA

David Dunson proposed research to develop sophisticated statistical analytical tools to analyze multiple endpoint data, as is being generated in reproductive and neurodevelopmental studies. In recent years there has been increasing concern that exposure to chemicals with endocrine disrupting properties during development may have irreversible effects on reproductive, immune, and central nervous system function. This concern was formalized in the NAS report, *Pesticides in the Diets of Infants and Children*, which called for better information on the effects of pesticide exposure during development. In response to this report, testing designs have been implemented in which pregnant dams are dosed for the week before and after birth, and then the pups are dosed through puberty. Animals are tested at various points in the dosing period to ascertain effects on a variety of neurobehavioral, immunological, and reproductive outcomes.

Standard approaches for characterizing risk from toxicological studies are not ideal for multigenerational and developmental studies, in which multiple correlated endpoints are measured, as well as effects occurring across generations of related individuals. If the outcomes are considered separately and no adjustment is made for multiple comparisons, analyses will often detect some differences among dose groups (at, for example, $p = .05$) even if the chemical has no effect. However, standard adjustments for multiple comparisons make it very difficult to detect real effects, if present, because of the small numbers of animals tested and the large number of endpoints. An additional complication is that sick animals often die prior to being measured for outcomes that occur later in development. Such survival effects can produce biased estimates and misleading inferences. Another issue that arises in quantitative risk assessment is how to estimate a benchmark dose or virtually safe dose based on multiple correlated endpoints that are measured on different scales.

The purpose of this program would be to develop new approaches for assessing and characterizing risk in toxicology studies with multiple endpoints that are potentially measured on a variety of scales (*e.g.*, continuous, binary, ordinal). In particular, methods would be considered for reducing the dimensionality of the analysis, possibly by using a few "latent variables" underlying a set of measured outcomes (Dunson 2000). In addition, methods would be developed for estimating a dose level associated with a designated "acceptable" level of risk. One possibility would be to estimate the change with dose in the proportion of animals with a lower level of function than an average untreated animal (*e.g.*, by using a latent variable approach). The dose associated with a small change (*e.g.*, 1%) could be useful for policy makers deciding on permissible levels of exposure. A major objective would be to formulate a method that is readily interpretable by both toxicologists and risk managers.

Some important questions that need to be considered when developing this kind of method are

1. Should there be an adjustment for informative censoring from the deaths of sick animals before the end of the study?

2. Does the method have good operating characteristics in the small samples typical of these types of developmental toxicology studies?

3. Does the model account for correlation between different endpoints measuring a similar trait (*e.g.*, related neurobehavioral functions)?

4. Does the model account for correlation between endpoints measuring different traits (*e.g.*, unrelated neurobehavioral functions)?

5. Is the correlation structure realistic, but simple enough that the analysis yields information about all of the parameters, even in small samples?

6. Is there prior information that can be incorporated to improve the method? How can this information be combined with data from the current study?

7. Does the method account for different measurement scales for the different outcomes and for outcomes that do not follow standard distributions?

8. Can a latent variable or factor analytic type approach be used to simplify modeling and risk assessment (*e.g.*, by incorporating the dose effect relationship on one latent variable underlying the observed outcomes)?

9. If the latent variable approach is used, how many latent variable are there underlying the observed outcomes. How can the data be used to provide information about this choice?

10. Is the method robust (*e.g.*, to outliers, choice of parametric forms, etc.)

11. Is the model identified by the data? Can reasonable constraints be added to ensure identifiability and improve efficiency?

12. Is the model easy enough to implement and interpret for non-statisticians to be used widely?

NEW MECHANISTIC MODELS OF CARCINOGENESIS

Two major topics were discussed under this heading:

- The implications for risk assessment modeling of genomic instability and webs of molecular pathological pathways to carcinogenesis (instead of a completely sequential set of k ordered stages in which particular genetic changes are acquired in a fixed order)

- The need to develop procedures (*e.g.* Markov chain Monte Carlo analysis) to represent the available biological information on signaling pathways and cellular responses mediating carcinogenesis)

Chao Chen highlighted some of the research opportunities related to this area

It is now widely accepted that cancer results from the accumulation of mutations and other genetic changes in the genes that directly control cell division, cell death, and differentiation (Bishop 1987; Weinberg 1989; Sugimura 1992; Williams *et al.*

1996). These include positive changes in growth/division signaling systems via "protooncogenes" and abrogation of specific growth/division control functions in tumor suppresser genes (Barrett 1993). For some human cancer sites, molecular biological tools have been used to elucidate the specific sequence of alterations in genetically determined functions both in full-fledged tumors and in tissue where the process is not yet complete (Fearon and Vogelstein 1990; Shi 1999; Mao *et al.* 1997; Grossman and Lefell 1997; Cordon-Cardo *et al.* 1994; Sekido *et al.* 1998).

Given these recent advances in molecular research, and increasing understanding of the different kinds of influences (*e.g.*, "nongenetic" events such as cell proliferation) on carcinogenesis that can be exerted by chemical exposures, risk assessors face major challenges in adapting risk assessment models to accommodate new mechanistic understanding. For example TCDD, acting in part through Ah receptors, triggers a variety of biological responses that can be divided into two broad categories: (1) metabolic changes associated with uptake and subsequent binding with other proteins, and (2) mitogenic processes manifested as DNA replication, cell division, and alterations in differentiation. Which of these are relevant in what ways quantitatively for predicting both cancer risks and a variety of non-cancer risks? Schwarz *et al.* (2000) recently reviewed several proposed possible influences on signaling pathways in relation to TCDD-induced antiapoptotic activities and concluded that many of the proposed pathways are not plausible. It would be highly misleading if a putative causal pathway were chosen for modeling that turned out not to be relevant to the endpoints of ultimate interest. As a general approach a model could be constructed that incorporates rate-limiting factors that are known to be in a relevant signaling pathway, and a black box could be used to represent unknown or uncertain steps. This type of situation is best handled with a newly developed type of mathematical tool that can be easily used to construct a stochastic model that incorporates all available biological information from activation of signaling pathways to cellular responses and tumor incidence (Tan and Chen 1998). Application of this kind of tool to risk assessment problems is a promising area for further exploration.

COMBINING DATA OF DIFFERENT TYPES IN RISK ANALYSES

This was a recurring theme emphasized by several participants in the course of the discussions. The main themes were to encourage

- Development of hierarchical Bayesian frameworks for cancer risk assessment to:

  —combine data from multiple sources of a variety of types (*e.g.*, data from different animal bioassays covering animals of different cancer sites, genders and species, and human epidemiology)

  —integrate data on the effects of a single chemical via different modes of action

  —develop probabilistic distributions of hazard potential

- Application of a control theory framework to

  —understand the influences of population heterogeneity in metabolic pathway activities on susceptibility and risk.

EXPLORING EVIDENCE AND MODELS FOR COMPLEX DOSE RESPONSE RELATIONSHIPS IN THE CONTEXT OF HOMEOSTASIS

Both Kenny Crump and Ralph Kodell submitted proposals for research on the toxicological phenomenon called hormesis, whereby a substance causes deleterious effects at high doses but a stimulatory response in the opposite direction at low doses (Calabrese *et al.* 1999). Dr. Crump's proposal was to develop and apply rigorous statistical tools to evaluate the pervasiveness of hormesis, while Dr. Kodell's proposal was to develop biologically plausible mathematical models that predict less-than-background risk of cancer at low doses. Under Dr. Crump's proposal, a toxicological data base comprised of data sets meeting a set of minimal criteria would be examined rigorously, controlling the false positive error rate while making the power to detect hormetic effects as large as possible. Meta-analytic procedures would be used to estimate the prevalence of hormesis in the data base, such as the procedure used by Crump *et al.* (1999) to estimate proportions of carcinogenic and anticarcinogenic chemicals in bioassays conducted under the National Toxicology Program. Under Dr. Kodell's proposal, biological data would be collected and biomathematical models of hormesis would be developed, to build on hypotheses (Andersen and Conolly 1998; Lutz 1998) and models (Bogen 1998; Lutz and Kopp-Schneider 1999) that have been proposed to support U-shaped or J-shaped dose response relationships for cancer. If, for example, toxic substances can affect homeostatic processes in such a way as to produce dose response relationships exhibiting less-than-background risk at low doses, then the default assumption that either genotoxicity or additivity to background will necessarily lead to low-dose linearity needs to be re-evaluated. Some others in the group are skeptical that a careful and rigorous analysis would lead to this result (Hattis 1997).

OVERALL CONCLUSION

The suggestions for research described above indicate the potential for greatly improved contributions from different fields (toxicology, molecular epidemiology, mathematical modeling, etc.) to improve dose response modeling. This can in turn improve the estimation of potential health risks and the consequences of different risk control choices. The field is the focal point of the dynamic interaction of data and theory relevant to important social policy concerns.

REFERENCES

Allen BC, Covington TR, and Clewell HJ. 1996. Investigation of the impact of pharmacokinetic variability and uncertainty on risks predicted with a pharmacokinetic model for chloroform. Toxicology 111:289-303

Andersen ME and Conolly RB. 1998. Mechanistic modeling of rodent liver tumor promotion at low levels of exposure: An example related to dose response relationships for 2,3,7,8-

tetrachlorodibenzo-p-dioxin. Human Experimental Toxicol 17, 683-90; discussion 701-4, 708-18

Bailer AJ and Dankovic DA. 1997. An introduction to the use of physiologically-based pharmacokinetc models in risk assessment. Stat Methods Medical Res 6:341-58

Bailer AJ, Stayner LT, Smith RJ, *et al.* 1997. Estimating benchmark concentrations and other non-cancer endpoints in epidemiology studies Risk Anal 17:771-80

Baird JS, Cohen JT, Graham JD, *et al.* 1996. Noncancer risk assessment: Probabilistic characterization of population threshold doses. Human Ecol Risk Assess 2(1):79-102

Barrett JC. 1993. Mechanisms of multistep carcinogenesis and carcinogen risk assessment. Environ Health Perspect 100:9-20

Bishop JM. 1987. The molecular genetics of cancer. Science 235:305-11

Bogdanffy MS and Jarabek AM 1995. Understanding mechanisms of inhaled toxicants: Implications for replacing default factors with chemical-specific data. Toxicol Lett 82/83:919-32

Bogdanffy MS, Sarangapani R, Plowchalk DR, *et al.* 1999. A biologically-based risk assessment for vinyl-acetate induced cancer and noncancer inhalation toxicity. Toxicol Sci 51:19-35

Bogen KT. 1998. Mechanistic model predicts a U-shaped relation of radon exposure to lung cancer risk reflected in combined occupational and US residential data. Human Experimental Toxicol 17:691-6

Bogen KT and Spear RC. 1987. Integrating uncertainty and interindividual variability in environmental risk assessment. Risk Anal 7:427-36

Bois FY. 1999. Analysis of PBPK models for risk characterization. Ann NY Acad Sci 895:317-37

Bois FY, Gelman A, Jiang J, *et al.* 1996. Population toxicokinetics of tetrachloroethylene. Arch Toxicol 70(6):347-55

Bois FY, Krowech G, and Zeise L. 1995. Modeling human interindividual variability in metabolism and risk: the example of 4-aminobiphenyl. Risk Anal 15:205-13

Bois FY, Zeise L, and Tozer TN. 1990. Precision and sensitivity of pharmacokinetic models for cancer risk assessment: Tetrachloroethylene in mice, rats, and humans. Toxicol Applied Pharmacol 102:300-15

Calabrese EJ, Baldwin LA, and Holland CD. 1999. Hormesis: a highly generalizable and reproducible phenomenon with important implications for risk assessment. Risk Anal 19:261-81

California Environmental Protection Agency. 2000. Draft Public Health Goal for Tetrachloroethylene in Drinking Water. Office of Environmental Health Hazard Assessment, Sacramento, CA, USA

Chemical Industry Institute of Toxicology (CIIT). September 28, 1999. Formaldehyde: Hazard Characterization and Dose Response Assessment for Carcinogenicity by the Route of Inhalation. Revised Edition. Chemical Industry Institute of Toxicology, Research Triangle Park, NC, USA

Chien JY, Thummel KE, and Slattery JT. 1997. Pharmacokinetic consequences of induction of CYP2E1 by ligand stabilization. Drug Metab Dispos 25:1165-75

Clewell HJ. 1995. The use of physiologically based pharmacokinetic modeling in risk assessment: A case study with methylene chloride. In: Olin S, Farland W, Park C, *et al.* (eds), Low-Dose Extrapolation of Cancer Risks: Issues and Perspectives. ILSI Press, Washington, DC, USA

Clewell HJ, Gearhart JM, Gentry PR, *et al.* 1999. Evaluation of the uncertainty in an oral Reference Dose for methylmercury due to interindividual variability in pharmacokinetics. Risk Anal 19:541-52

Clewell HJ and Jarnot BM. 1994. Incorporation of pharmacokinetics in non-carcinogenic risk assessment: Example with chloropentafluorobenzene. Risk Anal 14:265-76

Clewell HJ, III and Andersen ME. 1996. Use of physiologically-based pharmacokinetic modeling to investigate individual versus population risk. Toxicology 111:315-29

Cordon-Cardo C, Dalbagni G, Sarkis AS, *et al.* 1994. Genetic alterations associated with bladder cancer. Important Advances in Oncology 1994:71-83

Crump KS, Hoel DG, Langley CH, *et al.* 1976. Fundamental carcinogenic processes and their implication for low dose risk assessment. Cancer Res 36:2973-9

Crump KS, Krewski D, and VanLandingham C. 1999. Estimates of the proportion of chemicals that were carcinogenic or anticarcinogenic in bioassays conducted by the National Toxicology Program. Environ Health Perspect 107:83-8

Cullen AC and Frey HC. 1999. Probabilistic Techniques in Exposure Assessment—A Handblook for Dealing with Variability and Uncertainty in Models and Inputs. Plenum Press, NY, NY, USA

Dankovic D and Bailer AJ. 1994. The impact of exercise and intersubject variability on dose estimates for dichloromethane derived from a physiologically-based pharmacokinetic model. Fund Applied Toxicol 22:20-5

Dietert RR, Etzel RA, Chen D, *et al.* 2000. Workshop to identify critical windows of exposure to children's health: Immune and respiratory systems work group summary. Environ Health Perspect 108(suppl. 3):483-90

Dourson M, Maier A, Meek B, *et al.* 1998. Boron tolerable intake: Re-evaluation of toxicokinetics for data-derived uncertainty factors. Biol Trace Elem Res 66:453-63

Dunson DB. 2000. Bayesian latent variable models for clustered mixed outcomes. J R Stat Soc B 62 (part 2):355-66

Fearon ER and Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. Cell 61:759-67

Finkel AM. 1987. Uncertainty, Variability and the Value of Information in Cancer Risk Assessmentt. Ph.D. Dissertation. Harvard University, Harvard School of Public Health, Cambridge, MA, USA

Finkel AM. 1995. A quantitative estimate of the variations in human susceptibility to cancer and its implications for risk management. In: Olin S, *et al.* (eds), Low-Dose Extrapolation of Cancer Risks: Issues and Perspectives, pp 297-328. ILSI Press, Washington, DC, USA

Fosbroke DE, Kisner SM, and Myers JR. 1997. Working lifetime risk of occupational fatal injury. Am J Ind Med 31:459-67

Gilbert SJ, Bailer AJ, Stayner LT. 1998. Years of Potential Life Lost due to Occupational Fatal Injury in the United States. Human Ecol Risk Assess 4(6):1321-35

Grossman D and Leffell DJ. 1997. The molecular basis of nonmelanoma skin cancer: new understanding. Archives Dermatol 133(10):1263-70

Hasselblad V and Jarabek AM. 1996. Dose-response analysis of toxic chemicals. In: Berry DA and Stangle DK (eds), Bayesian Biostatistics. Marcel Dekker, Inc., NY, NY, USA

Hattis D. 1996. Human interindividual variability in susceptibility to toxic effects–from annoying detail to a central determinant of risk. Toxicology 111:5-14

Hattis D. 1997. Invited comment on Heitzman, M., and Wilson, R. low dose linearity: The rule or the exception. Belle Newsletter 6:21-4

Hattis D. 1998. Strategies for assessing human variability in susceptibility, and variability to infer human risks. In: Neumann DA and Kimmel CA (eds) Human Variability in Response to Chemical Exposure: Measures, Modeling, and Risk Assessment, pp 27-57. CRC Press, Boca Raton, FL, USA

Hattis D and Anderson E. 1999. What should be the implications of uncertainty, variability, and inherent 'biases'/'conservatism' for risk management decision making? Risk Anal 19:95-107

Hattis D, Banati P, Goble R, *et al.* 1999. Human interindividual variability in parameters related to health risks. Risk Anal 19:705-20

Hattis D and Barlow K. 1996. Human Interindividual Variability in Cancer Risks—Technical And Management Challenge. Health Ecol Risk Assess 2:194-220

Hattis D and Burmaster DE. 1994. Assessment of variability and uncertainty distributions for practical risk analyses. Risk Anal 14:713-30

Hattis D and Minkowitz WS. 1996. Risk evaluation: criteria arising from legal traditions and experience with quantitative risk assessment in the United States. Environmental Toxicol Pharmacol 2:103-9

Hattis D and Silver K. 1994. Human interindividual variability—a major source of uncertainty in assessing risks for non-cancer health effects. Risk Anal 14:421-31

Integrated Risk Information System (IRIS). October 1999. Toxicological review of ethylene glycol monobutyl ether (EGBE) CAS no. 111-76-2. National Center for Environmental Assessment, U.S. Environmental Protection Agency, Washington, DC, USA

Integrated Risk Information System (IRIS). May 2000. Toxicological review of vinyl chloride CAS no. 75-01–4. National Center for Environmental Assessment, U.S. Environmental Protection Agency, Washington, DC, USA

Internation Programme for Chemical Safety (IPCS). 1998. Boron. Environmental Health Criteria. 204. World Health Organization, Geneva, Switzerland

Jarabek AM. 1995a. Interspecies extrapolation based on mechanistic determinants of chemical disposition. Human Ecol Risk Assess 1(5):641-62

Jarabek AM. 1995b. The application of dosimetry models to identify key processes and parameters for default dose response assessment approaches. Toxicol Lett 79:171-84

Jarabek AM. 1999. Application requirements for biomarkers in risk assessment. Presented at Biomarkers: Taking stock – An EPA/NIEHS In-house Workshop on applying biomarker research, 30-31 August 1999, Research Triangle Park, NC, USA (Manuscript in preparation)

Jarabek AM. 2000. Mode of Action: Framework for dosimetry model development. Presented in: Mode-of-Action Dosimetry: An interagency project to develop models for inhalation, oral and dermal disposition. Annual meeting of the Society for Risk Analysis, 4 – 6 December, Arlington, VA, USA (Manuscript in preparation)

Jonsson F and Johanson G. 2000. Improving the reliability of parameters in PBTK models using Markov Chain Monte Carlo simulation. (Abstract No. 422) Toxicologist 54:90

Kohn MC. 1995. Achieving credibility in risk assessment models. Toxicol Letters 79:107-14

Kohn MC. 1997. The importance of anatomical realism for validation of physiological models of disposition of inhaled toxicants. Toxicol Appl Pharmacol 147:448-58

Krewski D, Rai SN, Zielinski JM, *et al.* 1999. Characterization of uncertainty and variability in residential radon cancer risks. Annals NY Academy Sciences 895:245-72

Leisenring W and Ryan L. 1992. Statistical properties of the NOAEL. Reg Toxicol Pharmacol 15(2):161-71

Lutz WK. 1998. Dose response relationships in chemical carcinogenesis: superposition of different mechanisms of action, resulting in linear-nonlinear curves, practical thresholds, J-shapes. Mutation Res 405:117-24

Lutz WK and Kopp-Schneider. 1999. Threshold dose response for tumor induction by genotoxic carcinogens modeled via cell-cycle delay. Toxicol Sciences 49:110-5

Manton KG and Stallard E. 1979. Maximum likelihood estimation of a stochastic compartment model of cancer latency: lung cancer mortality among white females in the U.S. Computers and Biomedical Res 12:313-25

Manton KG, Stallard E, and Vaupel J. 1986. Alternative models for the heterogeneity of mortality risks among the aged. J Am Stat Assoc 81:635-44

Mao L, Lee JS, Kurie JM, *et al.* 1997. Clonal genetic alterations in the lungs of current and former smokers. J National Cancer Institute 89(12):857-62

Meek ME, Ohanian E, Renwick A, *et al.* 2000. Guidelines for application of data-derived uncertainty factors in risk assessment. Submitted.

National Research Council. 2000. Acute Exposure Guideline Levels for Selected Airborne Chemicals, Volume 1. National Academy Press, Washington, DC, USA

National Research Council. 2001. Standing Operating Procedures for Developing Acute Exposure Guideline Levels for Hazardous Substances. National Academy Press, Washington, DC (*in press*).

Occupational Safety and Health Administration (OSHA). 1997. Occupational exposure to methylene chloride; final rule. Fed Reg 62(7):1493-619

Price PS, Keenan RE, Swartout JC, *et al.* 1997. An approach for modeling noncancer dose responses with an emphasis on uncertainty. Risk Anal 17(4):427-37

Renwick AG. 1991. Safety factors and establishment of acceptable daily intakes. Food Add Contam 8(2):135-50

Renwick AG and Lazarus NR. 1998. Human variability and noncancer risk assessment—an analysis of the default uncertainty factor. Regul Toxicol Pharmacol 27(1 Pt 1):3-20

Schlosser PM and Bogdanffy MS. 1999. Determining modes of action for biologically based risk assessments. Regul Toxicol Pharmacol 30:75-9

Schwarz M, Buchmann A, Stinchcombe S, *et al.* 2000. Ah receptor ligands and tumor promotion: survival of neoplastic cells. Toxicol Letters 112-113:69-77

See K and Bailer AJ. 1998. Estimates of lifetime risk of occupational fatal injury from age-specific rates. Human Ecol Risk Assess 4(6):1309-19

Sekido Y, Fong KM, and Minna JD. 1998. Progress in understanding the molecular pathogenesis of human lung cancer. Biochimica et Biophysica Acta 1378:F21-59

Shi ST, Yang FY, Wang LD, *et al.* 1999. Role of p53 gene mutations in human esophageal carcinogenesis; results from immunohistochemical and mutation analyses of carcinomas and nearby non-cancerous lesions. Carcinogenensis 20:591-7

Slob W and Pieters MN. 1997. A Probabilistic Approach for Deriving Acceptable Human Intake Limits and Human Health Risks from Toxicological Studies: General Framework. 620110005. Rijksinstituut voor Volksgezondheid en Milieu, National Institute of Public Health and the Environment, The Netherlands

Sugimura T. 1992. Multistep carcinogenesis: A 1992 perspective. Science 258:603-8

Swartout JC, Price PS, Doruson ML, *et al.* 1998. A probabilistic framework for the reference dose. Risk Anal 18(3):271-82

Tan WY and Chen C. 1998. Stochastic modeling of carcinogenesis: Some new insights. Mathematical Computer Modeling 28(11):49-71

Thompson KM. 1999. Developing univariate distributions from data for risk analysis. Human and Ecol Risk Assess 5:755-83

USEPA (U.S. Environmental Protection Agency). October1994. Methods for Derivation of Inhalation Reference Concentrations and Application of Inhalation Dosimetry. EPA-600/8-90-066F. Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, Research Triangle Park, NC, USA

USEPA (U.S. Environmental Protection Agency). 1996a. Air Quality Criteria for Particulate Matter – Volume II. EPA-600/P-95-001bF. National Center for Environmental Assessment, Research Triangle Park, NC. Available from NTIS, Springfield, VA, USA; PB96-168224.

USEPA (U.S. Environmental Protection Agency). 1996b. Proposed guidelines for carcinogen risk assessment. Federal Register 61(79):17960-8011

USEPA (U.S. Environmental Protection Agency). 2000. Exposure and Human Health Reassessment of 2,3,7,8-Tetrachlorodibenzo-p-Dioxin (TCDD) and Related Compounds. Part II: Health Assessment for 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) and Related Compounds. Chapter 8. Dose response Modeling for 2,3,7,8-TCDD. NCEA-I-0835, May 2000, SAB Review Draft.

Wassell JT. 1998. Probability Models of Occupational Injury. Human Ecol Risk Assess 4(6):1275-83

Wienberg RA. 1989. Ocogenes, antioncogenes, and the molecular bases of multistep carcinogenesis. Cancer Res 49:3713-21

Williams JR, Russell J, Dicello JF, *et al.* 1996. The genotype of the human cancer cell: Implications for risk analysis. Mutation Res 365:17-42

Human and Ecological Risk Assessment: Vol. 8, No. 6, pp. 1445-1487 (2002)

Review of Information Resources to Support Human Exposure Assessment Models

Catherine Petito Boyce[1] and Michael R. Garry[2]
[1]Gradient Corporation, 9725 SE 36th Street, Suite 404, Mercer Island, WA 98040;
[2]Exponent, Inc., 15375 SE 30th Place, Suite 250, Bellevue, WA 98007

ABSTRACT

Efforts to model human exposures to chemicals are growing more sophisticated and encompass increasingly complex exposure scenarios. The scope of such analyses has increased, growing from assessments of single exposure pathways to complex evaluations of aggregate or cumulative chemical exposures occurring within a variety of settings and scenarios. In addition, quantitative modeling techniques have evolved from simple deterministic analyses using single point estimates for each necessary input parameter to more detailed probabilistic analyses that can accommodate distributions of input parameters and assessment results. As part of an overall effort to guide development of a comprehensive framework for modeling human exposures to chemicals, available information resources needed to derive input parameters for human exposure assessment models were compiled and critically reviewed. Ongoing research in the area of exposure assessment parameters was also identified. The results of these efforts are summarized and other relevant information that will be needed to apply the available data in a comprehensive exposure model is discussed. Critical data gaps in the available information are also identified. Exposure assessment modeling and associated research would benefit from the collection of additional data as well as by enhancing the accessibility of existing and evolving information resources.

Key Words: exposure assessment, exposure modeling, input parameters, information resources, probabilistic, review.

BACKGROUND

Efforts to model human exposures to chemicals and consequent health risks are growing more sophisticated and encompass increasingly complex exposure sce-

[1] Corresponding author: cpboyce@gradientcorp.com, Tel(voice): 206-275-4774, Tel(fax): 206-275-4775
Received December 1, 2001; revised manuscript accepted February 6, 2002

1080-7039/02/$.50

narios. The applications and goals of quantitative exposure models have grown in complexity. In addition, the scope of such analyses has increased, growing from assessments of single exposure pathways to complex evaluations of aggregate or cumulative chemical exposures occurring within a variety of settings and scenarios. For example, as a result of passage of the Food Quality Protection Act (FQPA) in 1996, a number of models have been developed to assess aggregate exposures to pesticides through use on food crops and from home applications and other exposure sources (see, *e.g.*, Price *et al.* 2000, 2001; Shurdut *et al.* 1998; ILSI 1998). Models have also been used to examine cumulative exposures to multiple chemicals that may have a similar mode of action (see, *e.g.*, USEPA 1997c).

The passage of this act, as well as the current or upcoming public availability of substantial amounts of screening level toxicity and exposure data, has raised interest in developing methods for conducting comprehensive exposure assessments for non-pesticide chemicals as well. Initiatives in these areas include the Toxics Release Inventory and associated efforts to evaluate the potential exposure posed by the listed emissions (*e.g.*, Hertwich *et al.* 1999; Scorecard 2000), the toxicity data that will be generated as part of the High Production Volume chemical testing program and associated efforts to provide some exposure-related context (*e.g.*, ACA 2000), the Children's Chemical Evaluation Program (BNA 2000), the National Exposure Report Card summarizing exposures of the U.S. population to a variety of chemicals based on biomonitoring data collected by the Centers for Disease Control and Prevention (CDC) (CDC 2001b; Fox 2000), and environmental justice evaluations (*e.g.*, Waller *et al.* 1999; Sexton and Adgate 1999).

These changes in the scope of quantitative exposure models have been accompanied by enhancements in the tools used to model exposures. Quantitative modeling techniques have evolved from simple deterministic analyses that use single point estimates for each input factor to more detailed probabilistic analyses that can accommodate distributions of input factors and assessment results. Use of probabilistic techniques, such as Monte Carlo analyses, in conducting exposure and risk analyses has become more prevalent, as has the sophistication of such analyses. By using distributions of values for various input factors, rather than selecting single values as is done in deterministic analyses, probabilistic techniques can more easily incorporate a broader range of the available data needed for exposure and risk analyses. In addition, by generating a distribution of exposure or risk estimates, probabilistic analyses provide a more direct means of quantitatively assessing uncertainties and the degree to which the results of the analyses are applicable to specific segments of the potentially exposed populations (*e.g.*, Whitmyre *et al.* 1992a). Growth in the use of probabilistic techniques has been driven in part by the interests of risk assessment practitioners in developing such techniques (see *e.g.*, Cullen and Frey 1999). The U.S. Environmental Protection Agency (USEPA) and other regulatory agencies have also encouraged the use of such techniques through guidance, policy statements, and other efforts supporting development of such techniques (*e.g.*, USEPA 1992c, 1995b, 1996c, 1997b, 1999d; ODEQ 1998).

To date, such efforts have focused primarily on developing more refined estimates of exposure, rather than incorporating distributions for assessing toxicity. Although some techniques for developing distributions of toxicity factors have been applied, and other potential methods exist (see, *e.g.*, Petito Boyce 1998; Baird *et al.*

1996), various elements of exposure assessment have proven more amenable to distribution development. Thus, exposure assessment has been the primary focus of probabilistic applications in risk assessment.

Paralleling the growth in application of probabilistic techniques is increasing interest in defining technically sound distributions of values for use in such analyses and in understanding the types of information reflected in such distributions. Some distributions are designed to reflect actual interindividual variability in the potential values for specific input parameters (*e.g.*, natural variations in body weight among individuals). Intraindividual variability is also being increasingly recognized, *e.g.*, an individual's fish ingestion rate may vary over time, and at different stages in life (*e.g.*, Harrington *et al.* 1995; Price *et al.* 1996). Some distributions also reflect uncertainty regarding the true value of an input parameter, or some combination of uncertainty and variability. Probabilistic approaches and interpretation, including the use of two-dimensional techniques that explicitly segregate sources of variation in results, can distinguish among these influences on the results of the analyses.

PROJECT APPROACH

As one component of an overall effort to develop a comprehensive framework for modeling human exposures to chemicals, available information resources needed to derive input factors for human exposure assessment models were identified and compiled, including existing default exposure factors. The factors evaluated in this study were grouped into the following five categories: individual physical and physiological factors, intake rates and related factors, behavioral factors related to activity patterns, demographic factors, and environmental modeling factors. This information was critically reviewed, including evaluating the adequacy of the available data to assess uncertainty and variability and assessing the degree to which available default values are representative of the overall data sets. Ongoing research in the area of exposure factors was also identified. These efforts focused on exposure factors that would be of interest for non-pesticide chemicals. The results of these efforts are summarized, and other relevant information that will be needed to apply the available data in a comprehensive exposure model is discussed. Critical gaps in the available information are also identified.

This work focused primarily on the most relevant data for assessing exposures to individuals in the United States, on exposures to the general population, and on parameters of general applicability (rather than parameters that are chemical-specific or situation-specific). Some elements of this focus reflect, in part, limitations in currently available information. For example, although an extensive amount of information is available regarding occupational exposures, these data historically have focused primarily on monitoring measurements of specific workplace chemicals and on inhalation exposures. Frequently, supplemental information is lacking to more specifically characterize the nature of the exposures or to support evaluations of other exposure pathways (*e.g.*, the activities workers were engaged in when the monitoring occurred, the duration of the exposures, or the presence of protec-

[1] This work was conducted with support from the American Chemistry Council (ACC). Copies of the full report upon which this article is based are available through the authors.

tive equipment that would modify exposures). As reflected in the proceedings of the November 1999 *Occupational Exposure Database Symposium* and reports in the February 2001 issue of *Applied Occupational and Environmental Hygiene*, efforts are underway both in the U.S. and abroad to develop more consistent, complete, and readily accessible databases of occupational exposure information to support health and safety monitoring and research; however, such information is not broadly available at this time (see, *e.g.*, ACGIH 2001; Morgan 2001; Boiano and Hull 2001; Van Dyke *et al.* 2001; Marquart *et al.* 2001; and Abell *et al.* 2001).

This research effort began with a review of compilations of those factors for which default parameters are available (*e.g.*, the USEPA *Exposure Factors Handbook* 1997a). Additional information resources were identified through searches of peer-reviewed scientific journals and gray literature, Internet searches of information regarding exposure assessment modeling, and research being conducted by regulatory agencies and other entities, and a call for information regarding unpublished and ongoing studies issued through a relevant Internet list server. In addition, approximately 45 exposure assessment researchers and other practitioners in the academic, government, trade association, and private sectors were directly contacted to identify ongoing or planned research that is relevant to exposure assessment and model development. These individuals were also polled regarding their perceptions of important gaps in the available information to supplement the data gap review conducted in this project.

OVERVIEW OF RESEARCH RESULTS

Resources Reviewed

Relevant information was identified from a number of sources. First, numerous secondary sources exist that compile a wealth of exposure-related information. Primary among these are USEPA's *Exposure Factors Handbook* (USEPA 1997a, 2001d); USEPA's draft *Child-Specific Exposure Factors Handbook* (USEPA 2000a), and the American Industrial Health Council's (AIHC) *Exposure Factors Sourcebook* (AIHC 1994). These sources are particularly useful because they were specifically designed to support exposure assessment. Guidance prepared by various regulatory agencies (*e.g.*, ODEQ 1998) and articles in the scientific literature (*e.g.*, Finley *et al.* 1994a; Gephart *et al.* 1994; Paustenbach 2000) also provide useful compilations of available information, in some cases including recommended point estimates or distributions of values for selected exposure parameters. The Society for Risk Analysis has also sponsored a textbook compiling information relevant for assessing exposures in residential settings (Baker *et al.* 2001; Driver 2001). The USEPA has also issued guidance on methods for deriving probability distributions based on the information available in the *Exposure Factors Handbook* (USEPA 2000l). Other resource compilations include a listing of federally sponsored databases that have been, or could be, used to support exposure assessment that was prepared in the early 1990s (Sexton *et al.* 1994).

In addition to these sources, which focus on U.S. populations, efforts have also been undertaken to compile information for conducting exposure assessments for European populations, *i.e.*; the *Exposure Factors Sourcebook for European Populations, With Focus on UK Data* (ExxonMobil 2000). This document includes a listing of

additional data sources that could be researched for exposure factor information tailored to specific regions of Europe. The European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) has also compiled exposure assessment resources focusing on European populations. These efforts include a software program (HAZCOM) to assess indirect exposures to chemicals and guidance regarding conducting exposure assessments for a wide variety of consumer products (ECETOC 1994). The Netherlands National Institute for Public Health and Environmental Protection has also developed a general framework for estimating human exposures to chemicals associated with consumer products (van Veen 1996) and has evaluated probabilistic approaches for assessing consumer exposures (van Veen *et al.* 2001). An initial evaluation of data needs for conducting probabilistic exposure assessments for German populations has also been prepared (Mekel and Fehr 2001). As necessary, information presented in these U.S. and other sources regarding specific factors was updated by using recent articles obtained from the scientific literature and other information sources.

Comprehensive exposure assessment studies and exposure modeling software packages also include relevant information for determining exposure parameters of interest in this study. For example, USEPA's National Human Exposure Assessment Survey (NHEXAS) is a comprehensive multi-site study of exposures to metals, pesticides, volatile organic chemicals, and polycyclic aromatic hydrocarbons that includes monitoring and biomarker data combined in some cases with activity pattern data from study participants (*e.g.*, Echols *et al.* 1999; Freeman *et al.* 1999; Gordon *et al.* 1999; MacIntosh *et al.* 2001; Pellizzari *et al.* 1995; Robertson *et al.* 1999; Scanlon *et al.* 1999; Sexton *et al.* 1995a,b; USEPA 2000j,k). The USEPA is also a collaborating agency in the National Health and Nutrition Examination Survey (NHANES), a comprehensive health study that is conducted by the National Center for Health Statistics, part of the Centers for Disease Control and Prevention (CDC). Conducted periodically since 1971, data collected through NHANES include descriptive and demographic information (*e.g.*, height, weight, race, and socioeconomic status), information regarding various behavioral practices that might influence exposure (*e.g.*, smoking, physical activity, environmental exposures, and eating habits), and biochemical measurements of health status or environmental exposures to specific agents (CDC 2001a; Perlin 2000).

The California Air Resources Board (CARB) has also conducted extensive studies of activity patterns in California populations (*e.g.*, breathing rates associated with various activities and driving patterns) (CARB 2000a), as well as other studies relevant to assessing human exposures (*e.g.*, monitoring studies focused on specific chemicals and emissions sources). Some of these resources are summarized in CARB's *Technical Support Document for Exposure Assessment and Stochastic Analysis* (CARB 2000b). A number of other recently completed and ongoing surveys by the USEPA (*e.g.*, surveys of children's exposures and factors influencing their exposures; Cohen Hubal *et al.* 2000a, USEPA 2000m), the U.S. Department of Energy (*e.g.*, statistical data on residential and commercial building characteristics, USDOE 1995), and the U.S. Department of Agriculture (*e.g.*, market basket or food consumption surveys, USDA 2000) also provide relevant information.

Two large database programs developed by USEPA's National Exposure Research Laboratory (NERL) provide access to much of the raw data from some of

the studies reviewed in this report, including data from the National Human Activity Pattern Survey (NHAPS; USEPA 1996a) and from studies conducted by CARB (Wiley *et al.* 1991; Clayton and Perritt 1993). The Total Human Exposure Risk Database and Advanced Simulation Environment (THERdbASE; USEPA 2000b) is an integrated database and modeling software system including data on human activity patterns, U.S. Census data, and other data relevant to modeling human exposure. The USEPA is currently developing a new database and modeling system (the Human Exposure Assessment Database System or HEADS) that will replace THERdbASE (Engelmann 2001). This new system is designed to be more flexible and to take advantage of capabilities offered by Internet access (*e.g.*, allowing users to more readily download or upload data into HEADS). Industry is also developing similar systems for coordinating exposure data and models, *e.g.*, the Cumulative and Aggregate Risk Evaluation System or CARES (Driver 2001).

In addition to the NHAPS, California, and U.S. Census data, NERL's Consolidated Human Activities Database (CHAD) includes activity pattern and other data from studies conducted in Baltimore, Washington, DC, Cincinnati, and Denver. The consolidated database and the raw data from individual studies can be downloaded or accessed online (McCurdy *et al.* 2000; USEPA 2000d). Development of this database is ongoing. CHAD also includes data from the University of Michigan's Child Development Supplement studies (University of Michigan 2000). These studies include activity pattern and time use data from approximately 2500 children and their caregivers. The raw data and the data collection tools are downloadable from the Institute for Social Research website (UM 2000).

USEPA's Office of Pollution Prevention and Toxics (OPPT) has also developed several specialized exposure databases and software tools capable of evaluating exposure to chemicals in the environment and/or in consumer products (USEPA 2000c). These tools typically include physiological, intake, activity pattern, demographic, and environmental modeling data from some of the studies summarized in this report. Four applications developed by OPPT assess consumer product exposure in some form, *i.e.*, the Multi-Chamber Concentration and Exposure Model (MCCEM), the Wall Paint Exposure Assessment Model (WPEM), the Exposure-Fate Assessment Screening Tool (E-FAST), and the Source Ranking Database (SRD). MCCEM and E-FAST are currently available to download from OPPT's website, and WPEM and SRD can be obtained from the USEPA (USEPA 2000c). USEPA's RISK software can calculate exposure to and risks from indoor air pollutants and allows for consideration of the effects of room-to-room airflows, air exchange with the outdoors, and air cleaners on the concentration-time history of pollutants (USEPA 2000c). Another relevant USEPA exposure model that draws upon some of the data sources reviewed in this report is the Dietary Exposure and Evaluation Model (DEEM), which can combine dietary intake data with data regarding chemical residues in food or drinking water to provide estimates of chemical intake via food ingestion (Tomerlin *et al.* 1997; Novigen 2000; USEPA 2000i). The USEPA is also providing integrated access to a wide variety of databases, models, and other resources at its Environmental Information Management System (EIMS) web site (USEPA 2000e).

Review of Available Exposure Data

The available data for each of the five categories of exposure factors considered in this review were compiled and critically reviewed. The first step in this process included identifying whether chemical-specific or standard default values are available for these factors and the general scope of the information available for each type of factor. Table 1 summarizes the general nature of the available information for each identified factor. This table reflects the primary objectives of this project, *i.e.*, to identify information resources, to provide an overview of the nature and extent of the available information, and to provide references to more detailed information sources.

The second step of the data review process included considering the adequacy of the available data for each exposure factor to support the development of a distribution of values reflecting uncertainty, interindividual variability, intraindividual variability, and various combinations of these elements. Conclusions regarding the overall adequacy of the data for each type of exposure factor are summarized in Table 2 using qualitative ratings of data adequacy (*i.e.*, high, medium, and low adequacy). In developing these ratings, published ratings of data quality (*e.g.*, Gephart *et al.* 1994; USEPA 1997a, 2000a) and the perspectives of experts consulted during this project were considered with the findings of this review. The data adequacy ratings were also assigned considering the adequacy of the data relative to that available for other parameters. Specifically, although envisioning additional data needs or refinements is almost always possible, the adequacy of an available data set was assigned a high rating if it was judged to have high adequacy relative to the data available for other factors.

During this stage of the review, information regarding potential correlations among factors was also identified. Certain correlations (*e.g.*, between body weight and skin surface area) are so well characterized that joint distributions have been established (*e.g.*, Burmaster and Murray 1998). Many other correlations have been qualitatively identified but not extensively quantified (*e.g.*, the correlations between body weight and water intake or food intake). Other correlations such as the autocorrelation among an individual's activities or intake rates over time are recognized (*e.g.*, McCurdy 2000a) but not well understood. The quantitative implications of correlations have also been examined in hypothetical modeling studies, most notably Smith *et al.* (1992) and Bukowski *et al.* (1995). In general, these studies have concluded that the quantitative impacts associated with combining factors from two correlated distributions are relatively small. However, the effects of assuming different distribution shapes can yield greater differences in the model results.

The third step of the review process summarized the quantitative information available regarding the exposure factors, including available default values and potential distributions of values. The amount of data available for commonly used exposure factors is far from consistent. Some exposure factors have been extensively studied, resulting in substantial available data. For example, a number of studies of fish consumption rates have been conducted and yielded data that allow exposure assessors to distinguish among various sizes of water body where fishing may occur, different categories of consumers (*e.g.*, subsistence fishermen), or different categories of fish (*e.g.*, fin fish versus shellfish). In some cases, sufficient data are available

Table 1. Summary of available exposure data.

Exposure Factor	Available Information	Basis	Selected Sources
Individual Physical/Physiological Factors			
Body weight	Defaults, means, percentiles, and distributions.	Large, national databases.	AIHC (1994); Burmaster and Crouch (1997); Burmaster and Murray (1998); ExxonMobil (2000); Finley *et al.* (1994a); Gephart *et al.* (1994); Leighton (2001); ODEQ (1998); USEPA (1985, 1989a,1989b, 1991b, 1997a, 2000a, 2001b, 2001c).
Skin surface area	Defaults, predictive equations, means, percentiles, and distributions.	Primarily one measurement study. Assumptions regarding the amount of exposed skin surface area associated with various activities are typically based on professional judgment.	AIHC (1994); Burmaster (1998a); ExxonMobil (2000); Finley *et al.* (1994a); Gephart *et al.* (1994); Leighton (2001); USEPA (1985, 1989a, 1992a, 1997a, 2000a, 2001c).
Life expectancy	Average values for expectation of life.	Census data.	ExxonMobil (2000); U.S. Bureau of the Census (1995); USEPA (1991b, 1997a, 2000a).
Gastrointestinal absorption	Varies by chemical.	*In vitro* and *in vivo* studies, evidence from toxicological and epidemiological studies or human observations.	Multiple, including Owen (1990); USEPA (1989c, 1992b).
Absorption via inhalation	Varies by chemical.	*In vitro* and *in vivo* studies, evidence from toxicological and epidemiological studies or human observations.	Multiple, including Owen (1990); USEPA (1989c, 1992b).
Dermal absorption	Varies by exposure scenario and chemical.	Default diffusion assumptions, laboratory and field studies.	Multiple, including Durkin *et al.* (1995); USEPA (1989c, 1992a, 1992b, 1995a, 1997b).
Factors influencing internal dose, including generic physiological factors and chemical-specific factors	General data regarding dimensions, composition, and distribution of various anatomical tissues and reference information for various physiological processes in humans. Chemical-specific data regarding human variability in contact rates, uptake or absorption, systemic dilution, elimination, half-lives, active site availability, and functional reserve capacity. Available data vary by chemical and parameter.	Anatomical and physiological studies; studies from pharmaceutical and other scientific literature reflecting human variability in response to specific substances.	Multiple, including Brown *et al.* (1997); Burin and Saunders (1999); Ginsberg (2000); Hattis (1996, 2000a, 2000b); Hattis *et al.* (1987, 1999a, 1999b); Hattis and Silver (1994); ICRP (1975); USEPA(1988, 2000b).

Exposure Factor	Available Information	Basis	Selected Sources
Intake Rates and Related Factors			
Drinking water ingestion	Defaults, means, percentiles, and distributions.	Primarily based on data from large, dated, national surveys (United States and Canadian). Small studies of specific subpopulations.	AIHC (1994); Burmaster (1998c); ExxonMobil (2000); Finley *et al.* (1994a); Gephart *et al.* (1994); ODEQ (1998); USEPA (1980, 1991a, 1991b, 1997a, 2000a).
Soil and dust ingestion	Defaults, means, percentiles, and distributions.	Multiple studies of trace elements in soil and in children's feces. Adult values loosely derived from children's studies based on professional judgment and results from one adult tracer study.	AIHC (1994); ATSDR (2000); Calabrese *et al.* (1997); Calabrese and Stanek (1992); ExxonMobil (2000); Finley *et al.* (1994a); Gephart *et al.* (1994); ODEQ (1998); Stanek and Calabrese (1992, 2000); USEPA (1989a, 1997a, 2000a).
Fruit and vegetable ingestion	Means, percentiles, and distributions. Per capita and consumer-only intake rates for total fruit and vegetable intake, and of intake of individual fruits and vegetables.	Large national surveys.	Berry (1997); Driver *et al.* (1996); ExxonMobil (2000); Gephart *et al.* (1994); Gregory *et al.* (1995); Hamey and Harris (1999); MDH (2000); Novigen (2000); ODEQ (1998); Tomerlin *et al.* (1997); USDA (2000); USEPA (1997a, 1999b, 2000a, 2000g, 2000i).
Fish and shellfish ingestion	Means, percentiles, and distributions. Varies with study.	General population and recreational marine fish intake rates are based on large, national surveys. Recreational fish intake rates are based on waterbody- or state-specific studies. Freshwater fish intake rates for Native Americans are based on population-specific studies.	AIHC (1994); Burger *et al.* (1999); Ecology (1997); ExxonMobil (2000); Finley *et al.* (1994a); Gephart *et al.* (1994); Jacobs *et al.* (1998); Murray and Burmaster (1994); ODEQ (1998); Ruffle *et al.* (1994); USEPA (1997a, 2000a, 2000f).
Meat and dairy ingestion	Means, percentiles, and distributions. Per capita and consumer-only intake rates for total meat and dairy intake, and of intake of individual meats.	Large national surveys.	AIHC (1994); ExxonMobil (2000); Gephart *et al.* (1994); ODEQ (1998); USEPA (1997a, 2000a).
Grain ingestion	Means, percentiles, and distributions. Per capita and consumer-only intake rates for total grain and individual grain products.	Large national surveys.	ExxonMobil (2000); USEPA (1997a, 2000a).
Ingestion of home-produced food items	Means, percentiles. Per capita and consumer-only mean intake rates and percentiles for total home-produced fruits, vegetables, meats, fish, and dairy products; individual home-produced food items; and various USDA classifications.	One large national survey.	AIHC (1994); ExxonMobil (2000); USEPA (1997a, 2000a).
Ingestion of breast milk	Mean and percentiles.	Small studies of infants between the ages of one month to one year.	ExxonMobil (2000); USEPA (1997a, 2000a).

Table 1. (continued)

Exposure Factor	Available Information	Basis	Selected Sources
Inhalation	Defaults, means, percentiles, and distributions for long-term exposure for adults and children and for short-term exposure associated with various activities.	Long-term exposure rates: primarily one large U.S. study. Short-term exposure rates: primarily a few small, Los Angeles-based studies using observations and predictive equations.	AIHC (1994); ExxonMobil (2000); Finley *et al.* (1994a); Layton (1993); Leighton (2001); ODEQ (1998); Rusconi *et al.* (1994); USEPA (1989b, 1991b, 1997a, 2000a, 2001c).
Soil adherence to skin	Mean and geometric standard deviation by activity and body region. Distributions.	One small recent study used field measurements categorized by activity, gender, age, field conditions, and clothing. Historical data from laboratory and field studies of adherence to hands.	Multiple, including ExxonMobil (2000); Finley *et al.* (1994a,1994b); Holmes *et al.* (1999); Kissel *et al.* (1996); ODEQ (1998); USEPA (1992a, 1997a, 2000a).
Transfer factors of chemicals to skin from various surfaces	Formulas and experimental data for estimating transfer rates of dislodgeable residues to skin from vegetation and other surfaces.	Experimental studies of human volunteers for multiple pesticides and other chemicals.	Multiple, including Durkin *et al.* 1995; Brouwer *et al.* (1999); USEPA (1997d, 1999b).
Behavioral Factors Related to Activity Patterns			
Population mobility—residential	Average, percentiles, and distributions for U.S. population and various subpopulations.	Primarily U.S. Census data.	AIHC (1994); ExxonMobil (2000); Field *et al.* (1998); Finley *et al.* (1994a); Gephart *et al.* (1994); Israeli and Nelson (1992); Johnson and Capel (1992); ODEQ (1998); Sedman *et al.* (1998); U.S. Bureau of the Census (1993); USEPA (1997a, 2000b).
Occupational tenure	Medians for various categories (*e.g.*, age, gender, race, occupation, and earnings). Distribution.	Two studies using census data.	AIHC (1994); Carey (1988, 1990), as summarized in USEPA (1997a); ExxonMobil (2000); Finley *et al.* (1994a); Gephart *et al.* (1994); ODEQ (1998).
Time spent indoors/ outdoors	Means, standard deviations, and distributions.	Large national studies using diary techniques (*e.g.*, the National Human Activity Pattern Survey [NHAPS]).	AIHC (1994); ExxonMobil (2000); ODEQ (1998); USEPA (1996a, 1996b, 1997a, 1999b, 2000a, 2000b).
Time spent engaged in specific activities	Means, standard deviations, some distributions for typical daily activities.	Several national studies using diary techniques (*e.g.*, the NHAPS). Studies in several states (including California) using recall survey and other techniques. Surveys of activities of specific workplace populations, *e.g.*, janitors.	AIHC (1994); Clayton and Perritt (1993); CMI (2001); ExxonMobil (2000); Gephart *et al.* (1994); Hill (1985); IATUR (2000); Klepeis *et al.* (2001); McCurdy (2000a); McCurdy *et al.* (2000); ODEQ (1998); Robinson and Thomas (1991); Silvers *et al.* (1994); and others, as summarized in USEPA (1997a); Timmer *et al.* (1985); UNSD (2000); USEPA (1996a, 1996b, 1997a, 1999b, 2000a, 2000b, 2000d); Wiley *et al.* (1991); Wong *et al.* (2000); UM (2000); Wong *et al.* (2000).
Time spent smoking	Means, standard deviations, and percentiles.	Several national studies using diary techniques (*e.g.*, the NHAPS).	USEPA (1996a, 1996b, 1997a, 2000b).
Time spent bathing or showering	Means, standard deviations, percentiles, and distributions.	Several national studies using diary techniques (*e.g.*, the NHAPS).	AIHC (1994); Burmaster (1998b); ExxonMobil (2000); Finley *et al.* (1994a); James and Knuiman (1987); USEPA (1996a, 1996b, 1997a, 2000a, 2000b).

Exposure Factor	Available Information	Basis	Selected Sources
Time spent in specific locations	Means, standard deviations, percentiles, and distributions.	Several national studies using diary techniques (*e.g.*, the NHAPS). Studies in California using recall survey and other techniques.	AIHC (1994); CARB (1991a, 1991b); Clayton and Perritt (1993); ExxonMobil (2000); Field *et al.* (1998); Funk *et al.* (1998); Glickman (1986); Klepeis *et al.* (2001); Robinson and Thomas (1991); USEPA (1996a, 1996b, 1997a, 2000b); Wiley et. al. (1991).
Frequency, duration, and amount of use of specific consumer products	Means, standard deviations, and percentiles of use data for a variety of household products, including solvents, paints, personal care products, and household cleaners. Other relevant information available in general risk assessment approaches developed for consumer products.	Large- and small-scale studies conducted by industry groups and U.S. government agencies.	Abt (1992); Adams (2001); Api (2000); Driver (2000); ECETOC (1994); Feijtel *et al.* (2000); FEMA (2000); Hakkinen (1993); Hakkinen *et al.* (1991); Loretz (2000); MacPherson (2000); Montalbano (2001); RIFM (2000); Riley *et al.* (2001); Scheringer *et al.* (2001); SDA (2000); Thomas (2001); USEPA (1996a, 1996b, 1997a, 2000a); van Veen (1996); Weegels and van Veen (2001); Westat (1987a, 1987b, 1987c); Williams (2000).
Children's hand and mouthing activities	Mean, median, and ranges of rates (contacts per hour) of children's hand contact with their other hand, objects, mouth, clothing, dirt, and surfaces.	Study using videotaping and parent questionnaire.	Reed *et al.* (1999).
	Prevalence of mouthing behaviors in children on daily, weekly, or monthly basis, including estimates of frequency of contact and total mouthing time; and other studies exploring relationship between children's dietary and mouthing habits and exposures to specific chemicals (e.g., lead).	Several small studies of children up to 6 years old using parent and researcher observations, parent interviews, or videotape techniques.	Freeman *et al.* (1997); Reed *et al.* (1999); Stanek *et al.* (1998); USEPA (1997d, 1999c, 2000a); Zartarian (1999); Zartarian *et al.* (1997b); Zartarian and Leckie (1998).
Demographic Factors			
Sociodemographic characteristics	Detailed numerical data on size, distribution, and demographic characteristics of geographically defined populations in the United States.	Comprehensive national Census database.	U.S. Bureau of the Census (1995); USEPA (1999a).
Residence location and characteristics	Numerical data to support identification of residential factors that may lead to increased exposure and risk.	Variety of listings, surveys, databases, maps, and reports supported by federal and other agencies and other sources.	USEPA (1999a).
Nonresidential settings	Numerical data to support identification of populations affiliated with various nonresidential settings and characteristics of such settings.	Census and other survey data collected by federal and other agencies.	USEPA (1999a).
Activities affecting exposure	Numerical data to support identification of populations participating in various activities that can influence exposure and characteristics of those populations.	Census and other survey data collected by federal agencies and other entities.	USEPA (1999a).

Table 1. (continued)

Exposure Factor	Available Information	Basis	Selected Sources
Behavioral and cultural practices	Numerical data to support identification of populations participating in various behavioral or cultural practices that can influence exposure and characteristics of those populations.	Census and other survey data collected by federal agencies and other entities.	Multiple including CARB (1991a, 1991b); USEPA (1996a, 1996b, 1999a).
Drinking water and food sources	Numerical data to support identification of water and food sources for various populations and characteristics of those populations.	Census and other survey data collected by federal agencies and other entities.	USEPA (1996a, 1999a).
Socioeconomics	Numerical data to support identification of socioeconomic factors that may influence exposure potential for various populations and characteristics of those populations.	Census and other survey data collected by federal agencies and other entities.	USEPA (1999a).
Environmental Modeling Factors			
Uptake/transfer factors	Varies by chemical.	Predictive equations based on physical and chemical properties; chemical-specific and site-specific studies.	Specific values and techniques for estimating some values for environmental modeling provided in resources such as Lyman *et al.* (1990), SRC (2001), and Arms (1988), Verschueren (1996), EPA's Health Assessment Documents, ATSDR Toxicological Profiles, and chemical-specific studies.
Residential air exchange rates, air flow rates, and volumes.	Means, standard deviations, percentiles, and distributions. Raw data are available. Other relevant information available in individual studies, *e.g.*, Clothes *et al.* (1992) and Kerger *et al.* (2000).	Primarily two large studies of residential characteristics.	CCHT (2001); Koontz and Rector (1995); Murray (1997); Murray and Burmaster (1995); TIIE (2001); U.S. DOE (1995); USEPA (1997a, 2000b, 2000c, 2001b); VERSAR (1990).
Chemical transformation rates	Varies by chemical.	Predictive equations based on physical and chemical properties; chemical-specific and site-specific studies.	Specific values and techniques for estimating some values for environmental modeling provided in resources such as Lyman *et al.* (1990), SRC (2001), Verschueren (1996), EPA's Health Assessment Documents, ATSDR Toxicological Profiles, and chemical-specific studies.
Meteorological factors	Varies by site.	National data collection network.	National Weather Service, U.S. Department of Commerce.
Chemical distribution	Varies by site.	Site-specific and scenario-specific studies.	

Table 2. Summary of review of overall data adequacy.[a]

Exposure Factor	Overall Data Adequacy		
	High	Medium	Low
Physical/ physiological factors	Body weight	Skin surface area*	Absorption factors*
	Life expectancy		Factors influencing internal dose*
Intake rates and related factors	Water ingestion	Dietary ingestion	Soil and dust ingestion*
	Inhalation	Breast milk ingestion	Soil adherence to skin*
			Chemical transfer to skin
Behavioral factors related to activity patterns		Population mobility	Children's hand and mouthing activities*
		Occupational tenure	
		Time use	
Demographic factors	Sociodemographic characteristics	Residential and nonresidential settings	Activities affecting exposure*
			Water and food sources*
			Socioeconomics*
Environmental modeling factors		Residential modeling data	Uptake/transfer factors*
		Meteorological data*	Chemical transformation rates*

Notes:

[a] The data adequacy ratings were assigned considering published ratings of data quality, the perspectives of experts consulted during this project, and the adequacy of the data relative to that available for other parameters.

* An asterisk indicates that the adequacy of the data for the indicated factor was assigned a hybrid rating consisting of the rating category that the factor is listed under and the next highest category. For example, the data adequacy for skin surface area was assigned a medium/high rating.

to support distributions of factor values. By contrast, data for factors such as chemical uptake rates into beef or dairy cattle are so limited that only point estimates are typically available.

Regulatory agencies and others have recommended default deterministic values for many exposure factors and, in some cases (*e.g.*, ODEQ 1998; AIHC 1994; ExxonMobil 2000), have recommended default distributions. Factors for which default values exist include general factors (*e.g.*, body weight), intake rates (*e.g.*, soil or food ingestion rates), absorption factors (*e.g.*, from the gastrointestinal system), and temporal factors (*e.g.*, exposure duration or frequency). Distribution information is available for many of these factors, reflecting a broader perspective on the real variations that exist in these factors than can be accounted for by single point estimates. This step of the critical review included an assessment of the degree to which the available default values represent the probable actual values of the specific exposure factor when the available distribution information is considered.

Specific findings for each category of exposure factor are discussed below.

Individual Physical/Physiological Factors — Parameters in this category include physical parameters that describe receptor size (*i.e.*, body weight and skin surface area); life expectancy; absorption factors associated with gastrointestinal, inhalation, and dermal exposures; and pharmacokinetic and pharmacodynamic factors that influence the internal dose that reaches target tissues (*e.g.*, metabolic rate constants and partition coefficients). Data for the first three factors listed in Table 1 (*i.e.*, body weight, skin surface area, and life expectancy) are quite extensive and have benefited from data collection efforts conducted in areas other than environmental risk assessment (*e.g.*, medical and insurance-related research on body weights and life expectancy). Large national databases have been used to generate default values, summary statistics, and distributions for these parameters. For each of these factors, the available default values are likely to be adequate for general exposure scenarios. Evaluations of specific scenarios or subpopulations may merit analyses of the underlying data, however, to derive values that better reflect the exposed population of interest. The adequacy of these data is generally judged to be high. For these parameters (as well as virtually all the others), the data are least adequate for assessing intraindividual variability, where applicable. In general, any ongoing or proposed research regarding these parameters consists primarily of fine-tuning to update the existing values to reflect more recently collected data.

For skin surface area, a related factor is the amount of exposed skin through which absorption may occur during various activities. Little data exist to support values for this factor. Instead, assumptions for this factor are generally based on professional judgment. The adequacy of these data is assigned a lower rating. These data raise particular concerns regarding how well generic values apply to specific exposure situations. The degree to which differing skin areas are exposed and the types of exposures that occur can vary widely depending on many factors, including the types of activities engaged in, seasonal factors, and intraindividual variations from exposure event to exposure event (*e.g.*, what types of activities are engaged in each time, the intensity of the activity, or the time lapse between the event and cleanup activities). Inter-relationships between the various factors influencing dermal exposure are also poorly understood (*e.g.*, activity type, exposed skin surface area, and soil adherence), which presents additional concerns regarding the appli-

cation of default values. Research is ongoing as part of efforts to better characterize dermal exposure.

The remaining factors in this category include chemical-specific elements. Data for these factors are more limited and vary widely depending on the chemical. For the absorption parameters, default values have typically been set to reflect the maximum possible absorption (*i.e.*, 100% absorption in many cases; see, *e.g.*, USEPA 1999b). In some cases, alternative assumptions have been derived for specific chemicals or groups of chemicals based on theoretical predictions. Observations in humans, *in vivo* animal studies, and *in vitro* test systems have also been used to derive chemical-specific absorption values in some cases; however, data for most chemicals are limited. Even for the best-studied chemicals and exposure routes, data are typically insufficient to support development of distributions of potential absorption values in the general exposed population or specific exposed subpopulations; however, the potential range of absorption values is limited by definition to between 0 and 100%.

Chemical-specific studies have shown that absorption values can differ substantially from default values depending on such factors as the specific chemical or physical form, the matrix in which exposure occurs, and the characteristics of the receptor (*e.g.*, age or health status). As a result, situation-specific evaluations of absorption typically are merited when absorption assumptions may substantially affect exposure-modeling results. Research into both generic and chemical-specific elements of these factors is active, particularly for certain chemicals and exposure routes, *e.g.*, gastrointestinal absorption of various types of metals (see, *e.g.*, Ruby *et al.* 1999) and dermal absorption of persistent organic chemicals found in soil at contaminated sites such as pentachlorophenol (see, *e.g.*, Qiao *et al.* 1997).

The final set of factors listed in this category (*i.e.*, factors influencing internal dose) encompasses a wide range of generic physiological factors and chemical-specific factors that describe the uptake and disposition of chemicals in humans. For these factors, general data are available describing the dimensions, composition, and distribution of various anatomical tissues and critical aspects of various physiological processes in humans. Chemical-specific information is also available regarding uptake, metabolism, elimination, and other aspects of chemical disposition in humans. For some chemicals, *e.g.*, certain volatile chemicals such as chloroform and trichloroethylene, extensive data have been collected and applied in detailed pharmacokinetic and pharmacodynamic models. For most chemicals, however, few data have been collected, and little in the way of default values, summary statistics, or distributions is available. Efforts have been ongoing for many years to compile chemical-specific data reflecting interindividual variability in these parameters from literature reflecting environmental chemicals as well as pharmaceuticals. These data are being used to assess appropriate values for specific chemicals as well as to evaluate generic issues (*e.g.*, the general magnitude of variability in these parameters, or whether the default uncertainty factors used in deriving toxicity factors for noncancer health effects are adequate). Research is actively ongoing in this area, and the compiled data are being expanded to reflect data specific for children. The compiled data are available through the Internet.

For both the absorption factors and those factors influencing internal dose, the adequacy of the available data varies with the specific factor and chemical under

consideration. In general, these data are judged to have low/medium adequacy.

Intake Rates and Related Factors — Factors in this category include ingestion rates for water, soil and dust, food, and breast milk; inhalation rates; and dermal contact factors such as soil adherence to skin and transfer factors of chemicals to skin from various surfaces. Parameter estimates for inhalation rates and ingestion rates for water and most forms of food are generally based on large-scale national databases, which are generally considered dated to varying degrees. In most cases, these data have typically been used to derive default values, summary statistics, and distributions for use in estimating exposures for the general U.S. population. Data regarding these intake factors have also been collected in smaller studies of specific subpopulations (*e.g.*, inhalation rates for groups engaging in athletic activities or ingestion rates for persons raising home-grown produce or participating in recreational fishing at specific locations). In some cases, these data have been used to derive summary statistics and distributions for use with these subpopulations or at specific sites. Data on related factors are also frequently collected in these smaller studies, *e.g.*, on serving sizes or the frequency of engaging in specific activities.

The adequacy of data for these parameters is assigned a high overall rating, reflecting the fact that although some limitations exist in the available data for both of these factors, physiological bases exist for establishing reasonable bounds on the possible values for these parameters. In addition, the available default values for these factors are likely to be adequate for general exposure scenarios. Evaluations of specific scenarios or subpopulations may merit additional analyses, however, to derive values that better reflect the exposed population of interest. Ongoing research on these factors includes updates of the national food intake surveys (*i.e.*, the U.S. Department of Agriculture's Continuing Survey of Food Intake by Individuals) and subsequent analyses of these data to update related exposure factor estimates. Additional studies of specific subpopulations (*e.g.*, individuals that participate in recreational fishing at specific locations) are also being conducted.

The remaining factors (*i.e.*, soil and dust ingestion, ingestion of breast milk, and dermal contact factors) are supported by much more limited data. Although default values, summary statistics, and in some cases distributions of these values have been derived from the available data, questions exist regarding the applicability of the available data and how the factors may vary depending on specific exposure conditions (*e.g.*, how soil ingestion may vary with season, region of the country, or socioeconomic factors). Data are also very limited for specific aspects of these factors, *e.g.*, the prevalence and magnitude of breast milk intake in infants greater than 1 year old.

Medium overall ratings are assigned to the adequacy of the data supporting parameters for dietary intake and breast milk ingestion. In general, the adequacy of the dietary intake data is lower for assessing long-term intake rates than for assessing short-term intake rates. Data adequacy is also lower for assessing children's intake rates than for assessing adults' intake rates. Low/medium ratings are assigned to the adequacy of the data for soil and dust ingestion and dermal contact factors. These ratings reflect the substantial limitations in the available data. These limitations include the small number of participants in some of the underlying studies and the limited availability of data reflecting different exposure conditions. Data are also

sparse for characterizing soil and dust intake in adults and individuals with pica or for distinguishing between soil and dust intake.

The default values for dietary intake and breast milk intake are likely to be adequate for general exposure scenarios, particularly the central tendency values derived from these data. Questions exist, however, regarding the degree to which the upper-bound values derived from these data reflect actual exposure patterns and the degree to which data based on short-term observations adequately reflect long-term exposure patterns and current exposure practices. Specific populations or exposure scenarios of concern may also merit tailored analyses. For all of the factors considered in this section, data are considered most adequate for characterizing central tendency estimates and less adequate for characterizing upper-end estimates.

The adequacy of the available default values for soil and dust ingestion and dermal contact factors raises the most questions. Interpretation of the currently available data is subject to considerable debate. Moreover, these parameters can vary significantly depending on the specific exposed population and the types of activities engaged in. However, the magnitude and mechanisms of such variations are poorly understood. Limitations in the available data for specific aspects of these factors also raise questions regarding how well the standard default values reflect actual exposures.

Additional data collection on all of these factors has been proposed or is ongoing.

Behavioral Factors Related to Activity Patterns — This category includes factors related to the duration of time spent at specific residences or in specific occupations, the amount of time spent in specific locations or engaged in specific activities, and specific behaviors that influence the magnitude of exposure (*e.g.*, children's hand and mouthing activities). Data are available for the general U.S. population on most of the factors based on large national studies. Default values have been developed for certain of these factors, and summary statistics and distributions have also been compiled. In addition, many of the large databases are also available through the Internet either directly or linked with environmental models. Thus, tailored analyses of specific subgroups of the populations included in the studies are also possible. Data are also available on certain factors through smaller-scale studies or studies focused on specific aspects of the factor. Thus, availability of data for specific activities, products, exposed subpopulations, or chemicals is highly variable. In addition, some of the information is highly dispersed or is not readily available in a public format. For example, product use surveys conducted by private companies or trade associations may be kept as confidential business information or may only be accessible to members of the trade association.

For almost all of the factors included in this category, the overall data adequacy is assigned a medium rating. Factors enhancing the data adequacy include the large size and national basis of much of the data underlying these parameter values. However, these positive factors are tempered by questions regarding the applicability of these data, that were collected over a short time frame, for assessing potential exposures occurring over longer periods of time. In addition, other limitations in the available data include questions regarding whether the data are dated or whether they accurately reflect current and potential future exposure patterns. Questions also exist regarding how well the available data reflect the full range of

variability in activity patterns that may exist in the population and, in particular, how well individuals at the upper and lower ends of the distribution of activity pattern behavior are reflected in the available data.

Currently available data on more detailed aspects of behavioral factors influencing exposure (*e.g.*, children's hand and mouthing behaviors) typically are derived from small-scale studies. Some summary statistics are available for factors such as the number of contacts per hour of children's hands with objects, mouth, or surfaces; however, these data are generally obtained from a small number of observations (*e.g.*, only a few children). Work is ongoing to incorporate these types of factors in exposure assessment models; however, additional data are needed to better characterize associated exposure parameters (*e.g.*, chemical transfer rates associated with various types of contacts). The adequacy of the data for assessing children's hand and mouthing activities is assigned a low to medium overall rating. This rating reflects the generally small size of the studies of these factors, the preliminary nature of investigations into these factors, and limitations in the availability of modeling approaches and associated exposure data required to incorporate these factors in exposure modeling. Research in this area is ongoing.

Default values are available for selected factors included in this category. In general, the available default values are likely to be suitable for use in general exposure assessments. For many exposure evaluations, however, more specific analyses focusing on specific subpopulations or exposure scenarios of interest are likely to be warranted. Questions also exist regarding the application of some of the available default factors and data. For example, the data underlying estimates of time spent at a specific residence or in a specific occupation generally reflect the time spent to date at that location or type of work and may not adequately account for the additional time that will be spent after the survey information is collected. Concerns have also been raised regarding whether the use of these factors adequately accounts for similar exposures that may occur even after an individual changes residences or occupations. For those factors for which default values are not yet available, the available data need to be carefully evaluated to ensure their applicability to specific exposure scenarios of interest.

Demographic Factors — This category includes a variety of types of data to support evaluations of the prevalence of sensitive populations and demographic factors and other population characteristics affecting exposure. Most of the data available to support such evaluations are not readily summarized as default values and distributions. Instead, the available data consist largely of numerical and listing data that can be used to assess the overall prevalence of specific populations of interest or the likelihood of the presence of such populations at specific sites. Most of the available data are derived from large national surveys, such as the U.S. Census, and from other surveys, listings, databases, maps, and reports prepared by federal agencies and other entities. Some of these data sources are periodically updated (*e.g.*, the U.S. Census).

With respect to the adequacy of the available data, the factors included in this category fall into three subgroups. The overall adequacy of the data supporting evaluations of sociodemographic factors, such as race or household composition, is assigned a high rating because these factors parallel the types of data collected in the national Census, the goal of which is to provide a comprehensive enumeration

of the entire U.S. population. The adequacy of data supporting evaluations of residence and nonresidence locations and associated characteristics is assigned a medium rating overall. This rating reflects the fact that components of these data are determined in part from Census data and in part from other sources that may not be as comprehensive as Census data.

The adequacy of the data supporting the remaining factors included in this category is assigned a low/medium overall rating. This rating reflects the wide variation in the amount and type of data available to support specific factors. For example, data regarding participation in certain activities or behavioral practices has been collected by organizations supporting those activities, whereas no data are available for other activities or practices. For those activities for which data have been collected, the degree to which the data are comprehensive or represent the activity levels of all individuals who participate in the activity may be unclear. No default values are available for any of the factors in this category.

Environmental Modeling Factors — This category includes uptake/transfer factors, characteristics of the setting or building that influence exposure, chemical transfer rates, meteorological factors, and chemical distribution data. Many of these factors are either chemical-specific (*e.g.*, uptake factors and chemical transformation rates) or are highly dependent on the specific site or situation of interest (including building or setting characteristics [*e.g.*, Kerger *et al.* 2000] or chemical distribution in the environment). These factors and associated exposures can also be influenced by specific human activity patterns such as the use of air conditioning or of specific chemical-containing consumer products (*e.g.*, Clobes *et al.* 1992).

Data relevant for assessing some of these factors may be available or could be developed from studies conducted in test houses. For example, test houses are currently maintained by the USEPA (USEPA 2001b) and the Canadian Center for Housing Technology (CCHT 2001). These houses are used to conduct studies assessing the residential fate and transport of consumer products (*e.g.*, Sparks *et al.* 1999; Sparks *et al.* 1991, Guo *et al.* 1992) and the influence of new residential building materials and designs on residential air quality. The Texas Institute for the Indoor Environment at the University of Texas is also developing a test house system (TIIE 2001).

Default values, summary statistics, and distributions are available for some of these factors (*e.g.*, residential air exchange rates). For other parameters, default values may be available from various compilations or may be estimated by using various predictive equations. Default values or approaches are also provided in various exposure models. In general, little information is available regarding how some of these factors may vary under differing site-specific conditions (*e.g.*, how plant uptake of chemicals varies depending on soil conditions or the specific chemical or plant species of interest). Some updating of these values occurs periodically; however, this information typically is widely dispersed and may not be directly applicable in specific situations.

In this category, the data supporting meteorological factors are assigned a medium/high overall rating. This rating reflects the extensive network of supporting data tempered by limitations in data that may exist for specific sites. Questions also exist regarding the implications of meteorological conditions for other exposure parameters (*e.g.*, how weather conditions influence exposures to soil and dust). The

adequacy of the data supporting certain modeling factors (*e.g.*, residential air exchange rates, air flow rates, and volumes) is assigned a medium overall rating. This rating reflects the large national database that underlies evaluations for these factors; however, some questions remain regarding the degree to which the available data represent conditions in specific exposure situations. The adequacy of data supporting various chemical-specific uptake and transfer factors is assigned a low/medium rating overall. This rating reflects the wide variation in the amount of data available for specific factors and specific chemicals. In general, available data are sufficient to derive point estimates of these values but provide little information on the distribution of possible values.

Generic default values are not available for most of these parameters. In those few cases in which default data are available, they may not represent specific geographic, seasonal, or other differences reflected in exposure scenarios of interest. Therefore, the underlying data for these factors should be carefully reviewed before they are applied in specific exposure modeling exercises.

APPLICATION OF RESULTS

This review focuses primarily on the data available for developing point estimates and distributions of specific factors required as inputs for modeling human exposures to nonpesticide chemicals in a variety of settings. As such exposure models are developed and refined using these input parameter data, other aspects of the broader realm of research into exposure assessment will directly influence model development and will provide valuable context for modeling efforts. These broader contexts also provide additional perspective on the relative importance of various exposure factors and the priority for filling identified data gaps. These additional areas include the following:

- Developing a more detailed understanding of chemical disposition in the body and the biological mechanisms by which specific chemicals exert effects
- Evaluating whether model predictions are reasonable in light of monitoring data, data from animal studies, and other data sources
- Evaluating alternative methods for estimating overall exposures.

These areas are briefly discussed below.

Biological Mechanisms of Action — Among the factors included in this review are those influencing internal dose, such as metabolic rate constants and partition coefficients. The research efforts included in this review involve compiling available data on these factors, examining general trends in variability in these factors, and assessing their general implications for exposure and risk assessment (*e.g.*, Hattis *et al.* 1999a,b; Burin and Saunders 1999). In addition to these efforts, researchers are also applying these factors in the context of a variety of types of exposure information to develop detailed physiologically based pharmacokinetic (PBPK) models of the disposition and mechanisms of action of specific chemicals (*e.g.*, Wu and Schaum 2000). The degree of development of these models varies widely; some chemicals (such as several volatile solvents) have been the subject of extensive

research (*e.g.*, Barton *et al.* 1996; Allen *et al.* 1996), but many others have yet to be studied.

Researchers are also working to develop a better understanding of the underlying factors that influence variability in these factors and affect susceptibility to adverse health effects (*e.g.*, Chance and Harmsen 1998; Cohen Hubal *et al.* 2000a; Graeter and Mortensen 1996; Weaver *et al.* 1998; Grassman 1996). These efforts include work to develop consistent frameworks to define susceptibility (*e.g.*, Parkin and Balbus 2000; Pastino *et al.* 2000), particularly as regulatory concerns for susceptible subpopulations such as children have increased (*e.g.*, Clinton 1997; BNA 2000).

A number of techniques are being used to explore issues of how these factors vary and how they influence susceptibility to adverse health effects. These efforts include studies of cellular processes that may influence susceptibility and animal tests using inherently sensitive populations. Probabilistic modeling techniques and sensitivity analyses have been applied in dose-response evaluations to explore the impacts of assumptions regarding variability in PBPK modeling parameters on predicted toxic effects (*e.g.*, Kuempel *et al.* 2001; Sweeney *et al.* 2001; Allen *et al.* 1996; Cronin *et al.* 1995; and other studies reviewed in Petito Boyce 1998). Data linking variability in PBPK parameters and changes in adverse effects are limited, as are data regarding the pharmacodynamics of adverse health responses (Barton *et al.* 1996).

The interplay between exposure timing and effects is another area in which researchers are seeking a more detailed understanding of both the fundamental workings of toxicological processes and the mechanisms determining dose-response relationships for specific chemicals (*e.g.*, Boyes *et al.* 2000; Witschi and Hakkinen 1984). Research of this type has highlighted the importance of detailed information regarding mechanism of action in more accurately modeling potential exposures and risks. It also indicates a need for more detailed mechanistic toxicity data as an adjunct to the types of exposure modeling data explored in this review.

Developments in each of these areas of model refinement will enhance and complement the efforts directed at more generic developments in exposure assessment modeling that are the focus of this report. To the extent that any of these areas becomes more commonly applied in exposure assessment or becomes important for a specific chemical, it will also direct additional data collection needs.

Comparing Model Predictions with Empirical Observations — Another important context to be considered in applying the exposure factor data and determining additional data needs is the issue of evaluating model predictions in the context of relevant empirical observations (*e.g.*, monitoring or biomarker data). This context becomes particularly important as exposure assessment models become more complex and include greater numbers of exposure pathways, scenarios, and assumptions. In a regulatory framework, input assumptions are typically made conservatively to yield health-protective results. As the number of conservative assumptions increases, however, model results may increasingly diverge from a reflection of plausible results. Comparing model predictions with relevant empirical observations can provide a benchmark for assessing the degree to which the exposure model predictions are reasonable. Such comparisons can also yield insights regarding data collection needs for improving model predictions.

Existing data sources may provide a basis for evaluating predictive exposure models. For example, several researchers have used activity pattern data collected

in Denver, CO, and several other communities to develop probabilistic models of microenvironmental and total exposure to carbon monoxide (Law *et al.* 1997; Ott *et al.* 1988). Studies of exposures to pesticides and spray paints have also compared monitoring and modeling results (Krieger *et al.* 2001; Brouwer *et al.* 2001). Other resources that might provide data that could support similar studies include the ongoing data collection activities of the NHEXAS program (which includes personal monitoring data and biomarker data for certain chemicals, as well as data on certain exposure modeling factors such as activity patterns; see, *e.g.*, Sexton *et al.* 1995a), data collected as part of the TEAM study (see, *e.g.*, Wallace 1993), monitoring data collected in a number of regulatory contexts (*e.g.*, as compiled in USEPA's Envirofacts database; USEPA 2001a), and data contained within the exposure databases reviewed in Sexton *et al.* (1994). Alternatively, to conduct some types of comparative studies, it may be necessary to design and implement new research combining collection of relevant monitoring data and data regarding specific exposure factors of interest. Duan and Mage (1997) review issues in combining monitoring and modeling data.

Alternative Exposure Assessment Methods — A final area of context to consider is the overall framework being used to model exposures. Commonly used exposure modeling efforts and the factors reviewed in this report primarily focus on using contact with various media and duration of time spent in various locations as the measure of overall exposure. Other bases for modeling exposure are also under investigation. In addition, researchers are also working to ensure a clear and consistent set of definitions to guide exposure analyses, including consideration of physical components of steps in exposure and concepts related to exposure in time and space (see, *e.g.*, Zartarian *et al.* 1997a).

One alternative approach to modeling exposure uses metabolism, specifically energy expenditures, as the basis for assessing overall exposure (*e.g.*, McCurdy 2000a). Activity pattern data are evaluated not only in terms of the time spent in specific locations and activities but also in terms of the amount of energy expended in each activity. Total energy expenditures and patterns of energy expenditure can then be related to required inhalation rates, food and fluid intake, and associated exposures. Dermal exposure assessment is not readily addressed through this approach. For other exposure pathways, however, such an approach may provide a more biologically accurate and integrated perspective on aggregate exposures associated with multiple exposure routes. Moreover, this approach can account for the actual time pattern of exposure, rather than aggregating exposure data in ways that can obscure information critical to an accurate assessment of potential health impacts. This approach is being applied in the CHAD database (McCurdy 2000b) that compiles both activity pattern data and associated energy expenditure estimates drawn from sources of such data (*e.g.*, Ainsworth *et al.* 1993). Again, to the extent that this approach becomes widespread in exposure assessment or becomes important in assessments of specific chemicals, a different category of exposure data will need to be collected, compiled, and analyzed.

In addition to applying alternative measures to assess overall exposures, alternative techniques exist for quantitatively addressing distributional data. To date, the types of exposure factor data reviewed in this report have typically been applied in exposure and risk assessment models by using deterministic approaches or proba-

bilistic approaches relying on Monte Carlo techniques. Alternative approaches are being explored. For example, Petersen *et al.* (1994) presents a joint distributional analysis approach for assessing exposures occurring through ingestion of chemical residues in food. This approach allows intake through all dietary sources to be analyzed simultaneously, without the repeated simulations required in Monte Carlo analyses. The use of such alternative approaches may also modify future data requirements or how data are applied in exposure models.

REVIEW OF DATA GAPS

Overview of Review

This data compilation and review process also included identification of gaps in the available information to support development of exposure factors. Such gaps exist on several levels. First, in some cases, information regarding a certain exposure factor simply may not exist. For example, physiological parameters of the type applied in PBPK models have not been studied for many chemicals. In other cases, data may exist for a certain type of exposure factor, but the degree to which those data are applicable to a variety of exposure scenarios may be uncertain. For example, the data regarding children's incidental ingestion of soil that are commonly applied in risk assessments are derived from observations from a limited number of young children within only a few exposure settings. Questions exist regarding how representative these results are for children in other locations who take part in other types of activities or contact different types of soil. The implications of these data for other age ranges, for which soil ingestion data are even more limited, are also uncertain.

In other cases, data exist, but may not be well compiled. For example, a few sources of information regarding exposures of the general population to consumer products are available. While USEPA's *Exposure Factors Handbook* (USEPA 1997a) contains a limited amount of information regarding frequency and duration of use of certain consumer products (*e.g.*, spray paint and cleaning products), similar existing information for other products is more dispersed. Moreover, information regarding consumer use of specific products is often compiled by individual companies or affiliated researchers for marketing or other purposes, but this information is not always published, compiled in a central database, or made widely available.

The data gap evaluation considered both existing gaps and the relative importance and priority of the gaps (*i.e.*, which gaps would be most worthwhile to address). The priority of filling the gaps was assessed in part on the basis of the types of analyses for which the data may be required. For example, aggregate risk assessments are increasingly being conducted according to the mandates of the FQPA. Such analyses require an understanding of a broader spectrum of exposures that might be encountered in everyday life. The relative priority of identified data gaps also depends on the implications of the missing or inadequate data on exposure assessments. In identifying existing data gaps, this review process drew upon perspectives on existing data gaps presented in published sources (*e.g.*, USEPA 1997a, 2000a; Cohen Hubal *et al.* 2000a,b; ExxonMobil 2000; Whitmyre *et al.* 1992b) and

on perspectives provided by exposure assessment experts contacted during the course of this project.

In assessing the relative priority for filling the data gaps, this review considered the relative extent of the identified data gaps, the relative impact of the exposure factor on exposure modeling, and the general applicability and use of the parameter on modeling. In particular, factors that entail chemical-specific research rather than research supporting general exposure-modeling efforts for all chemicals or broad categories of chemicals were assigned a lower priority ranking. The relative timing of research on specific factors was also considered in assigning priority rankings. That is, identified research that might benefit from the results of additional research efforts in other areas was assigned a lower priority ranking.

In addition to considering limitations in the specific types of data included in this work, this review also considered limitations in the accessibility of the available information. These issues are also discussed below.

General Research Needs

Several general research areas were identified both in published assessments of data gaps and by numerous individuals contacted during this project. These included the following:

- Increasing our understanding of children's exposures to environmental chemicals, including developing better exposure factor estimates focused specifically on children, better methods for monitoring and modeling children's exposures, and better information regarding the relationship between children's activity patterns and exposures

- Designing and conducting exposure assessment studies that better reflect long-term variations, trends, and correlations in exposure, and developing better methods for extrapolating long-term exposure patterns from data collected in short-term studies[1-2]

- Conducting formal analyses (*e.g.*, using value of information and other decision analysis methodologies) or more qualitative analyses (*e.g.*, identifying sources of uncertainty associated with high-cost or high-consequence decisions) to determine priorities for collecting additional data or more detailed data to support exposure modeling

- Conducting confidence-building studies of existing or newly developed exposure assessment models by using existing data or data from specially designed studies to determine whether reasonable results are obtained from predictive models under a variety of exposure conditions

- Developing improved techniques for combining data when multiple studies exist for a specific factor.

---

[2] Issues in study design and extrapolating long-term exposure estimates from short-term data are presented in Buck *et al.* (1997), Price *et al.* (1998), and Wallace *et al.* (1994).

Another general data gap issue that was identified was the need to update certain default exposure parameter values and distributions (*e.g.*, for food consumption rates or consumer product use) either by collecting additional data or by analyzing recently collected data to derive updated values. This issue is discussed in the context of specific exposure parameters in the next section of this report.

The first three data gaps in the preceding listing were assigned a high priority for being filled. This ranking reflects the extent of the data gap, the pervasive need for such data in the types of exposure and risk analyses that are currently of regulatory and other interest, and the desire to provide a strong foundation for directing additional exposure assessment research. The need to compare the results of exposure assessment models with empirical observations was assigned a medium ranking after consideration of relative timing (*i.e.*, research in this area should be conducted after additional exposure assessment model development has occurred). The need for additional approaches for combining available data sets was assigned a low ranking because some meta-analysis techniques exist and the need to develop additional data in certain key areas was accorded a higher priority than needs associated with analysis of existing data.

Research Requirements for Specific Exposure Factors

An overview of the results of the data gap review for specific exposure factors is summarized in Table 3. These identified data gaps are discussed as follows based on the five general categories used as the framework for this review.

Individual Physical/Physiological Factors — Data gaps identified in this category were associated with four specific exposure parameter areas. The first two (*i.e.*, body weight and skin surface area) are related (*i.e.*, existing default assumptions for body weight need to be updated to reflect more recent data). This updating will affect assumptions for skin surface area. These two data gaps were assigned a low ranking because these modifications were seen primarily as fine-tuning of the existing data set. In addition, better information is needed characterizing actual skin surface areas subject to exposure during various activities. This data gap was assigned a high priority because of its substantial influence on estimates of dermal exposures.

The third data gap identified for this category is a general gap for the broad group of factors associated with gastrointestinal, inhalation, and dermal absorption. Substantial data gaps exist in this area regarding the absorption of specific chemicals via specific exposure routes as well as the general processes governing absorption, particularly for dermal absorption. Elements of this data gap that address generic processes of absorption were assigned a high-priority ranking, whereas those associated with specific chemicals were assigned a medium ranking. These rankings reflect a balance between the importance of this factor in exposure assessment and the chemical-specific nature of much of the required research.

The final data gap identified for this category is for factors influencing internal dose, including both generic physiological factors (*e.g.*, blood flow rates and tissue volumes) and chemical-specific factors (*e.g.*, metabolic rate constants and partition coefficients). Such parameters have not been researched or are poorly understood for many chemicals and chemical groups. Moreover, additional research is required to understand trends in variability reflected in these parameters and to develop

Table 3. Summary of factor-specific data gaps.

Exposure Factor	Priority for Filling Data Gaps[a]		
	High	Medium	Low
Physical/ physiological factors	Skin surface area—amounts exposed during specific activities	Absorption factors (chemical-specific elements)	Body weight
	Absorption factors (generic elements)	Factors influencing internal dose (generic elements)	Skin surface area—general
			Factors influencing internal dose (chemical-specific elements)
Intake rates and related factors		Ingestion factors	
		Inhalation	
		Dermal contact factors	
Behavioral factors related to activity patterns	Worker activities		Population mobility
	Time use		
	Consumer products		
	Children's hand and mouthing activities		
Demographic factors			Demographic factors
Environmental modeling factors		Uptake and transfer factors	Residential air exchange rates, flow rates, and volumes
		Room- and activity-specific factors influencing residential air exposures	
		Indoor dust factors	
		Meteorological factors	

Note:

[a] Priority rankings reflect consideration of the influence of the factor on exposure and risk assessment results, the priority accorded to the factor by published data gap listings and contacted experts, and the general applicability of the factor (*i.e.*, factors that will entail chemical-specific research rather than research supporting general exposure modeling efforts for all chemicals or broad categories of chemicals were assigned a lower priority ranking). Relative timing of research into specific factors was also considered (*i.e.*, identified research that might benefit from input from other identified research was assigned a lower priority ranking). Additional discussion of the priority rankings for each data group is presented in the text.

child-specific data regarding these parameters. The generic aspects of this data gap were assigned a moderate ranking, whereas the chemical-specific aspects of this data gap were assigned a low ranking. This ranking reflects the chemical-specific nature of much of the required research. In addition, the types of pharmacokinetic and pharmacodynamic modeling that are associated with these factors were seen as a chemical-specific refinement to the more generic exposure assessment modeling that is the focus of this review.

Intake Rates and Related Factors — Gaps were identified throughout the data underlying the exposure parameters in this category. For both drinking water ingestion and inhalation, additional data are required to characterize these factors for specific subpopulations and activity patterns. In addition, the existing data characterizing drinking water intake should be updated to reflect ongoing changes in fluid consumption patterns. For soil and dust ingestion, numerous limitations exist in the currently available data. These limitations include the scarceness of data regarding soil and dust intake in adults and individuals with pica; data addressing the influence of factors such as season, activity, and region on intake; and research distinguishing between the contributions of soil and those of dust to total intake.

For dietary intake, a general need exists for updated data, more extensive data for children, and data that better characterize long-term intake patterns. Specific needs also exist for better information regarding geographic sources of food for specific subpopulations and for better data or compilations of data characterizing intake patterns in specific subpopulations (*e.g.*, agricultural populations and people participating in recreational or subsistence fishing). Additional data are also required to characterize the prevalence and magnitude of breastfeeding.

In the area of dermal contact factors (including soil adherence to skin and transfer factors), additional research is needed to develop overall soil contact rates from activity-specific soil adherence estimates, data regarding activity frequency, and data regarding the degree to which chemicals can be dislodged from various surfaces and transferred to skin, particularly for nonpesticide chemicals. In addition, the existing data need to be expanded to reflect additional types of activities.

In all cases, these gaps were assigned a medium ranking. This ranking reflects the importance of these factors in determining exposure estimates, which is balanced by the substantial foundation of information that exists for many of these parameters.

Behavioral Factors Related to Activity Patterns — Significant gaps were identified in the information underlying most of the factors in this category. In particular, in many of the categories, the currently available data are substantially limited, and questions exist regarding whether the available data adequately characterize all relevant aspects of exposure. For example, for workers, information is generally available on the numbers of workers that participate in various industries or chemical concentrations measured in various workplaces. However, little information is available to characterize the specific activities that workers engage in, the frequency and duration of activities, the types of microenvironments they encounter, the degree of exposure that actually occurs, or how measured workplace concentrations may be influenced by these components of exposure.

Similarly, considerable information has been collected regarding activity patterns for the general population (*e.g.*, the amount of time spent in various locations).

Nevertheless, questions remain regarding whether all relevant factors influencing exposure are reflected in the available data (*e.g.*, the specific types of activities or contacts that occur during the time spent in various locations). Children's hand and mouthing activities and children's activity patterns in general are of particular interest. Similar issues exist for available data regarding use of specific consumer products. In all cases, questions exist regarding the degree to which the available data reflect current or long-term exposure patterns.

The final gap identified for factors in this category is associated with data available for assessing population mobility. In this case, questions exist regarding whether the currently available data appropriately reflect the duration of exposures in residential settings and whether the data have been appropriately interpreted.

Because of the magnitude of the data gaps identified for factors in this category, almost all of the gaps were assigned a high ranking. The only exception is the ranking for the data gap associated with the population mobility exposure factor. This data gap was assigned a low ranking because it is viewed as a fine-tuning effort rather than substantial new data collection.

Demographic Factors — In general, the data underlying the factors included in this category could be updated and made more comprehensive. For example, those factors that are characterized using Census data should be updated as new data become available. Similarly, additional data could be located or collected to characterize other factors that may influence exposure. For example, expanded surveys could be conducted to identify individuals participating in some of the activities that are currently identified in existing data compilations using surveys of individuals belonging to organized groups supporting such activities. In addition, surveys could be conducted to characterize participation in other activities that may not be included in existing data compilations. This general data gap was assigned a low ranking because of the high degree to which the necessary research depends on specific exposure scenarios and chemicals of interest.

Environmental Modeling Factors — Five broad groups of data gaps were identified in the factors included in this category. The first data gap is associated with the available data for chemical uptake factors (*e.g.*, from soil into plants and other biota) and transfer factors (*e.g.*, from outdoor soil to indoor dust). Data for these factors have many of the same limitations as those identified for absorption factors, *i.e.*, the data are limited for many specific chemicals as well as for understanding the general processes determining uptake and transfer in specific environmental settings. This data gap was assigned a medium ranking, reflecting a balance between the importance of these parameters and the chemical-specific nature of much of the required information.

To address the second data gap in this category, the data regarding residential air exchange rates, air flow rates, and volumes could be updated by using data collected triennially in the U.S. Department of Energy's Residential Energy Consumption Survey (RECS). The RECS study could also be expanded to include more types of buildings (*e.g.*, daycare centers). This data gap was assigned a low ranking because it primarily reflects fine-tuning of the existing database and additional data needs depend on specific exposure settings of interest.

The third data gap is the limited information regarding room- and activity-specific parameters influencing exposures via residential air. These include data on

room-to-room and room-to-outdoors air exchange rates, intraresidence variation in air exchange rates; variations in room sizes and volumes; and impacts of fans, heating and cooling systems, and other ventilation elements on air flow. Efforts to address this data gap include better compilation of existing information as well as development of new data. This data gap was ranked as a moderate priority, reflecting the importance of these data in assessing exposures to airborne chemicals in residential settings.

The fourth data gap in this category was identified for information supporting evaluations of the factors determining house dust generation, transport, and concentrations. This data gap was assigned a medium ranking, reflecting the importance of these factors in certain indoor exposure settings.

The final data gap identified for this category is associated with the use of meteorological data. Although substantial meteorological data are typically available for a wide range of geographic locations and time frames, substantial limitations exist in the understanding of how these data influence exposure-related activities (*e.g.*, time spent outdoors, soil ingestion rates, dermal contact with soil, or soil transfer into buildings). This data gap was assigned a moderate priority reflecting the importance of these data for modeling certain exposure scenarios.

Required Efforts to Improve Data Accessibility

As noted above, this review also considered limitations in the accessibility of the available information. Data for certain categories of exposure factors have been extensively compiled into databases or reference resources that are readily accessible and widely used (*e.g.*, the compilation of exposure factors commonly applied in risk assessments for contaminated sites in USEPA's *Exposure Factors Handbook*). Many of these resources also reflect extensive organization and analysis of the available data, *e.g.*, to derive point estimates or distributions of values for exposure parameters that are representative of specific population groups. In other cases, exposure factor data are more widely dispersed, and syntheses of the data are either not available or difficult to obtain (*e.g.*, data regarding use of various consumer products). As a result, researchers conducting analyses using these factors face substantial challenges in locating the appropriate data and cannot readily build upon previous efforts that have used the available data.

Similar limitations in building on previous efforts exist even for those data sets that have been relatively well documented and compiled. In addition, the accessibility and usefulness of some of these data could be improved if the data were available in formats that are more interactive and dynamic. Since the completion of the primary effort of this review, the USEPA has begun moving in this direction by supplementing the hard copy version of the *Exposure Factors Handbook* with a CD-ROM version (USEPA 1999e) and an on-line electronic version that will incorporate updated data and new data analyses as they become available (USEPA 2001d).

Thus, in addition to filling the data gaps identified above, exposure analyses could be improved by enhancements in the accessibility of the available data. Such enhancements could consist of modifications to the format of existing data compilations. More importantly, such enhancements could include the development of a central resource to consult to identify exposure assessment resources. This resource

could consist of a database that directly incorporates available information or a clearinghouse that could provide pointers to available information. Among the types of information that could be organized in this resource are existing databases of exposure parameters needed for exposure modeling, databases of exposure monitoring data, exposure models, and publications and other documentation of exposure assessment data, approaches, and applications. Models for such an effort include USEPA's Environmental Information Management System (EIMS), which is a tool that provides access to the agency's environmental resources, the bibliographic databases maintained by the National Library of Medicine to provide access to the scientific literature regarding health (*e.g.*, Medline and Toxline), and the compilation of exposure assessment resources currently being developed by the Alliance for Chemical Awareness. The USEPA has begun to compile such a resource in its Exposure Factors Program, which provides links to a variety of types of exposure information that the USEPA has generated (USEPA 2001e).

CONCLUSIONS

The results of this review demonstrate that a rich, although incomplete, database is available to support development of input factors for exposure assessment modeling. Values for some factors (such as body weight, life expectancy, and general demographic factors) can be developed based on extensive data sets, including substantial amounts of data that have been collected and documented in other contexts, such as the medical and insurance literature, and through comprehensive censuses. For other factors (such as those required to develop detailed models of children's activity patterns), exposure assessment researchers are exploring new areas and developing new data collection techniques for adequately reflecting variability in these factors and for identifying which aspects of these factors are most crucial for determining exposures.

As discussed above, the data available to support exposure factors vary widely in their quality and their adequacy to support detailed analyses of the variability and uncertainty inherent in specific exposure factors and scenarios. In general, the data are most adequate for characterizing average or "typical" exposures and less adequate for assessing exposures for individuals that most differ from the norm. Similarly, the available data most commonly are weakest in reflecting intraindividual variability in factors (*i.e.*, how an individual's exposure may vary from exposure event to exposure event) and in reflecting long-term exposure patterns. In many cases, the uncertainty inherent in the available data is not well bounded (*e.g.*, for many factors for characterizing the frequency with which individuals participate in various activities); however, in some cases, factors have been well characterized (*e.g.*, body weight) or the range of potential values for a factor has been defined by biological or other considerations (*e.g.*, for absorption values).

As is true of every scientific research area, additional data are desirable in numerous areas. These include broad needs, such as more data specifically reflecting children's exposures and data reflecting long-term exposure trends. Data needs associated with specific factors also exist, most prominently including data regarding soil and dust ingestion rates, dermal exposure, occupational activity patterns, and activity patterns for the general population reflecting a broader spectrum of

activities and more detailed information regarding those aspects of activities that most influence exposure.

In working toward filling the identified data gaps, enhancing mechanisms for compiling the developed data and making it more readily accessible can be as valuable to improving efforts in the field of exposure assessment as is new research and data collection. For example, establishing a clearinghouse for compiling exposure data related to the use of consumer products and other industry-generated data would provide a useful resource to support product- and chemical-specific exposure assessments. Development of a centralized database or clearinghouse for a broader spectrum of exposure assessment resources would offer correspondingly expanded benefits to exposure assessment research and applications. Evaluating model predictions in the context of empirical data is also critical in assessing the adequacy of existing data and in identifying additional research needs. Moreover, the value of additional research can be enhanced by identifying research priorities through a "big-picture" perspective that explicitly recognizes the value of specific research elements for improving overall exposure assessment modeling methods.

REFERENCES

Abell MT, Woebkenberg ML, Armstrong TW, *et al.* 2001. Research recommendations of the NORA exposure assessment methods team. Appl Occup Environ Hyg 16(2):331-3

Abt. 1992. Methylene Chloride Consumer Products Use Survey Findings. Prepared for the US Consumer Product Safety Commission, Bethesda, MD. Abt Associates, Inc., Cambridge, MA, USA

ACA (Alliance for Chemical Awareness). 2000. www.chemicalawareness.com

ACGIH (American Conference of Governmental Industrial Hygienists). 2001. International Symposium on OccupationalEexposure Databases and Their Application for the Next Millennium. Available at www.acgih.org/leaders/comput/oedb/oedb99.htm

Adams T. 2001. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on February 1, 2001, regarding trade association research into exposure assessment issues). Flavor and Extract Manufacturers Association, Washington, DC, USA

AIHC (American Industrial Health Council). 1994. Exposure Factors Sourcebook. Washington, DC, USA

Ainsworth BE, Haskell WL, Leon AS, *et al.* 1993. Compendium of physical activities: Classification of energy costs of human physical activities. Med Sci Sports Exerc 25(1):71-80

Allen BC, Covington TR, and Clewell HJ. 1996. Investigation of the impact of pharmacokinetic variability and uncertainty on risks predicted with a pharmacokinetic model for chloroform. Toxicology 111:289-303

Api AM. 2000. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on November 6, 2000, regarding trade association research into exposure assessment issues). Research Institute for Fragrance Materials, Hackensack, NJ, USA

ATSDR (Agency for Toxic Substances and Disease Registry). 2000. ATSDR Soil Pica Workshop. Division of Health Assessment and Consultation, Atlanta, GA, USA

Baird SJS, Cohen JT, Graham JD, *et al.* 1996. Noncancer risk assessment: A probabilistic alternative to current practice. Human Ecol Risk Assess 2(1):79-102

Baker S, Driver J, and McCallum D (eds.). 2001. Residential Exposure Assessment: A Sourcebook. Kluwer Academic/Plenum Publishers, New York, NY, USA

Barton HA, Flemming CD, and Lipscomb JC. 1996. Evaluating human varibility in chemcial risk assessment; hazard identification and dose-response assessment for noncancer oral toxicity of trichloroethylene. Toxicology 111:271-287

Berry MR. 1997. Advances in dietary exposure research at the United States Environmental Protection Agency-National Exposure Research Laboratory. J Exposure Anal Environ Epi 7(1):3-16

BNA (Bureau of National Affairs, Inc). 2000. EPA suggests 50 chemicals as candidates for testing to determine threat to children. Washington, DC, USA

Boiano JM and Hull RD. 2001. Development of a national occupational exposure survey and database associated with NIOSH hazard surveillance initiatives. Appl Occup Environ Hyg 16(2):1218-134

Boyes WK, Bushnell PJ, Crofton KM, *et al.* 2000. Neurotoxic and pharmacokinetic responses to trichloroethylene as a function of exposure scenario. Environ Health Perspect 108(2):317-22

Brouwer DH, Kroese R, and Van Hemmen JJ. 1999. Transfer of contaminants from surface to hands: Experimental assessment of linearity of the exposure process, adherence to the skin, and area exposed during fixed pressure and repeated contact with surfaces contaminated with a powder. Appl Occup Environ Hyg 14:231-9

Brouwer DH, Semple S, Marquart J, *et al.* 2001. A dermal model for spray painters. I. Subjective exposure modeling of spray paint deposition. Annals Occup Hygiene 45(1):15-23

Brown RP, Delp MD, Lindstedt SL, *et al.* 1997. Physiological parameter values for physiologically based pharmacokinetic models. Toxicol Ind Health 13(4):407-84

Buck RJ, Hammerstrom KA, and Ryan PB. 1997. Bias in population estimates of long-term exposure from short-term measurements of individual exposure. Risk Anal 17(4):455-66

Bukowski J, Korn L, and Wartenberg D. 1995. Correlated inputs in quantitative risk assessment: The effects of distributional shape. Risk Anal 15(2):215-9

Burger J, Stephens WLJr, Boring CS, *et al.* 1999. Factors in exposure assessment: Ethnic and socioeconomic differences in fishing and consumption of fish caught along the Savannah River. Risk Anal 19(3):427-38

Burin GJ and Saunders DR. 1999. Addressing human variability in risk assessment—the robustness of the intraspecies uncertainty factor. Regul Toxicol Pharmacol 30:209-16

Burmaster DE. 1998a. Lognormal distributions for skin area as a function of body weight. Risk Anal 18(1):27-32

Burmaster DE. 1998b. A lognormal distribution for time spent showering. Risk Anal 18(1):333-5

Burmaster DE. 1998c. Lognormal distributions for total water intake and tap water intake by pregnant and lactating women in the United States. Risk Anal 18(2):215-9

Burmaster DE and Crouch EAC. 1997. Lognormal distributions for body weight as a function of age for males and females in the United States, 1976–1980. Risk Anal 17(4):499-505

Burmaster DE and Murray DM 1998. A trivariate distribution for the height, weight, and fat of adult men. Risk Anal 18(4):385-9

Calabrese EJ and Stanek EJ. 1992. What proportion of household dust is derived from outdoor soil? J Soil Contam 1(3):253-63

Calabrese EJ, Stanek EJ, James RC, *et al.* 1997. Soil ingestion: A concern for acute toxicity in children. Environ Health Perspect 105:1354-8

CARB (State of California Air Resources Board). 1991a. Study of Children's Activity Patterns: Final Report. Sacramento, CA, USA

CARB (State of California Air Resources Board). 1991b. Activity Patterns of California Residents: Report. Sacramento, CA, USA

CARB (State of California Air Resources Board). 2000a. Assorted Databases and Research Results Available Through CARB Web Page, www.arb.ca.gov

CARB (State of California Air Resources Board). 2000b. Air Toxics "Hot Spots" Program Risk Assessment Guidelines Part IV: Technical Support Document for Exposure Assessment and Stochastic Analysis. Report dated October 27, 2000. Available at www.oehha.org/air/hot_spots/finalStoc.html

Carey M. 1988. Occupational tenure in 1987: Many workers have remained in their fields. Monthly Labor Review October, pp 3-12

Carey M. 1990. Occupational tenure, employer tenure, and occupational mobility. Occup Outlook Qtr Summer, pp. 55-60

CCHT (Canadian Centre for Housing Technology). 2001. http://www.ccht-cctr.gc.ca/ccht/e/index.htm

CDC (Centers for Disease Control and Prevention). 2001a. National Health and Nutrition Examination Survey. Available at www.cdc.gov/nchs/nhanes.htm

CDC (Centers for Disease Control and Prevention). 2001b. The National Exposure Report Card: Measuring Exposure of the U.S. Population to Toxic Substances. Available at www.cdc.gov/nceh/dls/one_pagers_pdfs/ reportcard_op.pdf

Chance GW and Harmsen E. 1998. Children are different: Environmental contaminants and children's health. Can J Public Health 89(Suppl. 1):S9-S13

Clayton A and Perritt R. 1993. Data Base Development and Data Analyses for California Indoor Exposure Studies, Volumes I and II. NTIS No. PB94-106903. Prepared for California Air Resources Board, Sacramento, CA, USA

Clinton WJ. 1997. Executive Order 13045—Protection of Children from Environmental Health Risks and Safety Risks. 62 Fed Reg 78, 19883-8

Clobes AL, Ananth GP, Hood AL, *et al.* 1992. Human activities as sources of volatile organic compounds in residential environments. In: Sources of Indoor Air Contaminants: Characterizing Emissions and Health Impacts. Annals NY Acad Sciences 641:79-86

Cohen Hubal EA, Sheldon LS, Burke JM, *et al.* 2000a. Children's exposure assessment: A review of factors influencing children's exposure, and the data available to characterize and assess that exposure. Environ Health Perspect 108(6):475-86

Cohen Hubal EA, Sheldon LS, Zufall MJ, *et al.* 2000b. The challenge of assessing children's residential exposure to pesticides. J Exposure Anal Environ Epi 10:638-49

Cronin IV WJ, Oswald EJ, Shelley ML, *et al.* 1995. A trichloroethylene risk assessment using a Monte Carlo analysis of parameter uncertainty in conjunction with physiologically-based pharmacokinetic modeling. Risk Anal 15(5):555-65

Cullen AC and Frey HC. 1999. Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs. Plenum Press, New York, NY, USA

Driver J. 2000. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on August 10, 2000, regarding exposure assessment resources). Infoscientific.com, Manassas, VA, USA

Driver J. 2001. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on March 5, 2001, regarding status of Society for Risk Analysis Residential Exposure Assessment Project). Infoscientific.com, Manassas, VA, USA

Driver JH, Ginevan ME, and Whitmyre GK. 1996. Estimation of dietary exposure to chemicals: A case study illustrating methods of distributional analyses for food consumption data. Risk Anal 16(6):763-71

Duan N and Mage DT. 1997. Combination of direct and indirect approaches for exposure assessment. J Exposure Anal Environ Epi 7(4):439-70

Durkin PR, Rubin L, Withey J, *et al.* 1995. Methods of assessing dermal absorption with emphasis on uptake from contaminated vegetation. Toxicol Indus Health 11(1):63-79

ECETOC (European Centre for Ecotoxicology and Toxicology of Chemicals). 1994. Technical Report No. 58: Assessment of Non-occupational Exposure to Chemicals. Brussels, Belgium

Echols SL, MacIntosh DL, Mammerstrom KA, *et al.* 1999. Temporal variability of microenvironmental time budgets in Maryland. J Exposure Anal Environ Epi 9(5):502-12

Ecology. 1997. Analysis and Selection of Fish Consumption Rates for Washington State Risk Assessments and Risk-based Standards. Draft. Washington State Department of Ecology, Olympia, WA, USA

Engelmann W. 2001. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on March 7, 2001, regarding the status of EPA's THERdbASE exposure database). U.S. Environmental Protection Agency, Las Vegas, NV, USA

ExxonMobil. 2000. Exposure Factors Sourcebook for European Populations, with Focus on UK Data. ExxonMobil Biosciences, Inc, Annandale, NJ, USA

Fehrenbacher C. 2001. Personal communication (telephone conversation and e-mail correspondence with C Petito Boyce, Exponent, Bellevue, WA, on January 29, 2001, and February 16, 2001, regarding exposure assessment resources used by EPA). U.S. Environmental Protection Agency, Washington, DC, USA

Feijtel TCJ, Webb SF, and Matthijs E. 2000. Predictive exposure modeling—a case study with a detergent surfactant. Food Chem Toxicol 38:S43-S50

FEMA (Flavor and Extract Manufacturers Association). 2000. FEMA 1995 Poundage and Technical Effects Update Survey. Washington, DC, USA

Field RW, Smith BJ, Brus CP, *et al.* 1998. Retrospective temporal and spatial mobility of adult Iowa women. Risk Anal 18(5):575-84

Finley BL, Proctor DM, Scott P, *et al.* 1994a. Recommended distributions for exposure factors frequently used in health risk assessment. Risk Anal 14(4):533-53

Finley BL, Scott PK, and Mayhall DA. 1994b. Development of a standard soil-to-skin adherence probability density function for use in Monte Carlo analyses of dermal exposure. Risk Anal 14(4):555-69

Fox M. 2000. CDC to report on chemical exposure in US. Article dated October 16, 2000. Environmental News Network. Available at www.enn.com/extras/ printer-friendly.asp?storyid=32585

Freeman NCG, Ettinger A, Berry M, *et al.* 1997. Hygiene and food-related behaviors associated with blood lead levels of young children from lead-contaminated homes. J Exposure Anal Environ Epi 7(1):103-18

Freeman NCG, Lioy PJ, Pellizzari E, *et al.* 1999. Responses to the Region 5 NHEXAS time/activity diary. J Exposure Anal Environ Epi 9:414-26

Funk LM, Sedman R, Beals JAJ, *et al.* 1998. Quantifying the distribution of inhalation exposure in human populations: 2. Distributions of time spent by adults, adolescents, and children at home, at work, and at school. Risk Anal 18(1):47-56

Gephart LA, Tell JG, and Triemer LR. 1994. Exposure factors manual. J Soil Contam 3(1):47-117

Ginsberg G. 2000. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on November 6, 2000, regarding development of database compiling information on interindividual variability in children in responses to chemicals). Connecticut Department of Public Health, Hartford, CT, USA

Glickman TS. 1986. A methodology for estimating time-of-day variations in the size of a population exposed to risk. Risk Anal 6(3):317-24

Gordon SM, Callahan PJ, Nishioka MG, *et al.* 1999. Residential environmental measurements in the national human exposure assessment survey (NHEXAS) pilot study in Arizona: Preliminary results for pesticides and VOCs. J Exposure Anal Environ Epi 9:456-70

Graeter LJ and Mortensen ME. 1996. Kids are different: Developmental variability in toxicology. Toxicology 111:15-20

Grassman JA. 1996. Obtaining information about susceptibility from the epidemiological literature. Toxicology 111:253-70

Gregory J, Collins D, Davies P, *et al.* 1995. The national dietary and nutritional survey: Children aged 1 $^1/_2$ to 4 $^1/_2$ years. HMSO. (as cited in Hamey and Harris [1999])

Guo Z, Mason MA, Gunn KN, *et al.* 1992. Adsorption and re-emission of ethylbenzene vapor from interior surfaces in an indoor air quality test house. Proceedings of Measurement of Toxic and Related Air Pollutants, Air Waste Management Association, Pittsburgh, PA, USA, pp. 51-6

Hakkinen PJ. 1993. Cleaning and Laundry Products, Human Exposure Assessments. In: Handbook of Hazardous Materials, pp 145-51. Academic Press, Inc., New York, NY, USA

Hakkinen PJ, Kelling CK, and Callender JC. 1991. Exposure assessment of consumer products: Human body weights and total body surface areas to use; and sources of data for specific products. Vet Human Toxicol 33(1):61-5

Hamey PY and Harris CA. 1999. The variation of pesticide residues in fruits and vegetables and the associated assessment of risk. Regul Toxicol Pharmacol 30:S34-S41

Harrington NW, Curry CL, and Price PS. 1995. The MicroExposure® event modeling approach to probabilistic exposure assessment. Paper No 95-TA42.03. In: Proc 88th Annual Meeting of the Air and Waste Management Association, San Antonio, TX, USA

Hattis D. 1996. Variability in susceptibility—how big, how often, for what responses to what agents? Environ Toxicol Pharmacol 2:135-45

Hattis D. 2000a. Human interindividual variability in parameters related to susceptibility for toxic effects. www2.clarku.edu/faculty/dhattis/. Accessed November 29, 2000. Last updated on November 28, 2000

Hattis D. 2000b. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on November 6, 2000, regarding development of databases compiling information on interindividual variability in adults and children in responses to chemicals). Clark University, Worcester, MA, USA

Hattis D and Silver K. 1994. Human interindividual variability—a major source of uncertainty in assessing risks for noncancer health effects. Risk Anal 14(4):421-31

Hattis D, Erdreich L, and Ballew M. 1987. Human variability in susceptibility to toxic chemicals—a preliminary analysis of pharmacokinetic data from normal volunteers. Risk Anal 7(4):415-26

Hattis D, Banati P, and Goble R. 1999a. Distributions of individual susceptibility among humans for toxic effects—how much protection does the traditional tenfold factor provide for what fraction of which kinds of chemicals and effects? Ann NY Acad Sci 895:286-316

Hattis D, Banati P, Goble R, *et al.* 1999b. Human interindividual variability in parameters related to health risks. Risk Anal 19(4):711-26

Hertwich EG, McKone TE, and Pease WS. 1999. Parameter uncertainty and variability in evaluative fate and exposure models. Risk Anal 19(6):1193-204

Hill MS. 1985. Patterns of time use. In: Juster FT and Stafford FP (eds), Time, Goods, and Well-Being, pp 133-66. University of Michigan, Survey Research Center, Institute for Social Research, Ann Arbor, MI. (not seen, as cited in USEPA [1997a])

Holmes KK, Shirai JH, Richter KY, *et al.* 1999. Field measurement of dermal soil loadings in occupational and recreational activities. Environ Res 80:148-57

IATUR (International Association for Time Use Research). 2000. IATUR Web page. Available at www.stmarys.ca/partners/iatur/iatur.htm

ICRP (International Commission on Radiological Protection). 1975. Report of the Task Group on Reference Man, ICRP Publication No. 23. Pergamon Press, NY, NY, USA

ILSI (International Life Sciences Institute). 1998. Aggregate Exposure Assessment. Available at http://www.ilsi.org/rsiaggexp.pdf. Accessed on August 10, 2000. International Life Sciences Institute, Washington, DC, USA

Israeli M and Nelson CB. 1992. Distribution and expected time of residence for US households. Risk Anal 12(1):65-72

Jacobs HL, Kahn HD, Stralka KA, *et al.* 1998. Estimates of per capita fish consumption in the US based on the continuing survey of food intake by individuals (CSFII). Risk Anal 18(3):283-291

James IR and Knuiman MW (as cited in Burmaster 1998b). An application of Bayes methodology to the analysis of diary records from a water use study. J Am Stat Assoc 832(399):705-11

Johnson T and Capel J. 1992. A Monte Carlo Approach to Simulating Residential Occupancy Periods and its Application to the General US Population. U.S. Environmental Protection Agency, Office of Air Quality and Standards, Research Triangle Park, NC, USA

Kerger BD, Schmidt CE, and Paustenbach DJ. 2000. Assessment of airborne exposure to trihalomethanes from tap water in residential showers and baths. Risk Anal 20(5):637-51

Kissel J, Richter K, and Fenske R. 1996. Field measurements of dermal soil loading attributable to various activities: Implications for exposure assessment. Risk Anal 16(1):116-25

Klepeis NE, Nelson WC, Ott WR, *et al.* 2001. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. J Exposure Anal Environ Epi. 11(3):231-52

Koontz MD, Rector HE, Fortmann RC, *et al.* 1988. Preliminary Experiments in a Research House to Investigate Contaminant Migration in Indoor Air. EPA 560/5-88-004. U.S. Environmental Protection Agency, Office of Pesticides and Toxic Substances, Washington DC, USA, as cited in USEPA 1997a

Krieger RI, Bernard CE, Dinoff TM, *et al.* 2001. Biomonitoring of persons exposed to insecticides used in residences. Annals Occup Hygiene 45(suppl 1):S143-53

Kuempel ED, Tran CL, Bailer AJ, *et al.* 2001. Methodological issues of using observational human data in lung dosimetry models for particulates. Sci Total Environ. 274(1-3):67-77

Law PL, Lioy PJ, Zelenka MP, *et al.* 1997. Evaluation of a probabilistic exposure model applied to carbon monoxide (pNEM/CO) using Denver personal exposure monitoring data. Air Waste Manage Assoc 47:491-500

Layton DW. 1993. Metabolically consistent breathing rates for use in dose assessments. Health Physics 64(1):23-36

Leighton T. 2001. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on March 1, 2001, regarding exposure assessment resources used by EPA, including the Pesticide Handler's Exposure Database). U.S. Environmental Protection Agency, Washington, DC, USA

Loretz L. 2000. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on November 6, 2000, regarding trade association research into exposure assessment issues). Cosmetic, Toiletries, and Fragrance Association, Washington, DC, USA

Lyman WJ, Reehl WF, and Rosenblatt DH. 1990. Handbook of Chemical Property Estimation Methods. American Chemical Society, Washington, DC, USA

MacIntosh DL, Kabiru CW, and Ryan PB. 2001. Longitudinal investigation of dietary exposure to selected pesticides. Environ Health Persp 109(2):145-50

MacPherson N. 2000. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on November 13, 2000, regarding databases for consumer products and activity patterns of product use). Lego

Marquart H, Maidment S, McClaflin JL, *et al.* 2001. Harmonization of future needs for dermal exposure assessment and modeling: A workshop report. Appl Occup Environ Hyg 16(2):218-27

McCurdy T. 2000a. Conceptual basis for multi-route intake dose modeling using an energy expenditure approach. J Exposure Anal Environ Epi 10:86-97

McCurdy T. 2000b. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on November 15, 2000, regarding CHADS database and other exposure assessment resources). U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

McCurdy T, Glen G, Smith L, *et al.* 2000. The national exposure research laboratory's consolidated human activity database. J Exposure Anal Environ Epi 10(6 Pt 1):566-78

MDH (Minnesota Department of Health). 2000. Comparative Risks of Multiple Chemical Exposures. Final Report for the Legislative Commission on Minnesota Resources. Available at www.health.state.mn.us/dirs/eh/esa/hra/children/lcmrept.pdf

Mekel OC, and Fehr R. 2001. Use of probabilistic methods in exposure assessment in Germany. Annals Occup Hygiene 45(Suppl. 1):S65-67

Montalbano A. 2001. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on January 22, 2001, regarding exposure assessment research conducted and compiled by the Consumer Product Safety Commission). Consumer Product Safety Commission, Bethesda, MD, USA

Morgan DA. 2001. Occupational exposure databases and their application for the next millennium: Symposium framework and workshop introduction. Appl Occup Environ Hyg 16(2):111-4

Murray DM. 1997. Residential house and zone volumes in the United States: Empirical and estimated parametric distributions. Risk Anal 17(4):439-46

Murray DM and Burmaster DE. 1994. Estimated distributions for average daily consumption of total and self-caught fish for adults in Michigan angler households. Risk Anal 14(4):513-9

Murray DM and Burmaster DE. 1995. Residential air exchange rates in the United States: Empirical and estimated parametric distributions by season and climatic region. Risk Anal 15(4):459-465

Novigen Sciences, Inc. 2000. Background document for the sessions: Dietary Exposure Evaluation Model (DEEM™) and DEEM™ Decompositing Procedure and Software. U.S. Environmental Protection Agency, Office of Pesticide Programs, Washington, DC, USA

ODEQ (Oregon Department of Environmental Quality). 1998. Guidance for Use of Probabilistic Analysis in Human Health Risk Assessments. Interim Final Oregon Department of Environmental Quality, Waste Management and Cleanup Division, Portland, OR, USA

Ott W, Thomas J, Mage D, *et al.* 1988. Validation of the simulation of human activity and pollutant exposure (shape) model using paired days from the Denver, CO, carbon monoxide field study. Atmos Environ 22(10):2101-13

Owen BA. 1990. Literature-derived absorption coefficients for 39 chemicals via oral and inhalation routes of exposure. Regul Toxicol Pharmacol 11:237-52

Parkin RT and Balbus JM. 2000. Variations in concepts of "susceptibility" in risk assessment. Risk Anal 20(5):603-11

Pastino GM, Yap WY, and Carroquino M. 2000. Human variability and susceptibility to trichloroethylene. Environ Health Perspect 108(2):201-14

Paustenbach DJ. 2000. The practice of exposure assessment: A state-of-the-art review. J Toxicol Environ Health Part B 3:179-291

Pellizzari E, Lioy P, Quackenboss J, *et al.* 1995. Population-based exposure measurements in EPA region 5: A phase I field study in support of the national human exposure assessment survey. J Exposure Anal Environ Epi 5(3):327-58

Perlin S. 2000. National Health and Nutrition Examination Survey (NHANES). Presentation materials. National Center for Environmental Assessment, Office of Research and Development, Washington, DC, USA

Petersen BJ, Barraj LM, Muenz LR, *et al.* 1994. An alternative approach to dietary exposure assessment. Risk Anal 14(6):913-6

Petito Boyce C. 1998. Comparison of approaches for developing distributions for carcinogenic slope factors. HumanEcol Risk Assess 4(2):527-77

Price PS, Curry CL, Goodrum PE, *et al* 1996. Monte Carlo modeling of time-dependent exposures using a microexposure event approach. Risk Anal 16(3):339-48

Price PS, Scott PK, Wilson ND, *et al.* 1998. An empirical approach for deriving information on total duration of exposure from information on historical exposure. Risk Anal 18(5):611-9

Price PS, Young JS, and Chaisson CF. 2000. Assessing Aggregate and Cumulative Pesticide Risks Using LifeLine™ Version 1.0. www.epa.gov/oscpmont/sap/2000/ september/ sap_report_lifeline. Accessed on October 16, 2000. Report dated August 31, 2000. Report prepared for U.S. Environmental Protection Agency, Science Advisory Board, Washington, DC, USA

Price PS, Young JS, and Chaisson CF. 2001. Assessing aggregate and cumulative pesticide risks using a probabilistic model. Annals Occup Hygiene 45(Suppl. 1):S131-42

Qiao GL, Brooks JD, and Riviere JE. 1997. Pentachlorophenol dermal absorption and disposition from soil in swine: Effects of occlusion and skin microorganism inhibition. Toxicol Appl Pharmacol 147:234-46

Reed KJ, Jimenez M, Freeman NCG, *et al.* 1999. Quantification of children's hand and mouthing activities through a videotaping methodology. J Exposure Anal Environ Epi 9(5):513-20

RIFM. 2000. Special issue IX: Monographs on fragrance raw materials. Food Chem Toxicol 38 (Suppl. 3)

Riley DM, Fischhoff B, Small MJ, *et al.* 2001. Evaluating the effectiveness of risk-reduction strategies for consumer chemical products. Risk Anal 21(2):357-69

Robertson GL, Lebowitz MD, O'Rourke MK, *et al.* 1999. The national human exposure assessment survey (NHEXAS) study in Arizona—introduction and preliminary results. J Exposure Anal Environ Epi. 9:427-34

Robinson JP and Thomas J. 1991. Time Spent in Activities, Locations, and Microenvironments: A California-National Comparison Project Report. U.S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory, Las Vegas, NV, USA

Ruby MV, Schoof R, Brattin W, *et al.* 1999. Advances in evaluating the oral bioavailability of inorganics in soil for use in human health risk assessment. Environ Sci Technol 33(21):3697-705, USA

Ruffle B, Burmaster DE, Anderson PD, *et al.* 1994. Lognormal distributions for fish consumption by the general US population. Risk Anal 14:395-404

Rusconi F, Catagento M, Gagliardi L, *et al.* 1994. Reference values for respiratory rate in the first 3 years of life. Pediatrics 94(3):350-5. (as cited in Exxon Mobil 2000)

Scanlon KA, MacIntosh DL, Hammerstrom KA, *et al.* 1999. A longitudinal investigation of solid-food based dietary exposure to selected elements. J Exposure Anal Environ Epi 9:485-93

Scheringer M, Vogl T, von Grote J, *et al.* 2001. Scenario-based risk assessment of multi-use chemicals: Application to solvents. Risk Anal 21(3):481-97

Scorecard. 2000. Scorecard home page. Available at www.scorecard.org

SDA (Soap and Detergent Association). 2000. www.sdahq.org/about/background.html

Sedman R, Funk LM, and Fountain R. 1998. Distribution of residence duration in owner occupied housing. J Exposure Anal Environ Epi 8(1):51-8

Sexton K and Adgate JL. 1999. Looking at environmental justice from an environmental health perspective. J Exposure Anal Environ Epi 9:3-8

Sexton K, Wagener DK, Selevan SG, *et al.* 1994. An inventory of human exposure-related data bases. J Exposure Anal Environ Epi 4(1):95-109

Sexton K, Kleffman DE, and Callahan MA. 1995a. An introduction to the national human exposure assessment survey (NHEXAS) and related Phase I field studies. J Exposure Anal Environ Epi 5(3):229-32

Sexton K, Callahan MA, Bryan EF, *et al.* 1995b. Informed decision about protecting and promoting public health: Rationale for a national human exposure assessment survey. J Exposure Anal Environ Epi 5(3):233-56

Shurdut BA, Barraj L, and Francis M. 1998. Aggregate exposures under the food quality protection act: An approach using chlorpyrifos. Regul Toxicol Pharmacol 28:165-77

Silvers A, Florence BT, Rourke DL, *et al.* 1994. How children spend their time: A sample survey for use in exposure and risk assessments. Risk Anal 14(6):931-44

Smith AE, Ryan PB, and Evans JS. 1992. The effect of neglecting correlations when propagating uncertainty and estimating the population distribution of risk. Risk Anal 12(4):467-74

Sparks LE, Tichenor BA, White JB, *et al.* 1991. Comparison of data from an IAQ test house with predictions of an IAQ computer model. Indoor Air, 1, 577-59

Sparks LE, Guo Z, Chang J, *et al.* 1999. VOC emissions from latex paint—Part 2. Test house studies and IAQ modeling. Indoor Air 99(9):18-25

Stanek III EJ and Calabrese EJ. 1992. Soil ingestion in children: outdoor soil or indoor dust? J Soil Contam 1(1):1-28

Stanek III EJ and Calabrese EJ. 2000. Daily soil ingestion estimates for children at a Superfund site. Risk Anal 20(5):627-35

Stanek III EJ, Calabrese EJ, Mundt K, *et al.* 1998. Prevalence of soil mouthing/ingestion among healthy children aged 1 to 6. J Soil Contam 7(2):227-42

Sweeney LM, Tyler TR, Kirman CR, *et al.* 2001. Proposed occupational exposure limits for select ethylene glycol ethers using PBPK models and Monte Carlo simulations. Toxicological Sciences 62(1):124-39

Syracuse Research Corporation. 2001. Environmental Fate Data Base. Available at http://esc.syrres.com/efdb.htm

Thomas T. 2001. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on January 25, 2001, regarding exposure assessment research conducted and compiled by the Consumer Product Safety Commission). Consumer Product Safety Commission, Bethesda, MD, USA

TIIE (Texas Institute for the Indoor Environment). 2001. University of Texas. Available at http://civil.ce.utexas.edu/prof/corsi/prospectus.html

Timmer SG, Eccles J, and O'Brien K. 1985. How children use time. In: Juster FT and Stafford FP (eds), Time, Goods, and Well-Being, pp 353-80. University of Michigan, Survey Research Center, Institute for Social Research, Ann Arbor, MI, USA

Tomerlin JR, Berry MR, Tran NL, *et al.* 1997. Development of a dietary exposure potential model for evaluating dietary exposure to chemical residues in food. J Exposure Anal Environ Epi 7(1):81-101

Travis CC and Arms AD. 1988. Bioconcentration of organics in beef, milk, and vegetation. Environ Sci Technol 22(2):271-4

United Nations. 2000. Time-use Surveys. www.un.org/Depts/unsd/timeuse/ tusresource.htm. Accessed on November 29, 2000. United Nations Statistics Division, Brussels, Belgium

University of Michigan. 2000. Child Development Supplement to the Panel Study oflncome Dynamics. Available at www.isr.umich.edu/src/child-development/ home.html

U.S. Bureau of the Census. 1993. American Housing Survey for the United States in 1991. U.S. Government Printing Office, Washington, DC, USA

U.S. Bureau of the Census. 1995. Statistical Abstracts of the United States. U.S. Government Printing Office, Washington, DC, USA

USDOE (U.S. Department of Energy). 1995. Housing Characteristics 1993, Residential Energy Consumption Survey (RECS). Report No EOE/EIA-0314 (93). Energy Information Administration, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1980. Water Quality Criteria Documents; availability. 45 Fed Reg 231:79318–79

USEPA (U.S. Environmental Protection Agency). 1985. Development of Statistical Distributions or Ranges of Standard Factors used in Exposure Assessments. EPA/600/8-85/010. Office of Research and Development, Office of Health and Environmental Assessment, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1988. Reference Physiological Parameters in Pharmacokinetic Modeling. EPA/600/6-88/004. Final Reports. Prepared for the US Department of Energy and the U.S. Environmental Protection Agency by Office of Risk Analysis, Oak Ridge National Laboratory, Oak Ridge, TN, USA

USEPA (U.S. Environmental Protection Agency). 1989a. Interim Final Guidance for Soil Ingestion. OSWER Directive 9850.4. Office of Solid Waste and Emergency Response, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1989b. Exposure Factors Handbook. EPA/600/8-89/043. Office of Research and Development, Office of Health and Environmental Assessment, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1989c. Risk Assessment Guidance for Superfund. Volume I. Human Health Evaluation Manual (Part A). Interim Final. EPA/540/1-89/002. Office of Emergency and Remedial Response, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1991a. National Primary Drinking Water Regulation; final rule. 56 Fed Reg 20, 3526–3597. January 30, 1991

USEPA (U.S. Environmental Protection Agency). 1991b. Human Health Evaluation Manual, Supplemental Guidance: Standard default exposure factors. OSWER Directive 9285.6-03. Office of Solid Waste and Emergency Response, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1992a. Dermal Exposure Assessment: Principles and Applications. EPA/600/ 8-9-91. Office of Research and Development, Office of Health and Environmental Assessment, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1992b. Absorption Factors for Superfund Chemicals. Environmental Criteria and Assessment Office, Office of Health and Environmental Assessment, Cincinnati, OH, USA

USEPA (U.S. Environmental Protection Agency). 1992c. Internal memorandum dated February 26, 1992, from FH Habicht II, Deputy Administrator to Assistant and Regional Administrators, regarding guidance on risk characterization for risk managers and risk characterization for risk managers and risk assessors. Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1995a. Assessing Dermal Exposure from Soil. EPA/903-K-95-003. Hazardous Waste Management Division, Philadelphia, PA, USA

USEPA (U.S. Environmental Protection Agency). 1995b. Internal memorandum dated March 21, 1995, from CM Browner, EPA Administrator, regarding EPA risk characterization program. Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1996a. Descriptive Statistics Tables from a Detailed Analysis of the National Human Activity Pattern Survey (NHAPS) data. EPA/600/R-96/148. Office of Research and Development, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1996b. Analysis of the National Human Activity Pattern Survey (NHAPS) Respondents from a Standpoint of Exposure Assessment: Percentage of Time Spent, Duration, and Frequency of Occurrence for Selected Microen-

vironments by Gender, Age, Time-of-Day, Day-of-Week, Season, and US Census Region. Final Report. EPA/600/R-96/074. National Exposure Research Laboratory, Office of Research and Development, Las Vegas, NV, USA

USEPA (U.S. Environmental Protection Agency). 1996c. Proposed guidelines for carcinogenic risk assessment. 61 Fed Reg 79, 17960–18011

USEPA (U.S. Environmental Protection Agency). 1997a. Exposure Factors Handbook. EPA/600/P-95/002F. Office of Research and Development, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1997b. Guiding Principles for Monte Carlo Analysis. EPA/630/R-97/001. Risk Assessment Forum, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1997c. Guidance on Cumulative Risk Assessment. Part 1. Planning and Scoping. Document dated July 3, 1997. Available at www.epa.gov/ORD/spc/cumrisk2.htm

USEPA (U.S. Environmental Protection Agency). 1997d. Standard Operating Procedures (SOPs) for Residential Exposure Assessments. Draft. Report dated December 19, 1997. Available at www.epa.gov/pesticides/trac/science/trac6a05.pdf

USEPA (U.S. Environmental Protection Agency). 1999a. Sociodemographic Data Used for Identifying Potentially Highly Exposed Populations. EPA/600/R-99/060. Office of Research and Development, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1999b. Exposure Data Requirements for Assessing Risks from Pesticide Exposure of Children. Draft. Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 1999c. Overview of Issues Related to the Standard Operating Procedures for Residential Exposure Assessment. Available at www.epa.gov/scipoly/sup/1999/september/resid.pdf

USEPA (U.S. Environmental Protection Agency). 1999d. Report of the Workshop on Selecting Input Distributions for Probabilistic Assessments. EPA/630/R-98/004. Available at www.epa.gov/ncea/input.htm

USEPA (U.S. Environmental Protection Agency). 1999e. Exposure Factors Handbook (CD-ROM). EPA/600/C-99/001. Office of Research and Development. Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 2000a. Child-specific Exposure Factors Handbook. Report dated June 2000. Available at www.epa.gov/ncea/ csefh2.htm

USEPA (U.S. Environmental Protection Agency). 2000b. Total Human Exposure Risk Database and Advanced Simulation Environment (THERdbASE). Available at www.epa.gov/heasd/therd-home.htm.

USEPA (U.S. Environmental Protection Agency). 2000c. Exposure Assessment Tools and Models. Available at www.epa.gov/opptintr/ exposure

USEPA (U.S. Environmental Protection Agency). 2000d. Consolidated Human Activities Database. Available at everest.sdc-moses.com/edr/chad/index.html

USEPA (U.S. Environmental Protection Agency). 2000e. EIMS—Environmental Information Management System of the USEPA home page. Available at www.epa.gov/eims/eimshome.html

USEPA (U.S. Environmental Protection Agency). 2000f. Methodology for Deriving Ambient Water Quality Criteria for the Protection of Human Health. Office of Water, Washington, DC, USA

USEPA (U.S. Environmental Protection Agency). 2000g. Available Information on Assessing Exposure from Pesticides in Food. Available at www.epa.gov/fedrgstr/EPA_PEST/2000/July/Day-12/6061.pdf

USEPA (U.S. Environmental Protection Agency). 2000h. Technical Workshop on Issues Associated with Considering Developmental Changes in Behavior and Anatomy When Assessing Exposure to Children. 65 Fed Reg 131, 41983

USEPA (U.S. Environmental Protection Agency). 2000i. Dietary Model Estimates Total Ingestion of DBPs and Identifies Contributing Food Groups and Items. Available at www.epa.gov/nerl/research/goal2.htm

USEPA (U.S. Environmental Protection Agency). 2000j. The National Human Exposure Assessment Survey (NHEXAS) Pilot Studies: Early findings. Available at www.epa.gov/nerl/research/goal8.htm

USEPA (U.S. Environmental Protection Agency). 2000k. National Human Exposure Assessment Survey home page. Available at www.epa.gov/nerl/nhexas.htm

USEPA (U.S. Environmental Protection Agency). 2000l. Options for Development of Parametric Probability Distributions for Exposure Factors. EPA/600/R-00/058. Available at www.epa.gov/ncea/paramprob4ef.htm

USEPA (U.S. Environmental Protection Agency). 2000m. Summary Report of the Technical Workshop on Issues Associated with Considering Developmental Changes in Behavior and Anatomy When Assessing Exposure to Children. EPA/630/R-00/005. Available at www.epa.gov/ncea/rat/wrkshops.htm

USEPA (U.S. Environmental Protection Agency). 2001a. Envirofacts Data Warehouse and Applications. Available at http://www.epa.gov/enviro/index_java.html

USEPA (U.S. Environmental Protection Agency). 2001b. Indoor Environment Management Branch: Facilities and Capabilities website. Available at http://www.epa.gov/appcdwww/crb/iemb/facilities.htm

USEPA (U.S. Environmental Protection Agency). 2001c. Pesticide Handler Exposure Database website. Available at www://epa.gov/opptintr/cbep/actlocal/phed.htm

USEPA (U.S. Environmental Protection Agency). 2001d. Exposure Factors Handbook website. Available at http://www.epa.gov/ncea/exposfac.htm

USEPA (U.S. Environmental Protection Agency). 2001e. Exposure Factors Program website. Available at http://cfpub.epa.gov/ncea/cfm/efprog.cfm?ActType=default

USDA (US Department of Agriculture). 2000. What We Eat in America, 1994–96, 1998 website. Available at www.barc.usda.gov/ bhnrc/foodsurvey/pdf/csfii98.pdf

Van Dyke MV, LaMontagne AD, Martyny JW, *et al.* 2001. Development of an exposure database and surveillance system for use by practicing OSH professionals. Appl Occup Environ Hyg 16(2):135-43

Van Veen MP. 1996. A general model for exposure and uptake from consumer products. Risk Anal 16(3):331-8

Van Veen MP, Van Engelen JG, and van Raaji MT. 2001. Crossing the river stone by stone: approaches for residential risk assessment for consumers. Annals Occup Hygiene 45(Suppl. 1):S107-18

Versar. 1990. Database of Perfluorocarbon Tracer (PFT) Ventilation Measurements: Description and User's Manual. U.S. Environmental Protection Agency, Office of Toxic Substances, Washington, DC, USA

Verschueren K. 1996. Handbook of Environmental Data on Organic Chemicals. Third edition. Van Nostrand Reinhold, New York, NY, USA

Wallace L. 1993. A decade of studies of human exposure: What have we learned? Risk Anal 13(2):135-39

Wallace LA, Duan N, and Ziegenfus R. 1994. Can long-term exposure distributions be predicted from short-term measurements? Risk Anal 14(1):75-85

Waller LA, Louis TA, and Carlin BP. 1999. Environmental justice and statistical summaries of differences in exposure distributions. J Exposure Anal Environ Epi 9:56-65

Weaver VM, Buckley TJ, and Groopman JD. 1998. Approaches to environmental exposure assessment in children. Environ Health Perspect 106(3):827-38

Weegels MF, and van Veen MP. 2001. Variation of consumer contact with household products: A preliminary investigation. Risk Anal 21(3):499-511

Westat. 1987a. Household Solvent Products–A National Usage Survey. Final Report. NTIS No. PB88-132881. Prepared for the USEPA, Washington, DC. Westat, Inc., Rockville, MD, USA

Westat. 1987b. National Usage Survey of Household Cleaning Products. Prepared for the USEPA, Office of Toxic Substances and Office of Pesticides and Toxic Substances, Washington, DC. Westat, Inc., Rockville, MD, USA

Westat. 1987c. National Household Survey of Interior Painters. Prepared for the USEPA, Office of Toxic Substances and Office of Pesticides and Toxic Substances, Washington, DC. Westat, Inc., Rockville, MD, USA

Whitmyre GK, Driver JH, Ginevan ME, *et al.* 1992a. Human exposure assessment I: Understanding the uncertainties. Toxicol Indust Health 8(5):297-320

Whitmyre GK, Driver JH, Ginevan ME, *et al.* 1992b. Human exposure assessment II: Quantifying and reducing the uncertainties. Toxicol Indust Health 8(5):321-42

Wiley JA, Robinson JP, Cheng Y-T, *et al.* 1991. Study of Children's Activity Patterns. NTIS No. PB94-106903. California Environmental Protection Agency, Air Resources Board Research Division, Sacramento; CA, USA

Williams M. 2000. Personal communication (telephone conversation with C Petito Boyce, Exponent, Bellevue, WA, on November 6, 2000, regarding trade association research into exposure assessment issues). Flavor and Extract Manufacturers Association, Washington, DC, USA

Witschi HP and Hakkinen PJ. 1984. The role of toxicological interactions in lung injury. Environ Health Perspect 55:139-48

Wong EY, Shirai JH, Garlock TJ, *et al.* 2000. Survey of selected activities relevant to exposures to soils. Bull Environ Contam Toxicol 65:443-50

Wu C and Schaum J. 2000. Exposure assessment of trichloroethylene. Environ Health Perspect 108(2):359–363

Zartarian V. 1999. Exposure assessment program updates: USEPA's National Exposure Research Laboratory (NERL). International Society of Exposure Analysis Newsletter 7.2:2-4

Zartarian VG and Leckie JO. 1998. Dermal exposure: The missing link. Environ Sci Technol 32:134A-137A

Zartarian VG, Ott WR, and Duan N. 1997a. A quantitative definition of exposure and related concepts. J Exposure Anal Environ Epi 7(4):411-37

Zartarian VG, Ferguson AC, and Leckie JO. 1997b. Quantified dermal activity data from a four-child pilot field study. J Exposure Anal Environ Epi 7(4):543-52