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.

concentration factors, and absorption and transformation coefficients can reduce the accuracy of estimates.

The third stage, *dose–response modeling,* determines the amount of each substance that causes harm. For carcinogens, this is typically called the "cancer potency." Most often, risk assessors use a mathematical model that describes the logarithm of the risk as a linear function of dose, without a threshold below which there is no impact. Both the assumptions of linearity and no threshold have come under criticism, as well as the implicit assumption that each compound acts independently of all other exposures.

In the fourth stage, *risk characterization,* the risk assessor combines the information from the three other stages into a single overall estimate of risk. For each chemical listed in *stage 1,* the predicted cumulative exposure is calculated for an average (or worst case) person over an entire average (or worst case) lifetime (*stage 2*) and multiplied by the potency estimated (*stage 3*) to derive a predicted risk of cancer for that individual. These risks for each compound are added up to get a total risk for the activity being evaluated, such as residing near a solid waste incinerator or inhaling benzene when filling up an automobile gasoline tank. Implicit in summing of these risks is the assumption of independent (not interactive) but additive (not multiplicative) effects. These assumptions are controversial.

The assumptions used in quantitative risk assessment highlight some of its limitations. Compounds may be omitted inappropriately or measured inadequately, and their effects may be modeled improperly. Errors may overestimate or underestimate the true risk. Advocates of quantitative risk assessment, while acknowledging these limitations, argue that the quantitative effects of these errors are small and that the benefits of the objectivity of the method far outweigh this concern. Yet, surprisingly little effort has been made either to quantify these potential errors or to validate the approach.

What has been done has revealed problems. For example, evaluations of the consistency of the animal data have shown substantial disparities.[3–6] Tests of carcinogenicity in animals of the same species but different sex are consistent with each other only about 65% to 85%; in those of the same sex but different species animals tests were consistent only 50% to 75%; and cancer occurred at the same anatomical site across species in only 35% to 50% of the tests. Without even addressing the potency of carcinogenic chemicals, which can vary by several orders of magnitude across animal species for the same chemical, these disparities raise questions about the general utility of quantitative risk assessment for predicting carcinogenicity.

Similarly, quantitative risk assessment results are not entirely compatible with human epidemiologic data.[7–9] In one study, the differences between risks predicted with the rat and mouse data and the results of human epidemiologic studies ranged over 10 orders of magnitude.[10] Even the rank orders of risks often did not agree.[11] Reasons postulated for such discrepancies include inadequate human exposure data, biased sampling (e.g., the healthy worker effect in occupational studies, partially "exposed" controls in case-control studies), uncontrolled confounding, genetic variability, variations in susceptibility, behavior and lifestyle, and interactive effects.[12,13]

For a few compounds, investigators have provided analyses, explanations, and sometimes adjustments for these apparent discrepancies,[12–16] but overall there has been no systematic evaluation and explanation of these discrepancies. In spite of this, most scientists agree that using these data is preferable to arbitrary or subjective decision making. Rather than invalidating the use of quantitative risk assessment, these problems pose challenges that must be investigated. By identifying and understanding the ambiguities and inconsistencies, we can increase our knowledge and ability to predict risk. However, we also must make clear the uncertainty and imprecision of the risk estimates derived with the current methodology.

In this Public Health Policy Forum, Hertz-Picciotto extends this approach. She proposes the use of epidemiologic data, where available, in place of the animal data typically used in the dose-response modeling stage of quantitative risk assessment and offers a protocol for evaluating the quality of epidemiologic studies. Her general proposal for the evaluation of epidemiologic data is not controversial. However, we believe that the use of epidemiologic data without additional caveats may compound rather than alleviate the shortcomings of quantitative risk assessment risk estimates.

For example, one limitation of much epidemiologic data is their insensitivity or low statistical power. Even if a true exposure-disease association exists, an epidemiologic study may fail to show an effect simply because the sample size is too small. To detect associations of regulatory importance, typically risks of 10^{-4}–10^{-5} or less, studies often need a minimum of several thousand subjects. Such studies of low risk are expensive, difficult to conduct, and hence relatively rare. Nonetheless, sanction for the use of epidemiologic data in quantitative risk assessment models may tempt many to dismiss the risks of carcinogenicity on the basis of studies without positive results but lacking power. Some have argued that with some epidemiologic data in hand, and in the absence of proof of hazard, one should assume relative safety. But, as Doll[17] notes, cautioning against the use of human data in preference to animal data in assessing human health risks, "Proof of absence of an effect by epidemiologic means . . . is, however, very much harder to achieve than proof that an effect is produced. . . . Negative human evidence may mean very little, unless it relates to prolonged and heavy exposure."

A second limitation of relying on epidemiologic data in quantitative risk assessment is that compounds believed to be hazardous cannot ethically be administered to human subjects. Thus, no human experimental studies of suspect chemicals can be undertaken; and in those situations in which people are knowingly being exposed to a suspect chemical, action must be taken to reduce exposure rather than study the consequences thereof.

A third limitation of using epidemiologic data in quantitative risk assessment is that while epidemiologic studies invite less extrapolation than animal experiments, they are more subject to bias, confounding, and effect modification (interaction). Often, sufficient data are not available to identify or fully control for these effects. While epidemiologic methods exist to adjust, in part, for these problems, most quantitative risk assessment models do not allow for this. The consequence of using biased or confounded data is that predicted risks may markedly overestimate or underestimate the true risk.

Using epidemiologic data with quantitative risk assessment models without adequate caution can lead to erroneous conclusions and could promote unsafe but preventable exposures. This concern is borne out by the many compounds thought to be safe on the basis of early epidemiologic studies which, upon further study, have been found to be hazardous (e.g., dioxin). Any rules or guidelines suggesting use of epidemiologic data in quantitative risk assessment must address this issue

head on. Negative and inconclusive epidemiologic studies (such as those that do not meet the standards of statistical significance) should be considered only if the power of each study is incorporated in the calculations, for example, by using the largest possible effect that could have gone undetected, along with a thorough assessment of bias and confounding. Discarding positive animal data in preference for limited, imprecise or confounded epidemiologic data must be avoided.

The Agent Orange litigation is an example of how misunderstanding epidemiologic concepts can lead to deleterious social effects. In the Agent Orange tort litigation, US District Judge Jack B. Weinstein made a detailed review of scientific evidence that Agent Orange exposure had caused particular diseases. The court rejected scientific opinion in support of causation because of "negative" epidemiologic studies.[18]

In recent years, the National Academy of Sciences, the Department of Veterans Affairs, and the Environmental Protection Agency have determined that a wide variety of adverse health effects are related to Agent Orange exposure.[19,20] Veterans brought product liability cases against the manufacturers of Agent Orange. The courts determined that the legal rights of these veterans had already been determined by the litigation decided by Judge Weinstein even though the veterans had not become ill and did not have legal claims until years after that decision. Thus negative epidemiologic information was used to infer safety, and that inference took legal precedence even after additional data were collected that showed hazard. Nonetheless, regulatory decisions need to be made about exposures and their likely, or possible, impact on human health. These decisions about the carcinogenicity and relative potency of chemicals are made not to advance scientific knowledge, but to reach decisions about public health protection.

Scientifically, we can gain most by validating the models we use and developing new ones to accommodate all available data. We must refocus efforts to assess the assumptions underlying quantitative risk assessment and the validity of the method instead of quantifying the uncertainty resulting from inadequate data, as is currently being done. We must use all animal and epidemiologic data together rather than selecting only one or a few data sets, possibly weighting each by source, quality, and statistical power. Heterogeneity between data sets should be understood as properties of the data rather than viewed as discrepancies or errors.

On the regulatory front, the best strategy is more complex. In the long term, the study of biological mechanisms and pharmacokinetics offer hope for resolving many discrepancies between species. In the meantime, we must acknowledge explicitly the limitations of our data and our models as well as the inconsistencies in the data for the specific chemical being studied. Rather than choosing one type of data over another, we must try to use all data to explain the heterogeneity and to accommodate it. Rather than presenting results of quantitative risk assessment as precise estimates of expected risk (i.e., with confidence intervals derived from Monte Carlo simulations that focus on assumptions about the exposure data distributions), we should conduct sensitivity analyses under broader sets of assumptions (for instance, using different exposure regimes, alternative dose–response models, alternative interspecies adjustments, etc.). When estimates diverge under different models, we should seek explanations for these divergences, not averages or arbitrary priorities for the models.

Only by owning up to the limitations of the methodology do we stand any hope of improving it. By accepting the incompleteness and conflicting nature of data, we can use it to make social decisions that reflect the political nature and scientific understanding of the alternatives being considered. The subjective and situation-specific input required is preferable to deluding ourselves about the accuracy and precision of the data and models used for rendering objective decisions. □

References

1. Wartenberg D, Chess C. Risky business: the inexact art of hazard assessment. *The Sciences.* 1992;March–April:16–21.
2. National Academy of Sciences/National Research Council. *Risk Assessment in the Federal Government: Managing the Process.* Washington, DC: National Academy of Sciences; 1983.
3. Purchase I. Inter-species comparison of carcinogenicity. *Br J Cancer.* 1980;41:454–468.
4. Haseman J, Humm J, Zeiger E, McConnell E. Comparative results of 327 chemical carcinogenicity studies. *Environ Health Perspect.* 1987;74:229–235.
5. Haseman J, Lockhart AM. Correlations between chemically related site-specific carcinogenic effects in long-term studies in rats and mice. *Environ Health Perspect.* 1993;101:50–54.
6. Gold L, Manley N, Ames B. Extrapolation of carcinogenicity between species: qualitative and quantitative factors. *Risk Anal.* 1992;12:579–588.
7. National Academy of Sciences Executive Committee. *Pest Control: An Assessment of Present and Alternative Technologies.* Washington, DC: National Academy of Sciences; 1975.
8. Crouch E, Wilson R. Interspecies comparison of carcinogenic potency. *J Toxicol Environ Health.* 1979;5:1095–1118.
9. Allen B, Crump K, Shipp A. Correlation between carcinogenic potency of chemicals in animals and humans. *Risk Anal.* 1988;8:531–544.
10. Hart R, Turturro A. Comments on Allen et al. *Risk Anal.* 1988;8:545–547.
11. Portier C. Species correlation of chemical carcinogens. *Risk Anal.* 1988;8:551–553.
12. Smith A. Epidemiologic input to environmental risk assessment. *Arch Environ Health.* 1988;43:124–127.
13. Shore R, Iyer V, Altshuler B, Pasternack B. 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.
14. 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.
15. Hertz-Picciotto I, Holtzman D. Issues in conducting a cancer risk assessment using epidemiologic data: arsenic as a case study. *Exp Pathol.* 1989;37:219–223.
16. Rowe J, Springer J. Asbestos lung cancer risks: comparison of animal and human extrapolations. *Risk Anal.* 1986;6:171–180.
17. Doll R. Purpose of symposium. In: Wald NDR. *Interpretation of Negative Epidemiological Evidence for Carcinogenicity.* Lyon, France: IARC; 1985. IARC publication 65:3–10.
18. In re Agent Orange Product Liability Litigation, 611 F Supp 1223 (ED NY 1985), *aff'd,* 818 F2d 187 (2d Cir 1987), *cert denied,* 487 US 1234 (1988).
19. 59 CFR 29723 (June 29, 1994).
20. Affidavit of Jenkins in *Ivy v. Diamond Shamrock Chemical Co.,* 781 F Supp 902 (1991), *aff'd,* 996 F2d 1425 (2d Cir 1993), *cert denied,* _ US _ (February 22, 1994).