Epidemiology and Quantitative Risk Assessment: A Bridge from Science to Policy

Irva Hertz-Picciotto, PhD

ABSTRACT

Quantitative risk assessment provides formalized scientific input to regulatory agencies that set occupational and environmental standards for potentially toxic exposures. Current practice relies heavily on statistical extrapolation from high-dose animal studies. Human data obviate the need for interspecies extrapolation and reduce the range of high-to-low dose extrapolation. This paper proposes a framework for classifying individual epidemiologic studies as to their adequacy for use in dose–response extrapolation.

The framework considers five criteria: (1) a stable positive association with an adverse health outcome; (2) high overall study quality; (3) no substantial confounding; (4) quantitative exposure assessment for individuals; (5) evidence of a dose–response relationship. With these criteria, studies can be categorized as (1) suitable to serve as a basis for extrapolation; (2) inadequate to be the basis for direct extrapolation but appropriate to use for evaluating the plausibility of animal-derived risk estimates; or (3) useful only for hazard identification, not for dose–response assessment. Methods for using studies in the first two categories are briefly described. The emphasis is not on establishing rigid rules, but rather on ensuring a consistent, reliable process that makes optimum use of available data. (*Am J Public Health.* 1995;85:484–491)

Introduction

Quantitative risk assessment provides the formalized scientific input to agencies that set occupational or environmental standards for regulating toxic exposures. As currently practiced, risk assessment relies primarily on animal data coupled with statistical extrapolation models. Use of epidemiological data for quantitative risk assessment has received scant attention in regulatory documents, while in the peer-reviewed literature, some chemical risk assessments using human data have been published but rarely has a discussion of principles appeared. A US Environmental Protection Agency program on improving health risk assessment essentially ignored epidemiologic data.[1] Several documents address criteria for qualitatively evaluating the weight of evidence from epidemiological studies, but they offer little guidance for incorporating epidemiological data into dose–response assessment.[2,3] Indeed, in 1980, the Occupational Safety and Health Administration declared that "epidemiologic studies, if accompanied by reliable data on exposure levels, may be useful in priority-setting, but are rarely, if ever, sensitive enough to be useful in setting acceptable levels of exposure."[2] As occupational epidemiology studies have mushroomed and the methodology has improved, this assessment is outdated. Ways to incorporate epidemiological data into the different stages of risk assessment have been described.[4,5] In practice, use of epidemiology has been inconsistent: poor data have been used in dose–response assessment and excellent studies have been ignored.

The purpose of this paper is to increase scientific rigor in quantitative risk assessment by proposing a framework for standardized classification of individual epidemiological studies as to their adequacy for use in dose–response extrapolation. This framework includes criteria that address both validity and utility and provides two possible roles for epidemiological data in dose–response assessment. Implementation of these guidelines would ensure appropriate use of epidemiological data, reduce uncertainty in risk estimates, and contribute to a more rational decision-making process. Although I focus on carcinogenic response, the principles apply, with only minor modification, to other health endpoints. The use of epidemiological data in hazard identification is not addressed, nor is reconciliation of conflicting studies or methods to combine studies in dose–response assessment.

What Is Risk Assessment?

The National Research Council has outlined four steps in risk assessment.[6] Hazard identification evaluates whether previous research indicates that the exposure may harm human health. Exposure assessment identifies the specific agents, determines the route of human exposure, and quantifies the amount and duration of exposure. Dose–response assessment uses published data to relate dose to adverse health response and then extrapolates to a (usually) lower environmental exposure. Finally, risk characterization combines exposure assessment with dose–

The author is with the Department of Epidemiology, University of North Carolina School of Public Health, Chapel Hill.

Requests for reprints should be sent to Irva Hertz-Picciotto, PhD, Department of Epidemiology, University of North Carolina School of Public Health, CB #7400, McGavran-Greenberg Hall, Chapel Hill, NC 27599-7400.

Editor's Note. See related editorial by Shore (p 474) in this issue.

response assessment to quantify, for a defined population, the risks predicted to result from the given exposure.

Hazard identification asks, Do published studies suggest that exposure to X will increase the risk of disease Y? However, the criteria for scientific consensus on causality are not necessarily the appropriate criteria for regulating. Analogously, criteria for establishing causality in a court of law also differ from scientific notions.[7] Recognition that the "rules of evidence" are different is a prerequisite for meaningful dialogue between scientists and the regulating community. If exposure is widespread and the consequences serious, a need for primary prevention may suggest that even a moderate degree of evidence justifies regulatory action. On the other hand, if the probability of human exposure is low and the adverse health effects uncertain, then the best policy may be collection of improved data.

While public health scientists often estimate risks, risk assessment poses a more specific question. Epidemiologists often address the question, What is the risk of disease Y in the presence of agent X relative to the risk of disease Y in the absence of X? The risk assessor asks, How many excess cases of disease Y will occur in a population of size Z due to exposure to agent X at dose level D? These questions differ on two counts at least. First, exposure is defined in quantitative terms for risk assessment, while epidemiologic investigation frequently relies on qualitative categories. Second, the outcome is defined as added risk (i.e., absolute risk, excess risk, additive risk, or risk difference) rather than the commonly used relative risk. Further, risk assessments may specify the exact population to which results are extrapolated, for example, a certain community located in a specific geographic locale.

The remainder of this paper is concerned with the dose–response assessment stage. It is assumed that a carcinogenic hazard has been identified and that epidemiological data were used, if appropriate, for such identification.

The Advantages of Epidemiological Data

Quantitative risk assessment emphasizes data from animal experiments.[2,8] This practice is defended on the grounds that epidemiological studies are too insensitive, involve uncertain measures of exposure as compared with well-controlled animal dosing, and are hampered by confounding and other biases—in short, that epidemiological studies fail to provide a controlled, randomized, experimental situation. Nevertheless, human data alone address the question of interest. The strength and validity of epidemiological evidence has been underestimated, and the ability to control confounding has been ignored. The net advantage of using well-conducted human studies far outweighs the disadvantages.

First, the magnitude of error is likely to be greater when animal data are used. The uncertainty stemming from interspecies extrapolation is far larger than the uncertainty resulting from uncontrolled bias or errors in exposure information in epidemiological studies.[9] Extrapolation from animals assumes similar rates of absorption, metabolic pathways, rates for activation or detoxification, and elimination rates. It also requires an assumption about equivalent exposures: should doses be scaled by milligrams per kilogram per day; milligrams per unit surface area per day (estimated as $mg/kg^{2/3}/day$ or $mg/kg^{3/4}/day$, both based on toxicities of anticancer chemotherapeutics[10,11]); or cumulative lifetime milligrams per kilogram? This choice (ignoring other pharmacokinetic considerations) can result in risk estimates that differ by a factor of as much as 10 to 100, and interspecies scaling continues to be controversial.[10,12] Ultimately, differences in breathing rates, organ sizes, basal metabolism, rates of cell turnover, and life spans make comparability difficult to achieve. In comparison with these vast uncertainties, the main uncertainty in human studies—inaccuracies in exposure data—is generally smaller.[9] Other uncertainties in human data, such as confounding, account for errors of a much lower magnitude, often around 10% to 50%, but rarely more than a factor of 2 or 3.[13,14]

A second advantage of human data is a smaller range of extrapolation. For instance, estimated exposures in the occupational study of ethylene dibromide[15] were about two orders of magnitude lower than the doses in animal studies[16,17] (Figure 1); environmental levels were one to two orders of magnitude below the lowest occupational exposures. Typically, use of human data reduces the range of extrapolation.

Third, the exposure experience in animals, although well controlled and measured, is a poor representation of human exposure scenarios. Patterns of variability differ. Exposures in a workplace setting begin in adult life, are intermittent in a way not replicated by animal experiments, and can vary in intensity both within a day and over a lifetime. Although environmental exposures may begin long before adulthood, they are frequently subject to patterns of variation that (similar to occupational exposures) reflect changes in ambient levels and in activities of the individuals. Because it is unclear how to minimize the impact of differences between laboratory and real-world dosing patterns, simplifying assumptions are made for translating animal exposures to human equivalents. This issue was recently reviewed[10] and will not be discussed further.

The context of exposure also differs markedly. Sequestering animals by sex and administering a single chemical in a laboratory has little in common with a scenario of multiple exposures via multiple routes (air, food, water, cosmetics), in which chemicals enter the body through the lungs, the gut, the skin, and so forth, and in which the lives of the organisms involve an intricate web of social interactions and biological cycles. The relevant point is that epidemiological data eliminate the need for many assumptions that ignore context and patterns of exposure in predicting health effects.

Fourth, the genetic diversity and the variability in other endogenous or host factors in the human population will be better represented in a human study than in an animal study. The single strain of rodents could be hypersensitive to the agent being tested, leading to overestimation of human risks, or hyperresistant, leading to underestimation. Furthermore, with most assays testing single chemicals only, the impact of other exogenous exposures on the carcinogenic potency of that substance remains unknown. Because endogenous and other exogenous factors alter susceptibility to disease, the controlled experiment with single strains of one or two species exposed to one chemical has less generalizability than any reasonably sized human study.

Human data also have limitations (discussed below). However, neither epidemiological studies nor animal bioassays can directly assess the levels of risk that are of interest to regulators, that is, increases in risk of 1 in 1 million or even 1 in 1000. To assess such risks requires extrapolation from studies in which a small population, either animal or human, experiences higher exposures.

These arguments should not be construed as an appeal to wait for human data when adequate animal studies demonstrate adverse health effects. The Inter-

TABLE 1—Summary of Classification Framework for the Use of Epidemiological Studies in Quantitative Risk Assessment

	Study Category		
	1	2	3
Use	Can serve as a basis for extrapolation	Can be used to check plausibility of an animal-based risk assessment	Can contribute to the weight-of-evidence determination of whether the agent is a health hazard
Criteria			
1. Moderate to strong positive association present	Necessary	Not necessary, and often this criterion is not met	If met, adds to weight of evidence for a hazard
2. Strong biases ruled out or unlikely	Necessary	Should be met, at least partially	If met, strengthens evidence regarding whether agent is or is not a hazard
3. Confounding controlled or likely to be limited	Necessary	Should be met, at least partially, or limits on confounding should be estimated	
4. Quantification of exposures linked to individuals	Necessary	Some quantification of exposures is needed, even if based on data external to study site	Usually not met
5. Monotonic dose–response relationship	Not necessary but adds certainty to risk estimates	Not necessary	May or may not be met
Summary of requirements	Criteria 1–4 should be met	Two of criteria 1–3 should be met	All other studies

national Agency for Research on Cancer posited that "in the absence of adequate data on humans, it is biologically plausible and prudent to regard agents and mixtures for which there is sufficient evidence of carcinogenicity [i.e., a causal association] in experimental animals as if they presented a carcinogenic risk to humans."[18] This conclusion has been disputed,[19] but in the absence of viable alternatives, it remains an eminently practical guide for decision making in the regulatory sphere. When studies of sufficient quality are available for both humans and animals, the human data are preferable as a basis for extrapolation.

Framework for Standardized Classification of Individual Epidemiological Studies

Description of Framework

Regulatory agencies and risk assessors have promulgated lists of desirable attributes of epidemiological studies but no clear criteria for using epidemiological data in dose–response assessment. The lack of systematic methodology has led to regulatory agencies' dealing with human data on an ad hoc basis, applying different criteria for each chemical. Based on a synthesis of empirical experience in this field, the approach proposed here would reduce uncertainty, enhance consistency, and hence foster credibility.

Epidemiological data can provide input to risk assessment in three ways. For some studies, designated category 1 studies, a dose–response relationship can be derived (usually from an occupational group) and used to set regulatory standards, either occupational or environmental. For other studies, including many null ones, data are inadequate to confidently derive a dose–response relationship but can be used as a check on the plausibility of an animal-based risk assessment. These are designated category 2 studies. Category 3 encompasses those studies that cannot contribute to dose–response assessment but can play a role in hazard identification. Thus a three-tiered evaluation is proposed, based on five criteria (Table 1): (1) a strong or moderate positive association that is statistically stable between cancer of one or more sites and the agent in question; (2) high overall quality (i.e., major biases in selection, follow-up, etc., can be ruled out); (3) no substantial uncontrolled confounding from other workplace exposures or lifestyle factors; (4) exposures that have been well characterized quantitatively and linked to the individuals in the study, and which are sufficiently variable; and (5) evidence for a dose–response relationship between exposure and outcome. Criteria 2 and 3 address validity of a study; the remaining criteria concern its utility.

Criteria 1 through 4 are necessary for category 1 studies. Criterion 5, though not critical, adds certainty to risk estimates. Category 1 studies are appropriate to serve as a basis for high-to-low-dose extrapolation. Approaches to fitting dose–response curves to such data are discussed in the section under the heading "Category 1 Studies." If no epidemiological study falls into category 1, animal data should be used for low-dose extrapolation.

Category 2 studies are deficient in, at most, one of criteria 1 through 3 and should at least partially satisfy the validity criteria (2 and 3). Although they may be deficient in criterion 4, some quantification of exposure is necessary, even if based on data external to the study site. Well-designed null studies fall in this category. These studies can be used in the dose–response phase of risk assessment for narrowing the range of uncertainty.

Studies substantially deficient in criteria 1 through 4 would, in most cases, fall into category 3; they are unlikely to be useful in quantitative risk assessment, though they may play a role in qualitative assessments (i.e., hazard identification). A notable exception to the category 2 designation may occur if even a crude quantification of exposure is not possible; in this case, a study that would otherwise fall into category 2 would be assigned to category 3.

Implementing the Criteria

1. A positive association with reasonable statistical precision is needed for category 1. When the risk among the exposed is not notably elevated, this criterion can still be met if a subgroup with higher or longer exposure does show

a clearly elevated risk. Additionally, there may be exceptional circumstances where the risk is high but not very precise, yet external data support a causal interpretation; in such a case, one might consider this criterion met. Studies not meeting this criterion may fall into category 2. Thus, excellent studies finding either no positive association or one that differs only marginally from a null association can be used for checking plausibility of animal-based risk estimates, as described below.

2. Overall quality of the study entails appropriateness of study design, control of potential sources of bias, proper statistical analysis, adjustment for confounding (see section 3 below), validity of measurement of exposure and of outcome, adequate study size for reasonable precision, and so forth. Because industrial settings often involve the most concentrated exposures in the population, occupational studies are highly sensitive, having identified about half the known human carcinogens.[20]

Considerations especially critical to occupational studies are (1) length and age distribution of follow-up time (Was the average length of follow-up as long as the latency for the cancer of interest?) and (2) appropriate use of exposure data (Were those with extremely short or low exposures or a low probability of exposure appropriately excluded from the "exposed" categories?). Shore et al. discuss exposure issues in relation to use of epidemiological studies for quantitative risk assessment.[4]

Study size and its corollary, precision, take on a particular meaning in occupational cohort studies.[21] A study with fewer workers (or person-years) may have greater power and precision than a larger study if the expected number of deaths is greater, which could occur because of an older age distribution or a greater prevalence of other factors predisposing to disease. Also, small studies with high exposures may be more powerful than larger ones with lower exposures.[21,22]

Bias could result from inappropriate or poorly defined selection criteria for establishing a cohort; ascertainment of outcomes that is differential with respect to the exposure (or ascertainment of exposures that is differential with respect to outcome); inadequate follow-up with respect to latency periods; systematic errors in exposure data taken from employment and industrial hygiene records; and so forth. Loss to follow-up can be a problem, particularly if related to duration of employment. Since exposure indices are usually constructed independently of vital status data, differential misclassification is often unlikely. Nondifferential misclassification frequently results in bias toward the null; however, when exposure is categorized and misclassification is not into adjacent categories, the bias can be away from the null.[23] The more frequent and widespread the measurements of exposure, the less likely that misclassification will cause substantial bias.

It should be emphasized that no epidemiological study is perfect and that well-conducted ones are unlikely to yield strong biases in effect measures. Nevertheless, with small effects (e.g., mortality ratios <2), the probability of bias should be lower if these effect measures are to be credible enough to support their use in low-dose extrapolation. Determination of limits on the degree of bias may be possible.

3. The condition that substantial confounding can be excluded may seem difficult to meet in occupational mortality studies, many of which do not collect much information on behavioral factors. However, for the putative confounder to explain a sizable positive association, the confounder characteristics of the exposed population would have to differ considerably from those in the unexposed population.[24] Where only external comparisons have been conducted, such differences are often plausible, though even for such strong risk factors as cigarette smoking in lung cancer studies, standardized morbidity or mortality ratios greater than 1.5 to 2.0 are unlikely to be explained by smoking differences.[13,14] This counterintuitive result is because smoking habits do not vary much within sex, age, and time period categories. Weaker risk factors for the disease, such as dietary determinants of cancer, are even less likely to produce nonnegligible confounding. Furthermore, when analyses use internal comparisons, either with an unexposed referent group employed at the same site or with comparisons among several exposure levels, lifestyle differences are even less likely.[25] Hence, direct data on confounding are often not necessary for this criterion to be met. Where other workplace exposures are present and have been shown to be related to the outcome of interest, firmer information on confounding may be needed, particularly since occupational exposures are sometimes highly correlated within a cohort. Confounding is plausible when effect measures are low (mortality ratios <2); however, quantitative limits on the magnitude of potential confounding can be calculated.[14] Moreover, the presence of other exposures that have no established association with the outcome is not persuasive evidence of confounding. For example, the presence of sodium dioxide in a study of lung cancer and arsenic does not detract from the validity of the estimated carcinogenic effect of arsenic, since there is no prior evidence that sodium dioxide is carcinogenic.[26]

4. At least crude measurements of the exposure are needed to conduct a dose–response assessment. Whereas the measurements themselves are almost always ecological (i.e., exposure is usually not measured in individuals but rather in areas of a plant), individual information on work history (job titles, departments, etc.) and periods of employment can be used to estimate individual exposures by time period and hence cumulatively.[27] While gaps in the exposure measurements are common (in some departments and some time periods), a certain amount of extrapolation and interpolation can be tolerated. Future developments involving the use of internal markers of exposure may improve upon measures such as air sampling, but this technology is not ready for application in dose–response assessments or the needed historical tissue or blood samples do not exist. Nevertheless, almost all observational studies use approximations for exposure measurements (e.g., dietary histories for the recent period are often used as surrogates for lifetime dietary intake in studies of cancer and nutrition). Considering the goal of risk assessment, intelligent use of existing records is likely to yield reasonably valid estimates of human health effects at dose levels observed in the study, a prerequisite for valid extrapolated risks.

Between a crude exposure definition (exposed/not exposed dichotomy) and a year-by-year exposure level for each individual worker lies a vast gray area encompassing qualitative exposure levels (low, medium, high) or scant measurements applied to all workers or broad job categories. Industrial hygiene measurements may be available for a plant other than the one where the occupational study was conducted (but one using similar processes) or in only one of several plants where a study was conducted (see, e.g., reference 15). Such studies, if otherwise well conducted, will fall into category 2. Similarly, where exposure measurements have been linked not with individual work histories by job type or

department, but only with duration of employment, the study would usually fall into category 2.

5. A dose–response relationship, that is, a monotonic rise in risk with increasing exposure, is useful. If the dose–response relationship is flat (no increased risk at any exposure level) or negative, then criterion 1 above will not be met. A monotonic dose–response relationship is not necessary, however. For instance, if at high doses an agent caused an increase in cardiovascular deaths with a shorter latency than for cancer, then the dose–response relationship for cancer might fall at high exposures. Alternatively, the shape of a dose–response relationship can be distorted by the healthy worker survivor effect,[28,29] and controlling for this bias is not straightforward.[29,30]

A monotonic dose–response relationship may also be obscured if confounding is differential by exposure level; there are substantial errors of measurement;[31] duration has been used as a surrogate for cumulative exposure; or the range of exposures is too narrow. For all these reasons, the dose–response criterion is not a prerequisite for the category 1 designation, and reliance on a trend test is not appropriate. Finally, if the published data do not provide several dose levels, a dose–response relationship will not be established; nevertheless, a category 1 designation is possible if mean exposure estimates are available.

Discussion of the Framework

One objection to the above framework is that well-conducted null studies appear to be penalized. First, unlike current regulatory practice, the proposed framework does not neglect these studies. Category 2 studies play an important role in integrating animal and human data and can contribute to the low-dose extrapolation by narrowing the range of uncertainty, as described below under the heading category 2 studies and in other publications. Second, this objection stems from not recognizing the purpose of these criteria. The criteria have no relation to a value judgment: a given category 1 study may or may not be of higher quality than a particular category 2 study. Category 1 studies are, however, better equipped to serve as a basis for extrapolation. An extremely well-conducted study may provide inadequate data to be the basis of extrapolation if no association between exposure and disease is found. For example, a study of methylene chloride–exposed workers had highly detailed historical industrial hygiene data, but no stable positive association was observed at cancer sites for which there was an a priori hypothesized relationship.[32] The validity criteria (2 and 3) were met, but two of the utility criteria (1 and 4) were not. In the course of a regulatory hearing in California, it was argued that quantitative use of this study should be relegated to an appendix. In contrast, under the proposed framework, this study would play a prominent role as a check on the plausibility of animal-based risk estimates, including those based on pharmacokinetic modeling.

These criteria are intended as guideposts, not rigid rules. For instance, if animal data are lacking or inconclusive and the human exposure information is crude or the study small, epidemiological data may be used for extrapolation to furnish a plausible range of risk estimates in the interim, pending more complete data.

In the past, risk assessors' decisions about epidemiological data have appeared arbitrary, while a simultaneous tendency has been to perform dose–response assessments as a series of simple algorithms. The spirit of these criteria is to foster consistency, but not rigidity. Each criterion should be considered in detail, but the overall goal must be to have a reliable process for making the best use of available data. No one should have illusions that this process can be carried out mechanically. On the other hand, the lack of category 1 human studies does not imply that a compound is not carcinogenic in humans. As is the current custom, use of animal data is warranted in such circumstances.

Methods for Use of Epidemiological Studies in Dose–Response Extrapolation

Category 1 Studies

Animal-based risk assessments require two extrapolations: one between species and the other from high to low dose. Epidemiological data eliminate the interspecies extrapolation. The high-to-low-dose extrapolation is still needed. A few fundamental points are presented here.

Because animal and human data differ fundamentally in their structure,[33] the statistical extrapolation models take different forms. For both, the aim is to derive a potency value that represents the increase in disease occurrence per unit of exposure. In animal studies, disease occurrence is measured as lifetime risk and exposure is often measured as daily concentration per unit of body weight. Potency then represents the increase in lifetime risk per unit of daily dose. Since humans are not observed for a full lifetime, disease occurrence is measured as rates (number of cases per person-time), and exposure may be measured by duration and concentration or in cumulative time at a concentration level. Thus, potency is in units of mortality rates or ratios for a given (often cumulative) exposure. To derive lifetime risk, a separate step is needed, in which potencies are applied to life tables constructed from age-specific death rate data[5,34] or to proportional mortality data.[35]

Models to derive potency values from occupational mortality data generally fall into two classes: (1) models in which the excess disease rate is a function (often linear) of exposure only (additive models) and (2) models in which the excess disease rate is a function of exposure and background disease rates or factors that determine these background rates (multiplicative models). Both types of models can be embellished to include thresholds, to incorporate the healthy worker effect, and so forth. A full discussion of such methods is beyond the scope of this paper (see, e.g., references 5, 9, and 33 through 37).

Category 2 Studies

On the basis of the five criteria outlined above, many epidemiological studies, including null studies, will fall into category 2. These studies can still be useful in quantitative risk assessment to check plausibility of models fitted to animal data. Exposures in occupational settings frequently fall between the doses administered in experimental bioassays and the low levels to which one wishes to extrapolate (Figure 1). Therefore, if the animal-based extrapolation predicts risks that are inconsistent with occupational study data, the predictions at environmental levels are probably also in error. Discrepancies between predictions and observations may permit rejection of models and a narrowing of the range of plausible risk estimates, as was achieved for ethylene dibromide.[38] Conversely, if predicted risks are consistent with observed mortality in the intermediate range of exposure, then greater confidence can be placed in the risks predicted at low exposures.[22,39,40] (Exposures from active smoking are also in an intermediate range

and can be similarly used to assess plausibility of risk predictions.[34])

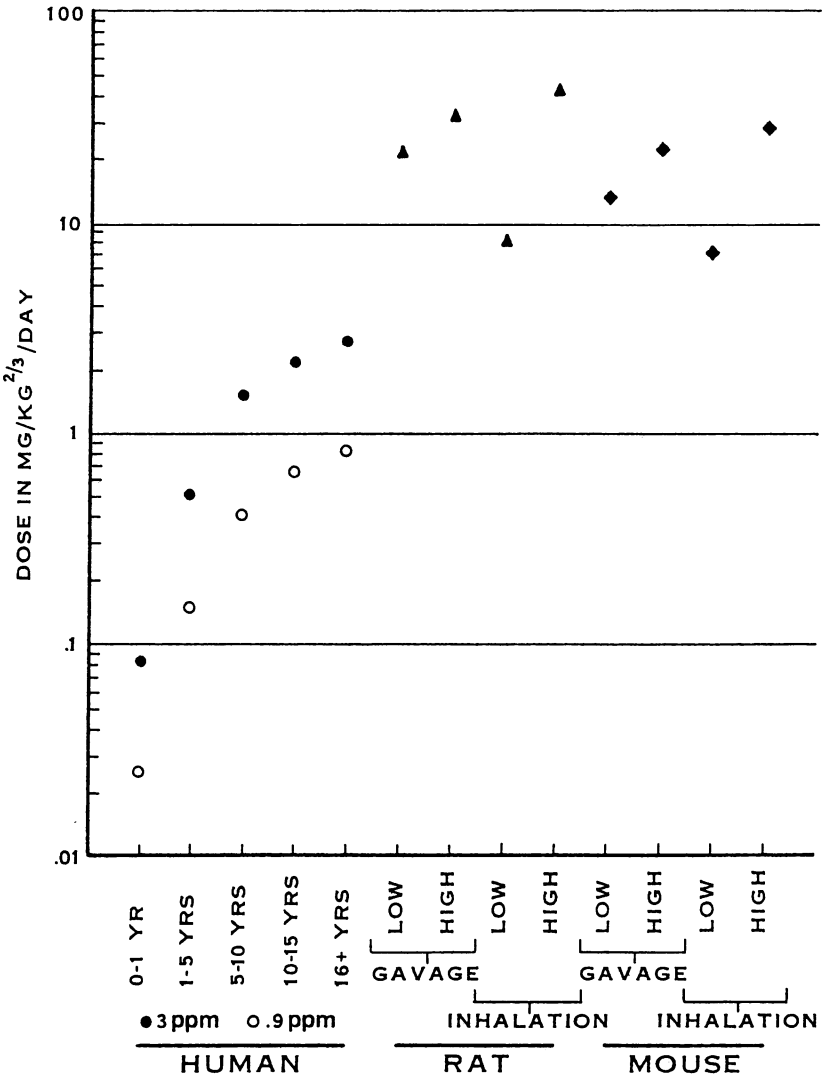
Five steps are needed to implement such comparisons: (1) Human exposures are converted to the units of exposure in the animal study. (2) The calculated unit risk is applied to the doses received by the persons studied. (3) Adjustment is made for partial lifetime observed. (4) Predicted numbers of deaths are compared with observed numbers. (5) Concordance is assessed and interpreted.

Example: In a cohort of 161 ethylene dibromide–exposed workers from two manufacturing plants, 8 cancer deaths were observed, 5.8 expected, for a marginally positive association with exposure.[15] Among those with the longest exposure and with 15 or more years since start of exposure, there were 4 observed and 1.6 expected deaths (criterion 1 possibly but not clearly met). Though completeness of follow-up was not 100%, its duration seemed adequate—30 years for nearly half the workers (criterion 2 met). Potentially carcinogenic exposures were present in one plant, and others may have been present at the other plant (criterion 3 not clearly met). Finally, a few exposure measurements had been made, but in only one of the plants and not at a variety of locations or times. The link to individuals was therefore based not on department or job category but solely on duration of employment (criterion 4 only partially met). Still, these data permitted a category 2 classification. The exposure rate of 0.9 ppm for 8 hours per workday was assumed for each worker, and translated into units of milligrams per kilogram per day over a lifetime; the lifetime risk estimated for 1 mg/kg/day based on a standard risk assessment derived from one animal bioassay was multiplied by this dose, then adjusted to the number of years of follow-up out of a 70-year lifetime. These individual risks were summed, resulting in an estimated 50 cancer deaths predicted by the animal data to occur in this cohort in the given follow-up period. The clear discrepancy between this prediction and the 8 deaths observed led to recognition of a major deficiency in the animal bioassay (very early deaths) and the need for a less simplistic statistical model for the analysis of the bioassay data.[38]

In other instances, this method will add very little additional information, reflecting the imprecision in the original epidemiology or the low predicted effect size, as in the case of saccharin.[21] Alternatively, the observed human data may be quite consistent with the animal-based predictions, as in the case of ethylene oxide,[22] thereby strengthening the credibility of the interspecies extrapolation. With null studies, the information from this exercise could be used to plan further epidemiological studies with more attention to the exposure levels and the likely range of risks. Clear-cut discrepancies between the animal-based predictions and the observed mortality should raise questions about the wisdom of instituting regulations based on those predictions.

This approach to category 2 studies provides a meaningful yardstick for evaluating the predictive validity of both the interspecies and high-to-low-dose extrapolation. Applications to positive,[22] equivocal,[38] and null[39] studies serve as an antidote to the tendency to dismiss nonpositive studies. If implemented as an integral part of risk assessment, these comparisons would improve quantitative estimates by narrowing the range of uncertainty.

Although a formal test of consistency would be desirable, the main sources of uncertainty are not statistical. For this reason, consistency is more appropriately evaluated in real-world terms. That is,



Note. Points represent doses of ethylene dibromide in studies of three species: rats,[17] mice,[16] and humans.[15] Interspecies exposure equivalence was based on mg/(kg body weight)$^{2/3}$ equivalence and doses are plotted on a log scale. The humans were exposed occupationally. Environmental exposures would be even lower, by one to two orders of magnitude, than the occupational exposures.

FIGURE 1—Ethylene dibromide exposures in three studies.

predictions may fall close to observed mortality and within its confidence bounds, suggesting compatibility of the animal-derived extrapolation with actual human experience and increasing confidence in the estimated low-level risk. A three- or fivefold difference may put predictions outside the statistical confidence bounds, but still within a range of tolerable uncertainty, considering the vast exposure differences and other uncertainties in animal-to-human extrapolation. Finally, observed risks may lie far above or far below the animal-based predictions, suggesting the need to revise the extrapolated risk estimates.

Conclusions

Risk assessment is a bridge between science and policy. A serious appraisal of current methods and practice suggests that much can be done to improve scientific rigor in risk assessment. This paper proposes a classification framework and methods for the use of epidemiologic data in dose–response assessment. Although solutions to the basic dilemma of estimating low-dose risks are not close at hand and controversies will continue,[41] implementation of the framework and methods described here would contribute to a firmer scientific foundation for low-dose risk estimates and the ensuing regulatory actions.

Epidemiology offers the most relevant data for the assessment of human health risks at low exposures to environmental agents. That epidemiology has played a small role in most quantitative risk assessments to date attests to the reluctance of experimental scientists to recognize the strength and validity of well-conducted observational studies. It also attests to the reluctance of epidemiologists to engage in risk assessment activities. Unfortunately, where epidemiologists have not been present to interpret human studies, other with less understanding of epidemiology have taken on that role. The result has been inconsistent evaluations of epidemiological evidence, inappropriate use of some human data, and unwarranted dismissal of other studies. The present paper is offered as a first step toward reversing this state of affairs. □

Acknowledgments

Partial support for this work was provided by grant 1P42-ES05948 from the National Institute of Environmental Health Sciences.

The author acknowledges Dr Raymond Neutra, for introducing her to the field, and Drs Donna Armstrong, Aaron Cohen, David Goldsmith, John Goldsmith, David Savitz, Jack Siemiatycki, Leslie Stayner, and Steve Wing for helpful comments on earlier drafts.

References

1. *Research to Improve Health Risk Assessments (RIHRA) Program.* EPA/600/9-90/038. Washington, DC: US Environmental Protection Agency; 1990.
2. Occupational Safety and Health Administration. Identification, classification and regulation of potential occupational carcinogens. *Federal Register.* 1980;45(January 22, 1980):5001–5296.
3. *Workshop Report on EPA Guidelines for Carcinogen Risk Assessment: Use of Human Evidence.* EPA/625/3-90/017. Washington, DC: US Environmental Protection Agency; 1989.
4. Shore RE, Iyer V, Altshuler B, et al. Use of human data in quantitative risk assessment of carcinogens: impact on epidemiologic practice and the regulatory process. *Regul Toxicol Pharmacol.* 1992;15:180–221.
5. Thomas D, Darby S, Fagnani F, et al. Definition and estimation of lifetime detriment from radiation exposures: principles and methods. *Health Phys.* 1992;63:259–272.
6. National Research Council, National Academy of Sciences. *Risk Assessment in the Federal Government: Managing the Process.* Washington, DC: National Academy Press; 1983.
7. Christoffel T, Teret SP. Epidemiology and the law: courts and confidence intervals. *Am J Public Health.* 1991;81:1661–1666.
8. US Environmental Protection Agency. Guidelines for cancer risk assessment. *Federal Register.* 1986;51(September 24, 1986):33992–34001.
9. Smith AH. Epidemiologic input to environmental risk assessment. *Arch Environ Health.* 1988;43:124–127.
10. US Environmental Protection Agency. Draft report: a cross-species scaling factor for carcinogen risk assessment based on equivalence of $mg/kg^{3/4}$/day. *Federal Register.* 1992;57(June 5, 1992):24152–24173.
11. Travis CC, White RK. Interspecific scaling of toxicity data. *Risk Anal.* 1988;8:119–125.
12. Davidson IWF, Parker JC, Beliles RP. Biological basis for extrapolation across mammalian species. *Regul Toxicol Pharmacol.* 1986;6:211–237.
13. Axelson O. Aspects of confounding and effect modification in the assessment of occupational cancer risk. *J Toxicol Environ Health.* 1980;6:1127–1131.
14. Flanders WD, Khoury MJ. Indirect assessment of confounding: graphic description and limits on effect of adjusting for covariates. *Epidemiology.* 1990;1:239–246.
15. Ott MG, Scharnweber HC, Langner RR. Mortality experience of 161 employees exposed to ethylene dibromide in two production units. *Br J Ind Med.* 1980;37:163–168.
16. *Carcinogenicity Bioassay of 1,2-Dibromoethane in F344 Rats and B6C3F1 Mice (Inhalation Study).* Bethesda, Md: National Cancer Institute; 1982. NCI Technical Report Series No. 210.
17. *Bioassay of 1,2-Dibromoethane for Possible Carcinogenicity.* Bethesda, Md: National Cancer Institute; 1978. NCI Technical Report Series No. 86.
18. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Volume 60: Some Industrial Chemicals.* Lyon, France: International Agency for Research on Cancer; 1994.
19. Ames BN, Gold LW. Too many rodent carcinogens: mitogenesis increases mutagenesis. *Science.* 1990;249:970–971.
20. *Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans, Suppl 7: Overall Evaluations of Carcinogenicity: An Updating of IARC Monographs Volumes 1–42.* Lyon, France: International Agency for Research on Cancer; 1987.
21. Hertz-Picciotto I, Neutra RR. Resolving discrepancies among studies: the influence of dose on effect size. *Epidemiology.* 1994;5:156–163.
22. Hertz-Picciotto I, Neutra RR, Collins JF. Ethylene oxide and leukemia. *JAMA.* 1987;257:2290.
23. Dosemeci M, Wacholder S, Lubin JH. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol.* 1990;132:746–748.
24. Cornfield J, Haenszel W. Some aspects of retrospective studies. *J Chron Dis.* 1960;11:523–34.
25. Siemiatycki J, Wacholder S, Dewar R, et al. Smoking and degree of occupational exposure: are internal analyses in cohort studies likely to be confounded by smoking status? *Am J Ind Med.* 1988;13:59–69.
26. Lubin JH, Pottern LM, Blot WJ, et al. Respiratory cancer among copper smelter workers: recent mortality statistics. *J Occup Med.* 1981;23:779–784.
27. Rice CH. Retrospective exposure assessment: a review of approaches and directions for the future. In: Rappaport SM, Smith TJ, eds. *Exposure Assessment for Epidemiology and Hazard Control.* Chelsea, Mich: Lewis Publishers; 1991:186–191.
28. Hertz-Picciotto I, Smith AH. Arsenic exposure and lung cancer: observations on theshape and magnitude of the dose–response curve. *Scand J Work Environ Health.* 1993;19:217–226.
29. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period: application to control of the healthy worker survivor effect. *Math Modeling.* 1986;7:1393–1512.
30. Arrighi HM, Hertz-Picciotto I. The evolving concept of the healthy worker survivor effect. *Epidemiology.* 1994;5:189–196.
31. Armstrong B. The effects of measurement error on estimates of exposure–response relationships. In: P Band, ed. *Occupational Cancer Epidemiology.* Berlin, Germany: Springer-Verlag; 1990:50–63.
32. Hearne FT, Grose F, Pifer JW, et al. Methylene chloride mortality study: dose–response characterization and animal model comparison. *J Occup Med.* 1987;29:217–228.
33. Hertz-Picciotto I, Holtzman DA. Issues in conducting a cancer risk assessment using epidemiologic data: arsenic as a case study. *Exp Pathol.* 1989;37:219–223.
34. Hertz-Picciotto I, Hu S-W. Contribution of

cadmium in cigarettes to lung cancer: an evaluation of risk assessment methodologies. *Arch Environ Health.* 1994;49:297–302.
35. Smith AH, Hopenhayn-Rich C, Bates MN, et al. Cancer risks from arsenic in drinking water. *Environ Health Perspect.* 1992;97:259–267.
36. Stayner L, Smith R, Bailer AJ, Luebeck EG, Moolgavkar SH. Modeling epidemiologic studies of occupational cohorts for the quantitative assessment of carcinogenic hazards. *Am J Ind Med.* 1995; in press.
37. Crump KS, Allen BC. Methods for quantitative risk assessment using occupational studies. *Am Stat.* 1985;39(4, pt2):442–450.
38. Hertz-Picciotto I, Gravitz N, Neutra R. How do cancer risks predicted from animal bioassays compare with the epidemiologic evidence? The case of ethylene dibromide. *Risk Anal.* 1988;8:205–214.
39. Alexeeff GV, Hertz-Picciotto I. Consideration of species concordance and pharmacokinetics in a risk assessment of methylene chloride. *Toxicologist.* 1990;10:351.
40. Goldman LR, Hayward D, Siegel DM, et al. Dioxin and mortality from cancer. *N Engl J Med.* 1991;324:1811. Letter.
41. Hertz-Picciotto I. Environmental risk assessment. In: Talbott EO, Craun G, eds. *An Introduction to Environmental Epidemiology.* Boca Raton, Fla: Lewis Publishers. In press.

Comment: Integrating Epidemiologic Data into Risk Assessment

Daniel Wartenberg, PhD, and Ronald Simon, JD

Policymakers use two principal tools in evaluating human health risk: epidemiology and quantitative risk assessment. Epidemiology is the gold standard because it assesses directly human health risk. However, when epidemiologic data do not exist or when epidemiologic studies are not conclusive, regulators often turn to quantitative risk assessment. For example, we know the dangers of smoking cigarettes and the benefits of low-fat diets and exercise through series of observational epidemiologic studies. We know the utility of a variety of drugs and surgical procedures through carefully controlled clinical trials. However, with compounds less well studied in humans, such as dioxin, quantitative risk assessment is relied upon to set regulatory policy. Using epidemiology to assess human health risk is not controversial; using quantitative risk assessment is. To explain this difference and the possible implications of using epidemiologic data in quantitative risk assessment, we explain the basics of quantitative risk assessment, point out some of its limitations, and raise some cautions on the use of epidemiologic data in quantitative risk assessment models.

Quantitative risk assessment is a statistical method designed to forecast human health risk where risk is hard to measure directly, as with people who shower in water contaminated with trichlorethylene or who dwell beside Superfund sites.[1,2] The basic tenet of quantitative risk assessment is that data on health effects detected in small populations of animals exposed to high concentrations of suspect chemicals can be used to predict health effects in large human populations exposed to lower concentrations of the same chemical. Most federal agencies conform to a 1983 National Academy of Sciences report[2] that defines quantitative risk assessment as a four-stage process. Though each stage has objective elements, each also requires some decisions based on subjective judgements into which personal values may enter. Disagreement and controversy often follow.

The goal of the first stage of quantitative risk assessment, *hazard identification,* is to identify all situations or substances that can, in any amount, pose a risk to human health as well as all the possible adverse health effects. Omission of compounds or specific health effects from consideration at this stage can undermine the validity of a quantitative risk assessment.

The goal of the second stage, *exposure assessment,* is to estimate for each material listed in the hazard identification stage the amount a typical person is likely to encounter. The three components to this step are determination of the source of the substance, the movement of the substance through the environment, and the uptake by people (i.e., ingestion, inhalation, and dermal exposure). Omission of sources, exposure pathways, bio-

Daniel Wartenberg is with the Environmental and Occupational Health Sciences Institute, UMDNJ–Robert Wood Johnson Medical School, Piscataway, NJ. Ronald Simon is with Connerton, Ray and Simon, Washington, DC.

Requests for reprints should be sent to Daniel Wartenberg, PhD, EOHSI, 681 Frelinghuysen Rd, PO Box 1179, Piscataway, NJ 08855-1179.

Editor's Note. See related editorial by Shore (p 474) in this issue.