TOXICOLOGIC PATHOLOGY, vol 29(Suppl.), pp 5–12, 2001

Assessment of Human Cancer Risk: Challenges for Alternative Approaches

GILBERT S. OMENN

The University of Michigan, Ann Arbor, Michigan

ABSTRACT

The ILSI/HESI Workshop on Alternatives to Carcinogenicity Testing aims to develop and apply new methods for assessment of potential carcinogenic risk to humans from various chemicals. The Workshop represents a major cooperative scientific effort. The long-term goals should be to greatly enhance the efficiency and reliability of such testing and to supplant, not just supplement, lifetime rodent bioassays. There are now well-established frameworks for risk assessment and risk management, putting risks into public health context and engaging stakeholders. The Lave-Omenn value-of-information model provides a useful way to assess the social costs and benefits of different strategies for testing large numbers of chemicals.

Keywords. Cancer risk assessment; risk management; cost-effectiveness analysis; public health context; value-of-information model; social costs.

INTRODUCTION

The International Life Sciences Institute's (ILSI) Alternatives to Carcinogenicity Testing (ACT) workshop brought together collaborators from many countries in an effort to advance the field of hazard identification for chemicals. The aims of the ILSI ACT Program may be summarized as follows:

1. Develop and apply new methods for assessment of potential carcinogenic risk to humans from a great variety of chemicals.
2. Stimulate a cooperative scientific effort.
3. Promote consideration of use of new tests and data in risk assessment.

The desired outcomes for this effort are to provide a better mechanistic basis for interpretation of short-term tests and rodent lifetime bioassay results and to generate results and interpretations of results more reliably related to risk in humans. I believe we should look forward to a time when new, better understood tests will supplant, not just supplement, rodent lifetime bioassays.

We have a large challenge. The public around the world finds scientists' conflicts over the interpretation of toxicological and epidemiological scientific evidence bewildering. There is also a common observation that the interpretation by scientists may be highly correlated with employment in industry versus employment in governmental regulatory agencies. Surely we can make progress in agreement on the nature of the testing to be done and the analysis of the results, both mechanistically and statistically. The decisions on what to do with that information rest on larger issues, statutes, policies, and economic interests, discussed later in this paper.

As we develop new test strategies and infer risk for humans there is a much neglected biological phenomenon that deserves more emphasis, namely, the crucial influence of heterogeneity. We should always speak of "cancers" in the plural for there are multiple causes, multiple pathogenetic pathways, and varied responses to treatments and preventive interventions. The phrase "The War on Cancer," which began in 1970 in the United States, was an illusion, as if there might be a unitary cause and a miraculous cure for all kinds, or at least most kinds, of cancers. Over these decades we have noted the emergence from basic science studies of new signaling pathways, many specific receptors, an ever-growing array of molecular targets, and now distinctive mRNA and protein expression patterns. This molecular heterogeneity and the corresponding epidemiological and clinical heterogeneity are highly relevant for research strategies on mechanisms, diagnosis, treatments, preventive interventions, and testing. In this Workshop, we are focused on the implications for testing.

Finally, we should recognize the ramifications of analyzing subpopulations with the new tools of pharmaco-genomics. Subpopulations may exhibit variation in biotransformation of pharmaceutical agents and variation in therapeutic or adverse effects of those agents at target sites. Many companies are concerned that the US Food and Drug Administration (FDA) and counterpart agencies in other countries might require just as extensive clinical trials in these subgroups as in larger populations, making the studies more difficult and more costly to conduct and possibly limiting the markets to more narrowly defined patient populations. On the other hand, a drug candidate that might be rejected due to infrequent adverse reactions might be approvable and widely used if those at risk for the adverse effect could be reliably

Address correspondence to: Gilbert S. Omenn, The University of Michigan, 7324 Medical Science I Building, Ann Arbor, MI 48109-0626.

0192-6233/01$3.00+$0.00

TABLE 1.—Major hazardous chemical laws in the United States.

EPA:	Air Pollutants	Clean Air Act 1970, 1977, 1990
	Water Pollutants	Fed WP Control Act 1972, 1977
	Safe Drinking Water	Safe DW Act 1974, 1996
	Pesticides	FIFRA 1972
	Ocean Dumping	Marine Protection Act, 1995
	Toxic Chemicals	TSCA 1976
	Hazardous Wastes	RCRA 1976
	Hazardous Waste Cleanup	CERCLA (Superfund) 1980, 1986
	Food Quality & Protection	FQPA, 1996
FDA:	Foods, Drugs, Cosmetics	FDC Acts 1906, 1938, 1962, 1977, 1997
CEQ:	Environmental Impacts	NEPA 1969
OSHA:	Workplace	OSH Act 1970
CPSC:	Dangerous Consumer Products	CPS Act
DOT:	Transport of Haz Materials	THM Act 1975, 1976, 1978, 1979, 1984, 1990

identified in advance. We shall have to find a balance in those regards.

Assessing Risks of Chemicals and Putting Risks in Public Health Context

As shown in Table 1, there are numerous statutes in the United States for identification and control of potentially hazardous chemicals. There is no overarching law covering all media and all uses. Thus there are specific laws for air pollutants, water pollutants, pesticides, foods, drugs, cosmetics, workplace exposures, consumer products, transport of hazardous materials, and other categories of products, effluents, and wastes. Similar laws exist at the state level in this country and in comparable agencies around the world. The different statutes have different rationales as reflected in Table 2.

We may note several important objectives of formal assessments of risks from exposure to potentially hazardous chemicals. The first is to balance risks and benefits. Exactly that strategy underlies our laws on pharmaceuticals and pesticides. These biologically active agents are often designed to kill living things, ranging from microbial agents and cancer cells to insects and other pests. However, such other statutes as the Clean Air Act and Safe Drinking Water Act are focussed on risk only; these statutes require the regulatory agency (EPA) to protect the public from hazards without regard to costs. In fact there is a long-standing argument over whether economic analyses and estimates of economic costs may be given weight in the risk management decisions, especially with regard to the Clean Air Act. Thus for air, water, and food it is necessary to set some target level for maximal exposure, and the attendant risk, that can be judged to be acceptable or, preferably, negligible.

TABLE 2.—Objectives of risk assessment.

1. Balance risks and benefits
   Drugs
   Pesticides
2. Set target levels of risk
   Food contaminants
   Water pollutants
3. Set priorities for program activities
   Regulatory agencies
   Manufacturers
   Environmental/Consumer organizations
4. Estimate residual risks and extent of risk reduction after steps are taken to reduce risks

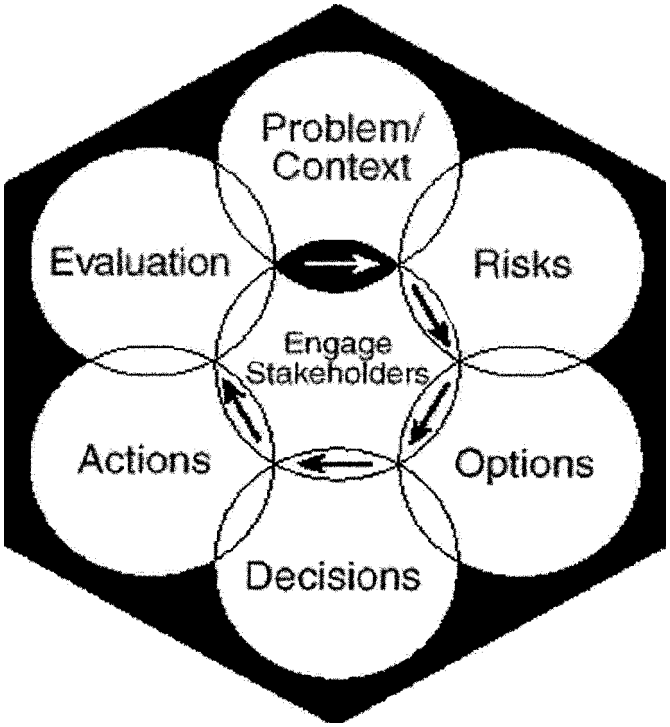

FIGURE 1.—Environmental health risk management framework (from 22).

The Risk Commission

Figure 1 shows the hexagon of the 6-stage Framework for Risk Management developed during 1994–1997 by the Presidential/Congressional Commission on Risk Assessment and Risk Management (Omenn Commission) (22). This commission was mandated by the Clean Air Act amendments of 1990, as was the National Academy of Sciences/National Research Council report *Science and Judgment in Risk Assessment* (16). That report, with its very meaningful title, was important input to the Risk Commission. The hexagon shows that the assessment of risks is a necessary step before the generation of options and the generation of decisions for actions. However, the first step is of extreme importance and is too often neglected. Each environmental problem—whether arising from a laboratory study, a cluster of cases in a neighborhood, the media report of findings elsewhere, or an analytical chemical study of effluents or contaminants—needs to be put into a public health and ecological context.

As shown in Table 3, such context has several components. First, for a specific chemical in development by the manufacturer or under review by the regulatory agency, there may be multiple sites of production, emissions, and uses. There usually will be multiple pathways of exposure in the environment. For example, contaminants of water may be volatilized into the air, and air pollutants will precipitate into bodies of

TABLE 3.—Putting problems into context.

- Multiple sources of same chemical
- Multiple media/pathways of exposure to same chemical
- Other causes of same endpoint(s) (attributable fraction)
- Multiple effects of same chemical

TABLE 4.—Biological endpoints.

• Cancers	• Neurobehavioral toxicity
• Mutations	• Organ-specific effects
• Birth defects	• Endocrine modulation/disruption
• Reproductive toxicity	• Ecosystem effects
• Immunological toxicity	

water and onto land. Next, there may be other causes of the same effect. In epidemiology we refer to the attributable fraction, the proportion of cases of lung cancer or birth defect or other endpoint shown to be due to the specific chemical among all cases. Finally, as outlined in Table 4, there may be multiple effects of the same chemical, causing multiple cancers or other biological endpoints.

As shown in the center of the hexagon (Figure 1), the Risk Commission's Framework emphasized the importance of actively engaging a broad array of stakeholders in the risk management process. We believe strongly that this process should start at the problem context stage, before risk analysis is launched, so that questions important to community groups, to workers, about infants or children, or other practical or technical matters will be identified and can be considered in the risk assessment, options development, and decision-making stages. Otherwise, fine work of the usual technical sort may neglect questions and make the stakeholders dissatisfied and distrustful of the conclusions of the regulatory agency and the inputs of the manufacturers. The Risk Commission published an entire report with examples of engagement of public groups, sometimes productive, sometimes less so (22).

Framework for Regulatory Decision-Making

In the 1970s, the US Environmental Protection Agency and the Food and Drug Administration, together with the Occupational Safety and Health Administration in the Department of Labor, the Consumer Products Safety Commission, the Food Safety program of the Department of Agriculture, and the

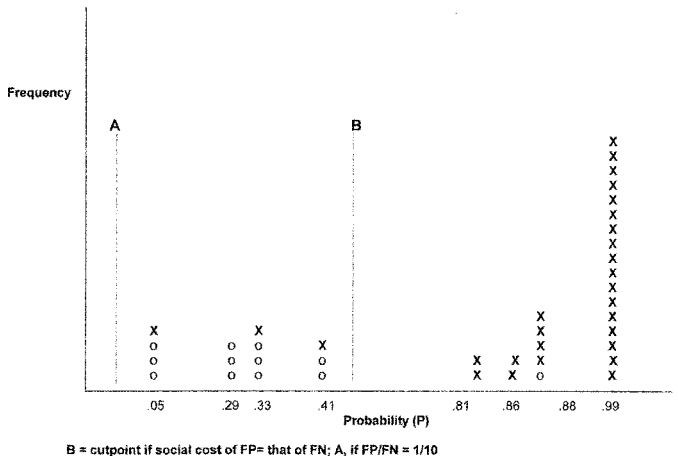

FIGURE 2.—Histogram showing value of predicted probabilities (P) of carcinogenicity for the set of 38 chemicals in the International Collaborative Program. Cutpoints corresponding to FN/FP values of 10 (A) and 1 (B) are indicated by dashed lines; the cutpoint must be greater than zero to accommodate false positive test results. X indicates a carcinogen, while O indicates a noncarcinogen (from 13).

TABLE 5.—Framework for regulatory decision-making.

Hazard identification	Epidemiology Lifetime rodent bioassays Short-term, in vitro/in vivo tests Structure/activity
Risk characterization	Potency (dose/response) Exposure analysis Variation in susceptibility
Risk reduction	Information Substitution Regulation/Prohibition

White House Office of Science and Technology Policy, developed the framework for regulatory decision-making shown in Table 5.

The three key elements for any agent under consideration are: 1) identification of potential hazard, 2) characterization of the risk, and 3) strategies for risk reduction, when warranted. We shall discuss the methods for hazard identification extensively in this paper and in the rest of the Workshop. The concept of risk characterization arose in a special report from the Office of Science and Technology Policy (2) to embrace both qualitative and quantitative assessment of risk. We felt then and I feel strongly still that the term risk assessment is too often considered synonymous with quantitative risk assessment. A risk estimate expressed as a single number, generally an upper bound estimate, with or without confidence intervals or Monte Carlo simulations, is not sufficient. Narrative is required to identify and explain the nature of the health risks, the consistency and weight of evidence, and the reversibility or preventability of the adverse effects. All of those elements are intended to be captured in the term "risk characterization." As shown in Table 5, risk characterization must address the potency of the agent, the details of sources and pathways of exposure, and the variation in susceptibility across identifiable subgroups in the population.

Next, the responsible parties must decide if action is warranted. The most immediate stimulus to action is information. Information may come from manufacturers, from regulatory agencies, or from environmental or consumer groups, generally through the media. Often, such information leads to behavior change to reduce exposures, to alter uses, or to change products. It generally takes longer to generate substitutions; we should be aware that we almost always have more information about a chemical that is the target of regulatory action than we do about any available substitute. In the cases of the flame-retardant chemical TRIS, the detergent nitrilotriacetic acid, the sweetener cyclamate, and the food colorant red dye #2, regulators and scientists knew much more about those agents than about their substitutes, all of which turned out to have similar or worse risks ("son of TRIS," phosphates, saccharin, red dye #40).

In the next part of this paper, I comment in some detail on the several methods for Hazard Identification outlined in Table 5.

Epidemiology

Epidemiology is the study of causes and risk factors of illness and injury in humans. The advantage is the direct study of the species of most interest for protection of human health. The disadvantage is that epidemiological studies require human exposures. For existing chemicals in widespread use,

epidemiology can be quite feasible. For new chemicals, of course, one would have to await exposures, perhaps decades of exposures, before definitive studies could be conducted of potential adverse effects in humans. There is also the limitation that observations in humans cannot be controlled in the same way that a protocol can control the diet and other exposures in rodents; therefore, epidemiologists must try to exclude confounding factors, most of which are unknown in specific situations. Depending upon the potency of the agent and the levels of exposure, epidemiological studies are usually insensitive, limited by sample size, by low levels or limited variation in exposures, and by uncertainties about mechanisms and confounding factors. Studies of working populations have the features of higher exposures and better exposure estimates, but usually small numbers of individuals with such exposures. Studies of the general population, in contrast, may have much larger numbers of people, but their exposures are generally lower, often much lower, and there is greater uncertainty about all features of their exposures.

One of the most important neglected features of the presentation of epidemiological results is the need to express "negative results" in terms of the size of effect that would have been needed in order for a difference from the null to have been detectable. For example, the observed result may be "no statistically significant difference," but the point estimate might be a relative risk of 1.5 when the study design would have required a relative risk of at least 1.8 (80% higher than the control group) in order to be statistically significant.

A. B. Hill (11) established criteria to consider in the evaluation of epidemiological associations that remain an excellent guide today. Thus, reviews of epidemiological findings should evaluate strength of association, consistency and specificity of findings, temporality of results (adverse effects occur after exposures, rather than the reverse), biological gradients or dose response relationship, and—very important for our emphasis on mechanisms in this Workshop—biological plausibility, and coherence with experimental evidence. These criteria require us to seek connections between observations in humans and results in animal bioassays.

Lifetime Rodent Bioassays

Standard protocols for the assessment of potential carcinogenicity in test species, generally rats and mice, have greatly advanced the fields of toxicology and risk assessment. The experimental control of variables and the large experience with common laboratory strains of rats and mice represent great advantages for this work. However, there are several limitations of animal bioassays. First, there needs to be clearer justification for the choices of species, strains, doses, route of exposure, and time course of exposure. Second, there is seldom investigation of the mechanisms of action and of differences in metabolism in the standard bioassay. Thus, the experiment is treated as a "black box," with inputs of exposures and counting of tumors or adverse effects during and after the exposure periods. There is also a problem that some of the tumors that arise in these rodents occur in odd sites or with unusual histology, complicating the extrapolation to risk in humans for particular cancers. Third, statistical criteria must be applied to determine whether the results of the hazard identification assay are positive or negative. This seemingly yes/no interpretation depends on conventions about statistical significance, making the qualitative judgment rest on quantitative criteria. For reference, that matter is illustrated in Table 6, which shows the number of animals that must be affected in the exposed group compared with the percent of animals observed to have similar effects in the unexposed or control group (10). Next, bioassay animal results require extrapolations from animals to humans and from high dose to low dose, sometimes over enormous dose ranges from those that produce adverse effects in 10 to 100% of animals in a positive bioassay (see Table 6) down to extrapolated risk of < 1/1,000 or < 1/1,000,000 (0.0001%) for humans. Finally, we must recognize the high costs (about $2 million per chemical) and the 3–5 years required to organize, conduct, analyze, and interpret lifetime rodent bioassays.

TABLE 6.—Bioassay results for statistical significance.

Controls (N = 50 per dose)		Exposed (N = 50)	
%	N	%	N
0	0	12	6
10	5	28	14
20	10	40	20
30	5	52	26
40	20	62	31
50	25	72	36

From Goldberg (10).

Are Rodent Bioassays a Worthy "Gold Standard" for Carcinogenicity Testing?

There are many problems with the lifetime rodent bioassay. Many compounds are carcinogenic only at doses at or near the "maximal tolerated dose." It is sometimes unlikely and often uncertain whether the effects at the maximal tolerated dose can justifiably be extrapolated linearly downward on the dose/response curve. High doses may exceed the capacity of host detoxification and other defense mechanisms that would be protective at lower exposures. It is noteworthy that results for chemical carcinogenicity testing are concordant only 70% of the time between rats and mice; it is unlikely that the concordance between rodents and humans would be higher. As already noted in Table 6, the observable range for the cancer endpoint requires at least 12% incidence among the exposed animals, whereas the regulatory agencies are mandated by statute or policy to protect below 0.0001% upper bound estimated risk. Thus, the modes of action and the models for extrapolation become crucial.

According to Joseph Contrera of the Food and Drug Administration (personal communication via James MacDonald), about 1,700 chemicals have been tested in multiple-dose, long-term cancer bioassays. In the database developed by Gold and Ames, there are about 3,000 compounds (8, 9). Nevertheless, even these numbers represent a limited database for regulation of the vast array of chemicals in widespread commercial use.

According to the Environmental Protection Agency (7), only 7% of high production volume (HPV) chemicals have a full set of studies for 6 basic toxicity endpoints; 43% of HPV chemicals have no publicly available studies for any of these 6 endpoints. There may be substantial data in the files of manufacturing companies or in the proprietary files protected by confidentiality at the regulatory agencies, but such data, by definition, are not publicly available. The Environmental

TABLE 7.—Rodent carcinogenic responses not likely to apply in humans.

Tumor site	Illustrative chemical agents
Male rat kidney	D-limonene, unleaded gasoline
Male bladder	Saccharin, nitrilotriacetic acid
Rat thyroid	EBDC fungicides, goitrogens
Forestomach-only (after gavage)	BHA, propionic acid, ethyl acrylate
Mouse liver	Certain classes of liver carcinogens

References (15, 5, 22, 18, 7).

Defense Fund, addressing this matter, generated an influential report called "Toxic Ignorance" (23). The OECD and EPA have stimulated new commitments by companies globally to test high production volume chemicals and generate standard data sets; that work is underway with cooperation from the chemical industry.

A very special problem with rodent bioassays as "the Gold Standard" for extrapolation to human risk assessment arises from exceptions to the general rule that we can and should rely on observations in rodents to predict potential hazard and risks in humans. As shown in Table 7, there are certain tumor responses in specific organs in rodents that seem to have no counterpart in humans. The best example is the development of nephropathy and tumors in the kidneys of male rats through induction of α_2-euglobulin by chemicals such as D-limonene and unleaded gasoline. No such response occurs in female rats or in other species. In the cases of saccharin and nitrilotriacetic acid, salts of these chemicals, when tested at doses that exceed their solubility product when concentrated in the urinary outflow tract, precipitate and produce crystals that irritate the lining of the bladder. This irritation leads to hyperplasia and, in a small percentage of test animals, cancers. The effect is limited to the males and does not occur at lower doses that would not exceed the solubility product. Other information about these very interesting mechanistic studies can be found in the references (5, 7, 15, 18, 22).

Short-Term Tests

Twenty years ago, I was hopeful that the rapidly emerging array of in vitro "short-term tests" would lead to scientific and regulatory strategies with batteries of such tests and tier-testing schemes that would greatly assist regulatory decision-making. The potential advantages are substantial: short time, usually measured in days or weeks; low cost (typically a few thousand dollars, compared with 1–2 million dollars for a lifetime rodent bioassay) and, very importantly, information about the mechanism of action. Understanding mechanisms would permit assessment of positive and negative results for coherence by mechanism and not just overall "positive" or "negative" results. However, the progress has been modest. There have been problems with validity and reproducibility of results. There is quite a variety of tests and of mechanisms. There has been lack of agreement on the decision rules or "cutpoint" to evaluate whether results should be called positive or negative. And the short-term test results have been added to regulatory requirements, rather than replacing the bioassays.

Figure 2 displays results of carcinogencity testing from the International Collaborative Program on short-term tests for carcinogens (3) in the form of a histogram of results for the 38 chemicals tested. The dashed vertical lines, A and B, indicate potential cutpoints for positive/negative conclusions, depending upon the relative valuation of false positives and false negatives. Cutpoint B at approximately 50% probability would be used when the consequences of a false positive (declaring a chemical carcinogenic when it is not, thereby losing some or all of its economic value) would be considered equal to the consequences of a false negative (declaring a chemical safe and thereby leading to health risks in the population that is unprotected). If the social decision were to value false negatives, for example, 10 times higher than false positives, the cutpoint for declaring the results predictive of carcinogenicity would be set at point A, below 5% probability, but still greater than zero. Economist Lester Lave and I developed a model to guide strategies for testing of chemicals based upon the social cost of testing and the consequences of false positives and false negatives (13, 14, 19).

The Lave-Omenn Value-of-Information Model

It is interesting that engineers and radiation biologists commonly use the principle of "ALARA" (as low as reasonably achievable) for practical control of exposures and for design of safety features. Instead of elaborate quantitative estimations of risks, they use this pragmatic approach. In principle, the same could be done for chemicals. Therefore, we created a "value-of-information" model to determine what the features would have to be for carcinogenicity testing in order to make more economically efficient and socially justifiable judgments with testing than without testing. Detailed presentation of the model and its applications to short-term tests and lifetime rodent bioassays can be found in the references.

Figures 3a and 3b illustrate the method. We assume that either 10% (Figure 3a) or 2% (Figure 3b) of a very large number of chemicals to be tested are truly carcinogenic to humans. Then, we estimate the costs of testing and the consequences of false positives and false negatives for short-term testing with very high specificity and sensitivity (90% each) and very modest testing costs of $8,000 per chemical for in vitro tests. These figures show that the results of testing when 10% of the chemicals are truly carcinogenic are approximately $10 billion in total social costs, compared with a social cost if no testing were done of $50 billion, clearly a very substantial benefit for testing under these assumptions. In Figure 3b, when only 2% of chemicals are assumed to be truly carcinogenic, the social cost of no testing falls to $10 billion for 50,000 chemicals, but the total social costs of conducting and interpreting the tests are still lower. Clearly, if the testing is much more expensive, as in the case of the transgenic mouse models, let alone the lifetime rodent bioassays, the proportion of chemicals truly carcinogenic must be much higher and the specificity and sensitivity must be very high to make testing cost effective. Elaborate figures illustrating the interaction of these factors are presented in Lave et al (13). The general formula for the information model is:

$$acn > [sc + (1 - b)acn + (1 - b')(1 - a)cm];$$

dividing by acn:

$$1 > [s/(an) + (1 - b) + (1 - b')(1 - a)m/(an)]$$

The crucial parameters are: a, proportion of true carcinogens among c chemicals; s/n, ratio of cost of testing to cost of a false negative; and m/n, ratio of cost of false-positive to cost of false-negative (19).

True biological activity

In vitro test results	Carcinogenic	Non-carcinogenic
Potentially carcinogenic	TP 4,500	FP 4,500
Not potentially carcinogenic	FN 500	TN 40,500

Social Costs ($B)
--of false positives
4,500 x $1 million = $4.5
--of false negatives
500 x $10 million = $5.0
Total = $9.5
Cost of testing $0.4
Total social cost $9.9
Social cost if no testing, 5,000 FNs
5,000 x $10 million $50.0

(a)

True biological activity

In vitro test results	Carcinogenic	Non-carcinogenic
Potentially carcinogenic	TP 900	FP 4,900
Not potentially carcinogenic	FN 100	TN 44,100

Social Costs ($B)
--of false positives
4,900 x $1 million = $4.9
--of false negatives
100 x $10 million = $1.0
Total = $5.9
Cost of testing $0.4
Total social cost $6.3
Social cost if no testing, 1,000 FNs
1,000 x $10 million $10.0

(b)

FIGURE 3.—Applications of Lave-Omenn Value-of-Information Model (from 13): a) 10% of 50,000 chemicals are true carcinogens; b) 2% of 50,000 chemicals are true carcinogens. Assumed specificity and sensitivity are 90% and cost of short term testing is $8,000 per chemical. These parameters can be modified to fit alternative test schemes, including use of transgenic assays.

The NIEHS/NTP Carcinogen Prediction Challenge

A decade ago, the National Toxicology Program conducted an admirable experiment utilizing, in the end, 40 chemicals that were entered in the routine rodent bioassay carcinogenicity testing program and inviting interested scientists to predict the results of the rodent bioassays based upon structural features of the chemicals and whatever short-term tests, subchronic tests, and mechanistic information were available (1, 25). In all, 10 sets of predictions were submitted and evaluated against the results in a workshop format. Nearly everyone predicted that 10 of the 40 chemicals would be carcinogenic; all proved positive in the rodent bioassay. There was a nearly unanimous prediction that 9 of the 40 chemicals would be noncarcinogenic; 6 were noncarcinogenic, whereas the other 3 had equivocal results. The problem was that 21 of the 40 had highly variable predictions and poor correlations.

We applied the Lave-Omenn model to these predictions of NTP carcinogenicity results (20). Figures 4a and 4b show 2 of the 10 sets of predictions in the tabular format that we utilized to show concordance or discordance of predictions with results. For these sets of predictions for 40 (or fewer) chemicals, we calculated the accuracy of the prediction, the number of false positives and false negatives and their ratio, and the sensitivity and specificity. Figure 4a shows results from the group who organized the project, Tennant and colleagues. Figure 4b shows the results from a group headed by Herbert Rosenkranz. This group utilized the MultiCASE prediction model based on structure/activity relationships. The predictions using this model were less accurate; the sensitivity and specificity were quite poor.

We also applied the Value-of-Information Model to the rodent bioassay (13) in light of our goal of risk management for

Predictions by Tennant et al.	NTP Bioassay Results		
	+	-	Totals
+	17	7	24
-	3	13	16
Totals	20	20	40

Chi square = 10.4 ($p > 0.1$)

(a)

Accuracy: 0.75 (30/40)

*FP/FN: 2.33 (07/03)

Sensitivity: 0.85 (17/20)

Specificity: 0.65 (13/20)

Social Cost (FP + FN): $37M*

Predictions with MultiCASE	Bioassay Results		
	+	-	Totals
+	10	9	19
-	9	7	16
Totals	19	16	35

Chi square = 0.0 ($p > 0.2$)

(b)

Accuracy: 0.49 (17/35)

*FP/FN: 1.00 (09/09)

Sensitivity: 0.53 (10/19)

Specificity: 0.44 (07/16)

Social Cost (FP + FN): $99M

FIGURE 4.—Analysis of predictions of NTP carcinogenicity bioassay results (from 20): a) predictions of Tennant et al; b) predictions of Rosenkranz et al, MultiCASE method. *Social cost $1M for FP and $10M for FN.

TABLE 8.—VOI model applied to rodent bioassay and risk management for Humans (13).

- IARC lists only 26 chemicals or groups of chemicals as showing definite evidence of carcinogenicity for humans: Either few chemicals are carcinogens, or few of the carcinogens are potent enough and reach enough people to have been recognized in humans as carcinogenic.
- If the likelihood a chemical is carcinogenic is 50% or more, use of the rodent bioassay(@$1M/bioassay) is more costly than classifying these chemicals as carcinogens without further testing.
- If concordance of rodent bioassay to true effect in humans is below 70%, or likelihood of chemical being carcinogenic is below 10%, social cost is less if classify all as noncarcinogenic.
- If testing were not done, it would be necessary to implement an "as low as reasonably achievable" (ALARA) approach to exposure reduction.

humans. Table 8 briefly summarizes illustrative guidance on whether or not social costs of testing specific sets of chemicals, including the consequences of false positives and false negatives, would justify the testing.

Expectations for the Future

For every test we will want to know the sensitivity and specificity of the findings, the likely mechanisms, the classes of chemicals most reliably detected, and the cost of testing. It is possible that the tests will be either too insensitive or too sensitive; we need criteria, as illustrated in Figure 2, to choose cutpoints to make that judgment. Of course, there are other strains that are highly sensitive, and that have been used in recent decades, including the strain A mouse with lung adenoma; the C3H and C57BL mice for liver foci; and the Sencar mice for 2-stage skin painting (Gary Stoner, personal communication, October, 2000) (21). Thus, it might be wise to compare the characteristics and costs of the new assays with use of such older strains. In any case, these transgenic assays can be subjected to the analysis shown in Figure 3.

It will be interesting to evaluate the correlation of results in the new assays against rodent results utilizing the Omenn et al (1995) approach to the results from the NTP Carcinogen Prediction Challenge. Then, we must try to put the results with this very interesting set of 21 chemicals in perspective for presumed human risks, because 12 of the 21 were chosen (24) as "rodent carcinogens/putative human noncarcinogens" (see Table 7). It would be natural to combine the output of the ILSI ACT initiative with the ILSI/HESI collaboration genomic and proteomic expression in hepatoxicity, nephrotoxicity, and genotoxicity, and with the NIEHS Environmental Genome Program (17).

The recommendation of the International Conference on Harmonization (12) to utilize one rodent bioassay, presumably the rat, together with one of the new assays, such as those being evaluated in these workshop proceedings, is a step in the right direction. It will be necessary to combine the results of these assays with the results from short-term tests, toxicokinetics, pharmacodynamics, genetic variation, and structure activity relationships. I hope that results from these new assays and from associated studies will enable us to eliminate the remaining rodent lifetime bioassay altogether in the coming years, so that we may have the resources to test many more high-production-volume existing chemicals, as well as new chemicals of great promise to our economy and our society.

REFERENCES

1. Ashby J, Tennant RW (1994). Prediction of rodent carcinogenicity of 44 chemicals: Results. *Mutagenesis* 9: 7–15.
2. Calkins DR, Dixon RL, Gerber CR, Zarin D, Omenn GS (1980). Identification, characterization, and control of potential human carcinogens: A framework for federal decision-making. *J Natl Cancer Inst* 61: 169–175.
3. deSerres FJ, Ashby J (eds) (1981). *Evaluation of Short-Term Tests for Carcinogens: Report of the International Collaborative Program*, Elsevier, New York.
4. EPA (1996a). *Food Quality Protection Act* (FQPA): Office of Pesticide Programs, Washington DC.
5. EPA (1996b). *Proposed Guidelines for Carcinogen Risk Assessment*. EPA/600/P-92/003C). Office of Research and Development, Washington DC.
6. EPA (1998). *Chemical Hazard Availability Study*. EPA Office of Pollution Prevention and Toxics, Washington DC.
7. Faustman EM, Omenn GS (2001). Risk assessment. In: *Casarett and Doull's Toxicology*, 6th ed, Klaassen CD (ed). McGraw-Hill, New York, pp. 83–104.
8. Gold LS, Manley NB, Slone TH, Rohrbach L (1999). Supplement to the Carcinogenic Potency Database (CPDB): Results of animal bioassays published in the general literature in 1993 to 1994 and by the National Toxicology Program in 1995 to 1996. *Environ Health Perspect* 107, Suppl 4: 527–600.
9. Gold LS, Slone TH, Ames BN (1998). What do animal cancer tests tell us about human cancer risk? Overview of analyses of the carcinogenic potency database. *Drug Metabolism Reviews* 30: 359–404.
10. Goldberg L (ed) (1973). *Carcinogenesis Testing on Chemicals*, CRC Press, Boca Raton, Florida.
11. Hill AB (1965). The environment and disease: Association or causation? *Proc Roy Soc Med* 58: 295–300.
12. International Conference on Harmonization, Expert Working Group on Safety (1996). Draft guideline on testing for carcinogenicity of pharmaceuticals. *Fed Reg* 61: 43298–43300.
13. Lave LB, Ennever FK, Rosenkranz HS, Omenn GS (1988). Information value of the rodent bioassay. *Nature* 336: 631–633.
14. Lave LB, Omenn GS (1986). Cost-effectiveness of short-term tests for carcinogenicity. *Nature* 324: 29–34.
15. McClain RM (1994). Mechanistic considerations in the regulation and classification of chemical carcinogens. In: *Nutritional Toxicolocy*, Kotsonis F, Mackey M, Hijele J (eds). Raven, New York, pp 278–304.
16. National Research Council (1994). *Science and Judgment in Risk Assessment*. National Academy Press, Washington, DC.
17. Omenn GS (2000). The post-genomic era: A crucial role for the public health sciences. *Environ Health Perspect* 108: 160–161.
18. Omenn GS, Faustman EM (2001). Risk assessment and risk management. In: *Oxford Textbook of Public Health*, 4th ed, Detels R, McEwen J, Baeglehole R, Tanaka H (eds). In press.
19. Omenn GS, Lave LB (1988). Scientific and cost-effectiveness criteria in selecting batteries of short-term tests. *Mutation Res* 205: 41–49.
20. Omenn GS, Stuebbe S, Lave LB (1995). Predictions of rodent carcinogenicity testing results: Interpretation in light of the Lave-Omenn Value-of-Information Model. *Molecular Carcinogenesis* 14: 37–45.
21. Pereira MA, Stoner GD (1985). Comparison of rat liver foci assay and strain A mouse lung tumor assay to detect carcinogens: A review. *Fund Appl Toxicol* 5: 688–699.
22. Presidential/Congressional Commission on Risk Assessment and Risk Management (1997a,b). A Framework for Environmental Health Risk Management (Vol 1); Risk Assessment and Risk Management in Regulatory Decision-Making (Vol 2). Government Printing Office, Washington, DC.
23. Roe D, Pease W, Florini K, Silbergeld E (1997). *Toxic Ignorance*. Environmental Defense Fund, Washington, DC.
24. Robinson D (1998). International Life Sciences Institute's role in the evaluation of alternative methodology for the assessment of carcinogenic risk. *Toxicol Pathol* 26: 474–475.
25. Tennant RW, Spalding JW, Stasiewicz S, Ashby J (1990). Prediction of the outcome of rodent carcinogenicity bioassays currently being conducted on 44 chemicals by the U.S. National Toxicology Program. *Mutagenesis* 5: 3–14.