REGULATORY TOXICOLOGY AND PHARMACOLOGY 7, 1–20 (1987)

Criteria for Identifying and Classifying Carcinogens, Mutagens, and Teratogens

Developed Jointly by an International Working Party of Experts,
Organized under the Auspices of the
Safety of Chemicals Committee of CEFIC
International Affairs Group of CMA/SOCMA
and
Canadian Chemical Producers Association[1]

Received May 29, 1986

AUTHORS

Cochairmen

B. BROECKER, *Head of Department Umweltchemikalien, Hoechst AG, Postfach 800320, D-6230 Frankfurt Am Main 80, West Germany.*

E. J. SOWINSKI, *Corporate Manager, Toxicology, Industrial Hygiene, Product Safety, Uniroyal Chemical Company Inc., Middlebury, Connecticut 06749.*

Carcinogenic Criteria

J. FACCINI, *Consultant Toxicologist, ECETOC, Brussels, Belgium.*

J. T. BARR, *Manager of Regulatory Response, Air Products and Chemicals, Inc., Allentown, Pennsylvania 18105.*

S. BEEKHUIZEN, *Referent Safety Department, Bayer, Leverkusen, Federal Republic of Germany.*

C. S. WEIL, *Corporate Fellow, Toxicology, Union Carbide Corporation (Retired), Past President Society of Toxicology, Consultant, 4326 McCaslin Street; Pittsburgh, Pennsylvania 15217.*

[1] Authors are listed according to study groups. Appendix I gives the various working parties.

0273-2300/87 $3.00

Mutagenic Criteria

H. P. GELBKE, *Certified Specialist in Toxicology DPHG, Director of Department of Toxicology, BASF, Ludwigshafen, Federal Republic of Germany.*

H. FLEIG, *Department of Toxicology, BASF, Ludwigshafen, Federal Republic of Germany.*

D. W. MATHESON, *Manager of In-Vitro Toxicology, Stauffer Chemical Company, Farmington, Connecticut 06032.*

G. C. GRANVILLE, *Manager, Toxicology and Material Safety, Shell Canada Ltd., Calgary, Alberta, Canada.*

W. ROZENBOOM, *Dangerous Substances, Regulatory Affairs Liaison, Shell International Petroleum, The Hague, The Netherlands.*

Teratogenic Criteria

D. M. FERGUSON, *Director of Scientific Affairs, ICI Americas Inc., Wilmington, Delaware 19897.*

I. F. CARNEY, *Central Toxicology Adviser, Imperial Chemical Industries plc Central Toxicology Laboratory, Cheshire, Great Britain.*

Epidemiologic Criteria

S. PELL, *Epidemiologist, E. I. DuPont de Nemours and Company (Retired), Consultant, 1416 Emory Road, Wilmington, Delaware 19803.*

A. MUNN, *Director, Medicine and Environmental Health, Monsanto Europe, Brussels, Belgium.*

J. C. AUBRUN, *Chief Medical Advisor, Rhone-Poulenc, Courbevoie, France.*

M. J. REAPE, *Director of Health Services, FMC Corporation, Philadelphia, Pennsylvania.*

Increasingly good science is being done in uncovering and identifying chemical hazards and risks. However, we lack an adequate language worldwide for communicating the conclusions of toxicological science with all its inherent uncertainty. Effective communication of chemical hazards and risks requires a basic framework of definitions or criteria, representative of current science, upon which adequate language for communicating conclusions can be based. This paper puts forth a set of working criteria offering consistent guidance for identifying and classifying carcinogens, mutagens, and teratogens. These criteria represent an attempt to (1) harmonize principles in the hazard identification of carcinogenic, mutagenic, and teratogenic substances; (2) provide guidance for the purpose of judging the relevance of experimental data, on a weight of evidence approach; and (3) serve as a constructive factor in discussions with authorities when differences in respective regulations are an issue. These criteria may serve as a consistent and harmonized basis upon which chronic health hazards are classified and identified for health protection purposes.

INTRODUCTION

A tripartite group of industrial experts from Europe, Canada, and the United States, dealing with toxicology, epidemiology, and chemical classification, has developed a set of working criteria offering consistent guidance for identifying and classifying carcinogens, mutagens, and teratogens. These criteria represent an attempt to (1) harmonize principles in the hazard identification of carcinogenic, mutagenic, and teratogenic substances; (2) provide guidance for the purpose of judging the relevance of experimental data, on a weight of evidence approach; and (3) serve as a constructive factor in discussions with authorities when differences in respective regulations are an issue. These criteria may serve as a consistent and harmonized basis upon which chronic health hazards are classified and identified for health protection purposes.

A main purpose of this exercise is to minimize international trade distortions caused by the use of differing classification schemes for carcinogens, mutagens, and teratogens.

The set of guidelines developed is based on existing relevant documents which exist worldwide and which vary in detail and scope (1–12). Appendix II is a comparison of some currently relevant approaches for classifying chemicals as carcinogens, mutagens, or teratogens.

The following papers were prepared by the Working Party:

"Criteria for Classifying Chemicals as Carcinogens,"
"Criteria for Classifying Chemicals as Mutagens,"
"Criteria for Classifying Chemicals as Teratogens," and
"Criteria for Inferring Causality from Epidemiological Studies."

The fourth paper, "Criteria for Inferring Causality from Epidemiological Studies," was included because standards for epidemiological findings are very often not adequately covered in existing identification and classification schemes.

It is not intended in this paper to comment directly on existing relevant documents, nor to initiate a completely novel set of concepts for organizing the relevant science. Moreover, these documents do not set a standard for classifying chronic health hazards but do reflect the thinking of the group of scientists who participated in this effort.

GENERAL PRINCIPLES FOR CLASSIFICATION: DISCUSSION

Carcinogens

The purpose for classifying carcinogens is to provide guidance for judging human and animal data on proven and suspected carcinogens so that appropriate and consistent interpretations as to their relevance to humans can be made.

No attempt has been made to review the current state of the science of carcinogenicity. This has been done recently by the U.S. Office of Science and Technology Policy (OSTP) and others (13, 14). It is assumed, however, that all available scientific data relevant to the carcinogenicity of a substance have been assembled before begin-

ning an assessment of this substance, and that the nature and quality of the data are considered in its acceptance.

The definition of a carcinogen implies that it causes malignant tumors; however, it is recognized that a precise distinction between benign and malignant tumors is not always possible and that in certain instances, e.g., some endocrine tumors, it is valid to combine them when assessing incidence in a particular organ (7). Nevertheless, the finding of an increase in malignant tumors is the hallmark of carcinogenicity and for this reason it is stressed in this text. Similarly, for a substance to be considered a suspected carcinogen to humans or a proven animal carcinogen of potential relevance to humans under expected conditions of exposure, it will normally need to be genotoxic, as well as giving a positive result in appropriately performed animal carcinogenicity bioassays. The exact relevance to man of chemicals that are positive in a bioassay but negative in a battery of short-term tests for genotoxicity is not clear and these substances particularly—as with all chemicals—need to be considered on a case-by-case basis for hazard classification in the groups below. The very large variations in potency and latency of carcinogens have not been utilized in classifying substances for hazard identification purposes, but these properties should be considered when making risk management decisions.

In evaluating the available data, it has to be realized that no single piece of evidence by itself provides the definitive answer. It is only when all the evidence is evaluated that implications to human health can be defined. It is particularly important when considering the relevance of ambient exposures to humans that the mechanism of action and the probable shape of the dose–response curve at very low doses be considered.

The majority of classifications currently employed, e.g., European Economic Community (EEC) (1), Environmental Protection Agency (EPA) (2), International Agency for Research on Cancer (IARC) (5), European Chemical Industry Ecology and Toxicology Centre (ECETOC) (9) envisage three broad categories of carcinogens. Some regulations utilize classifications made by others, e.g., Occupational Safety and Health Administration (OSHA) and Workplace Hazardous Materials Information System (WHMIS) (3, 4). The expanded classification used in this document is intended to be compatible with these other systems but does not correspond exactly with any of those systems, which, incidentally, do not correspond precisely with each other.

Mutagens

The classification of chemicals into various categories according to their mutagenic properties as described here is based on a weight of evidence consideration of available test results. The goal of this exercise is to detail the criteria for this classification and compare them with existing systems (1, 2, 4).

Many test systems are used today to provide indications of the mutagenicity of chemicals, including highly complex intergenerational animal studies with the potential for direct indication of a heritable effect. Most of the test systems, however, only give results relating to part of the mutagenic spectrum. This reality is reflected in the following classification scheme by the inclusion of multiple endpoint criteria.

The current literature attributes up to 10% of human ill health to defective or disease predisposing genes and further shows that 1% of all live newborns inherit genetic

defects, some of which can be ascribed to new mutations in the parental germ line. Approximately another 1% of all newborns bear chromosome abnormalities of some type (15, 16). This high background incidence of genetic damage and problems in the identification and classification of mutagenic effects in humans makes the epidemiologic investigation of heritable effects caused by one specific agent very difficult (17). It should be noted that direct evidence for the occurrence of a causal relationship linking chemical exposure and an increase in the frequency of occurrence of a heritable effect in a human population has never been shown. The absence of this sort of evidence in humans does not rule out the potential for human DNA and chromosomes to be affected in ways comparable to those that have been observed experimentally in laboratory animals and, therefore, there must be concern about the possible hazards of chemicals causing heritable effects. Animal studies and supporting *in vitro* studies are currently used to predict genetic hazard associated with chemical exposure.

Predictions of relative hazard are generally made from a weight of evidence approach. Since evidence of germ cell mutagenicity is of greatest concern, animal tests measuring this are placed in the highest possible category. Other evidence results in the classification of a substance in a category of lower human concern.

Some further general remarks on classifying mutagens are listed below:

(1) In the classification of mutagens only the potential hazard for humans is relevant, mammals being taken as predictive substitutes.

(2) For classification purposes, the mutagenic event of concern is the transmission of heritable effects through the germ cells to the next generation. Evidence for mutagenicity in somatic cells may provide useful information relating to this concern.

(3) Gene mutation and chromosomal aberrations are the specific endpoints to be considered. These should be regarded as separate mutagenic endpoints, with each endpoint being evaluated independently.

(4) In evaluating evidence of mutagenicity, greater weight should be placed on tests conducted with germ cells than on tests conducted with somatic cells, on tests performed *in vivo* rather than *in vitro*, on tests in mammalian species rather than submammalian species, and on tests in eukaryotes rather than prokaryotes.

(5) Mixed test results may exist for a particular agent and endpoint. In those cases the preponderance of the evidence, weighted as described above, should determine the potential for *in vivo* mammalian germ cell mutagenicity rather than the number of positive results obtained.

(6) For positive results in particular assays there may be evidence to indicate that the positive result is not applicable to *in vivo* germ cell mutagenicity. Such evidence may include pharmacokinetic or metabolism data or considerations which are clearly not relevant to an intact mammal. In these cases it is not appropriate to consider the positive results in the weight of evidence approach.

(7) Other information about the chemical under consideration, such as structure–activity relationships, can be used for the indication of possible mutagenic activity within certain well-studied chemical classes. However, at the present time there is no established scientific basis to reliably predict the mutagenic activity of a chemical from its structure alone and it should not be used as the sole criterion for classification purposes.

(8) Only results from valid tests, conducted according to established scientific procedures, should be used to categorize chemicals for hazard identification purposes.

EEC, U.S., and Canadian documents for classifying mutagenicity data were reviewed (1, 2, 4). Similarities between the systems were identified and a single new system was developed for classifying substances according to relative strength of evidence for potential human germ cell mutagenicity.

Five categories indicating relative levels of concern have been developed for this new classification scheme. Substances are assigned to one of these categories based on the best match between the test results profile, the category criteria, and the remarks listed above under General Principles for Classification: Discussion. In assigning a substance to a category, the quality of the test results and the appropriateness of the test must be carefully considered. A description of many of the mammalian germ cell mutagenicity assays currently in use and factors to be considered in their application and interpretation has been developed (23).

Teratogens

There is a need to develop criteria for classifying those chemicals that cause adverse effects on the reproductive process (1, 11). As a first step, this document offers guidance for classifying chemicals on the basis of their ability to cause specific adverse effects on the developing conceptus. This limitation in scope is appropriate since such effects are, at present, the only endpoints that can be reproducibly assessed. It is also consistent with some existing classification schemes (e.g., EEC) which do not address all potential effects on the reproductive process. At a later date other adverse effects on reproduction may be addressed.

A basic problem is one of definition. In this field, many of the terms in use have been—and are—applied loosely and this has led to misunderstanding. Given the emotive nature of the subject, it is crucial that the terminology used should be precise to ensure consistency in interpretation and evaluation of human or experimental animal data. A number of expert groups have defined terms in this area, but such definitions are often acknowledged to be valid only for the particular document in question (18).

In this case, the following two definitions have been agreed upon by many specialists in the fields of toxicology and occupational medicine and are not believed to conflict with those derived by other expert groups:

Reproductive toxicity: adverse effects of chemicals/agents that interfere with the ability of males or females to reproduce.

Developmental toxicity: adverse effects of chemicals/agents on the developing conceptus associated with exposure during pregnancy. These effects may be manifested in the embryonic or fetal periods, or postnatally.

One specific manifestation of developmental toxicity is the production of structural malformations which are usually deleterious. Thus, a "teratogen" may be defined in the strict sense as an agent causing irreversible, deleterious structural malformations in a conceptus as a consequence of exposure of the mother during pregnancy.

Whether or not a specific malformation is deleterious is a matter of judgment in each particular case. This definition excludes agents causing other manifestations of

developmental toxicity, i.e., death of the conceptus, growth retardation, or functional deficits. It is consistent with the approach outlined in the EEC Labelling Guide which refers specifically to teratogens (and requires the phrase "May cause birth defects") (1). Thus the other types of developmental toxicity remain unclassified for the present since morphological change is "the only reproducible assessment currently available" (19).

Epidemiology

In establishing hazard identification procedures for carcinogenic, mutagenic, and teratogenic health hazards, categories have been established to classify substances according to how much evidence exists that a particular substance is a human carcinogen, mutagen, or teratogen (1, 2).

Criteria have also been developed to assist in deciding into which category a substance should be placed. Some of these criteria, however, are not sufficiently specific as to how they relate to humans. Criteria that are too general may allow so much room for differing interpretations that they can lead to widely differing conclusions. For example, if there are to be harmonized warnings on chemical labels, criteria for categories of evidence of hazards would have to be uniform, and if they are to be uniform, they have to be sufficiently specific.

On the other hand, criteria should not be so rigid that they limit the exercise of individual judgment. There is a need to strike a balance between criteria that are too specific and those that are too general.

If a substance is to be classified as a proven carcinogenic, mutagenic, or teratogenic health hazard in humans, that classification must be based on unequivocal human evidence derived from epidemiological studies. Therefore, an assessment of these studies must be made to decide whether the substance should be categorized as a proven human carcinogen, mutagen, or teratogen, or placed into another category indicating less certainty.

In making this statement, there is a need for a set of criteria for inferring causality from epidemiological studies. Certain limitations of some epidemiological data can make it difficult at times to determine whether there is a cause-and-effect relationship between chronic diseases and exposure to a particular substance. Examples are as follows:

(a) An excess incidence of death rate among exposed persons may be due to confounding factors that are unknown or known but cannot be controlled.
(b) The sample size may not be large enough to provide the statistical power needed to detect excess risks.
(c) Information on level and duration of exposure may be inadequate.
(d) The study period may not allow enough time for a latent period to elapse.

In setting standards for suspected human carcinogens or in classifying substances according to evidence of carcinogenicity, regulatory agencies and expert scientific panels have had to deal with two major areas of controversy in the interpretation of epidemiological data. One is the interpretation of observed excess of cases where it is uncertain whether the excesses are chance occurrences or due to confounding factors. The other pertains to substances that have been found to be carcinogenic in certain

species of laboratory animals, but not carcinogenic in humans, as indicated by negative epidemiological studies. Differing interpretations of these data could be a major obstacle in developing harmonized hazard warnings (12).

The major question raised by both of these issues is, to what extent do the epidemiological studies indicate a causal relation between cancer and exposure to the substance in question? Therefore, if there is to be a resolution of these issues, a uniform set of criteria for inferring causality from epidemiological studies would have to be adopted by all agencies involved in hazard identification.

The principles outlined in this paper for inferring causality from epidemiological studies apply equally to studies of mutagenicity, teratogenicity, and carcinogenicity in humans.

CRITERIA FOR CLASSIFYING CHEMICALS AS CARCINOGENS

Category 1: Proven Human Carcinogenic Substances

Evidence for inclusion in this category is provided by formal epidemiological studies. For details we refer to the section on epidemiology.

Category 2: Suspected Human Carcinogenic Substances

There is sufficient evidence that human exposure to the substance may result in the development of cancer because of:

(1) suggestive epidemiological data not sufficient to satisfy the criteria for establishing causality, as described in the paper on epidemiology, *and*

(2) proven evidence from animal studies carried out under conditions which are relevant to human exposure as defined in Category 3 below.

A causal relationship suspected from limited data on humans may be better understood if specific evidence from animal studies is available, particularly if cancers have been induced in animals by routes and levels of exposure which are relevant to the expected human exposure. Similarly, studies of genotoxicity and of metabolic processes can assist in this understanding. The evidence from animal studies, however, must be strong evidence as defined in Category 3 below.

Category 3: Proven Animal Carcinogenic Substances of Potential Relevance to Humans

If the evidence is not adequate to classify a substance in Categories 1 and 2, consideration should be given to Category 3 if the strength of the evidence is sufficient and was obtained under conditions relevant to the expected human exposure. In this case, it may be assumed that exposure to the substance is of potential relevance to humans.

The following findings increase the justification for a substance being classified as a proven animal carcinogen with relevance to man when obtained under exposure conditions which correspond to those in humans:

(a) A clear statistical and biologically significant increase in the incidence of malignant tumors in an organ with a low spontaneous tumor incidence, e.g., less than 4% (6, 10). If a substance induces malignant tumors only in an organ of high spontaneous tumor incidence, it would be considered for inclusion in this category if the incidence was well in excess of that in the control group including historical controls.

(b) The induction of a statistically significant excess of malignant tumors at more than one site.

(c) The existence of a clear dose–response relationship in the number or time to appearance of the malignant tumors.

(d) The induction of an excess of malignant tumors in more than one species or strain. (If found in more than one strain, and especially in more than one species, the tumors are regarded as more significant if they are of the same type.)

It is not necessary that all of the above requirements be fulfilled for a classification as proven animal carcinogen nor are they all equal in strength. Each case should be considered on its merits. However, at least one of the following conditions should be met (5):

There should be a clear statistical and biologically significant increase in the incidence of malignant tumors

(i) in more than one species or strain, or
(ii) in multiple experiments preferably with different routes of administration or at different dose levels, or
(iii) to an unusual degree with regard to incidence, site or type, or age at onset.

The data would not normally be regarded as sufficient for a proven animal carcinogen classification if

(i) only benign tumors are induced, or
(ii) an excess only of tumors such as hepatic nodules in rats or mice, or only pulmonary tumors in mice is induced, or
(iii) an excess of malignant tumors is induced only in an organ which has a high spontaneous incidence, or
(iv) an excess of malignant tumors is induced only by an irrelevant route of administration (see comments under Category 4 below), or
(v) the dose level required to produce tumors in experimental animals is so high that it adversely affects the normal physiology of the experimental animals due to its bulk or physicochemical properties [see comment below Category 5a i], or
(vi) no tumors are produced by exposure at or above those levels to which humans are likely to be exposed, even it tumors are found at exposures which produce chronic injury in the tissues or organ in which tumors later appear.

Supporting evidence. Information other than that from the long-term animal studies can influence the relevance to the human situation and aid a decision on categorization.

(a) *Metabolic data.* Studies indicating similarity in types and rate of metabolite formation can support decisions to infer the significance of data from other species to humans.

(b) *Short-term tests.* Positive results from a battery of properly validated short-term tests, while not adequate for classification when taken alone, may serve as supporting evidence for animal data and increase the confidence for classifying a substance as a proven animal carcinogen. Failure to find such supporting evidence could lead to a lowered classification.

(c) *Structure–activity relationship.* Structure–activity relationships (SARs) are not at present reliable enough to play a role in decisions which have implications for human health and they cannot be used to support the classification of a substance as a carcinogen, except in a few cases where particularly strong SARs have been established for a series of chemicals which are very closely related in structure.

Category 4: Suspected Animal Chemical Carcinogenic Substances of Potential Relevance to Humans

In this case there is limited evidence from animal studies carried out under conditions which are possibly relevant to humans, and there is no other strong supporting evidence. Some experimental results which might lead to this categorization include:

(a) A small increase in the incidence of malignant tumors of significance, e.g., where background data suggest that this incidence could have occurred by chance, particularly if this is only toward the end of the animal's natural life span.

(b) An increased incidence of malignant tumors in only one species or strain, which does not meet the biological and statistical criteria necessary to be classified as a proven animal carcinogen as described in Category 3.

(c) An increased incidence of malignant tumors only in organs for which the natural incidence is high or variable.

(d) Malignant systemic tumors are induced only by routes of exposure which are not relevant to human exposure (see comment below).

(e) Other information suggests that limited animal data are not relevant to humans, for example, when extensive and valid epidemiological studies have given no evidence of a carcinogenic effect.

(f) A treatment-related increase is observed only of nonmalignant tumors which are well known to progress to malignancy.

It should be noted that results from some routes of administration, e.g., subcutaneous, intravenous, intramuscular, and intraperitoneal treatment, are typically not reliable indicators of carcinogenicity. A very careful interpretation of results obtained by such exposure routes is necessary. The subcutaneous route particularly gives false-positive findings and much caution should be exercised in interpreting the data especially when only local tumor formation is induced. A careful interpretation of results obtained by intragastric intubation (gavage) is also necessary.

Category 5: Substances Nonclassifiable with Regard to Carcinogenicity

This category includes substances for which some experimental evidence exists, but the evidence is limited in strength and/or is irrelevant to the human situation. Some examples of such inadequate evidence are:

TABLE 1

COMPARISON OF CLASSIFICATION SCHEMES FOR CARCINOGENS

This document	EEC	EPA	IARC
1. Proven human carcinogenic substance	1	A	1
2. Suspected human carcinogenic substance	2	B^1	2A
3. Proven animal carcinogenic substance of potential relevance to humans	2	B^2	2B
4. Suspected animal carcinogenic substance of potential relevance to humans	3	C	3
5. Substances nonclassifiable with regard to carcinogenicity	—	D	4
6. Negative evidence	—	E	—
7. No evidence	—	—	—

Note. A direct comparison of categories is not always possible. This comparison is proposed as a reasonable approximation.

(a) The only positive evidence of carcinogenicity is from animal studies

(i) at excessively high doses which result in altered physiological conditions such as tissue damage (necrosis, chronic irritation) or a change in metabolic pathway, e.g., through overloading, to which the positive result is attributable (this phenomenon is illustrated by the case of carcinoma of the bladder epithelium induced by the presence of calculi); some groups have suggested dosages above which no practical significance should be attached (8);

(ii) under exposure conditions which do not occur when the substance is handled or used by humans, e.g., "solid state" sarcomas from plastic implantation; or

(iii) under exposure conditions in which the substance was used in a physical form to which humans are not exposed.

(b) The experimental evidence is equivocal, whereas there is valid negative evidence from epidemiological studies.

(c) The only positive evidence is from experimental studies which were improperly designed, improperly conducted, or could not be reproduced.

(d) Metabolic data indicate that the metabolic products or rates of formation of those products are grossly different in the experimental animal from those of humans.

(e) Failure to obtain supportive short-term results in cases where the animal data are not consistent with classification in Category 3.

Category 6: Negative Evidence

"This category is used for agent(s) that show no evidence for carcinogenicity in at least two adequate (and appropriate) animal tests in different species or in both epidemiologic and animal studies" (2). A more extensive discussion of this subject is provided in the epidemiology section.

Category 7: No Data

In this case no relevant data are available.

Table 1 is a summary of the comparison of the proposed classification system for carcinogens presented here and those of EEC, EPA, and IARC. It is appropriate to note that a material could be classified as having negative evidence of carcinogenicity when animal studies have been carried out under conditions which are relevant to human exposure and have positive evidence under inappropriate or exaggerated experimental conditions. The appropriate negative experimental and epidemiological evidence clearly supercedes the data collected from inappropriate studies.

The methodology for classification of carcinogens is not to be applied routinely in all situations, but is to be used as a guideline for classification, after which appropriate labels and warning statements may be developed.

CRITERIA FOR CLASSIFYING CHEMICALS AS MUTAGENS

Category 1: Evidence for Substances with Human Germ Cell Mutagenicity

There is sufficient evidence to establish a causal connection between human exposure to chemicals and heritable genetic effects.[2] This evidence can be developed only by appropriate epidemiology studies. For details refer to the "Criteria for Inferring Causality from Epidemiological Studies."

Category 2: Evidence for Substances with Mammalian Germ Cell Mutagenicity

There is evidence for mutagenicity and chemical interaction with the genetic material in mammalian germ cells *in vivo* by

(1) valid positive results from an *in vivo* mammalian germ cell study that measures mutations transmitted to offspring, or

(2) evidence that the chemical interacts[3] with the genetic material of mammalian germ cells *in vivo plus* clearly positive results in at least one valid *in vivo* study assessing either gene mutation or chromosome aberrations in somatic cells of humans or other mammals.

Category 3: Evidence for Substances with Somatic Cell Mutation in Mammals without Evidence for Germ Cell Interaction

Evidence of mutagenic activity is provided when a substance gives clearly positive results in at least one valid *in vivo* mutagenicity study assessing either gene mutation or chromosome aberrations in somatic cells of humans or other mammals.

[2] At this time, no causal relationships have been described.

[3] Artifacts such as incorporation of radiolabel through normal metabolites or tritium exchange should be excluded.

TABLE 2

COMPARISON OF CLASSIFICATION SCHEMES FOR MUTAGENS

This document	EEC	EPA
1. Evidence for substances with human germ cell mutagenicity	Category 1	—
2. Evidence for substances with mammalian germ cell mutagenicity	Category 2	Sufficient evidence
3. Evidence for substances with somatic cell mutation in mammals without evidence for germ cell interaction	Category 3	Suggestive evidence
4. Inadequate evidence for classifying mutagenicity	—	Limited evidence
5. Negative evidence	—	—

Note. A direct comparison of categories is not always possible. This comparison is proposed as a reasonable approximation.

This category is to be used if no evidence is available to permit one to determine whether interaction occurs with the genetic material of the mammalian germ cell *in vivo*.

Category 4: Substances with Inadequate Evidence for Classifying Mutagenicity

This category is for chemicals for which there are insufficient data available to meet the criteria described above, for example:

(1) positive test results only for endpoints without established clinical relevance and for which human genetic health hazards are not defined, such as DNA damage, sister chromatid exchange, and DNA binding in somatic cells, or

(2) positive findings of gene mutations or chromosome aberrations from *in vitro* studies only.

Category 5: Negative Evidence

In a weight of evidence approach to classifying chemicals it is important to distinguish between chemicals for which no evidence exists and those for which appropriate negative evidence is available. This category is therefore necessary to indicate that there is evidence that a chemical gives negative results in appropriate tests to measure gene mutation and chromosome aberration or that there is appropriate evidence to indicate that a chemical does not interact with the genetic material of mammalian germ cells *in vivo*.

These categories for mutagenicity show concordance with the EEC classification scheme and also share elements with the EPA categories of sufficient, suggestive, and limited evidence for mutagens (1, 2); see Table 2.

CRITERIA FOR CLASSIFYING CHEMICALS AS TERATOGENS

Three categories of substances are proposed in this document. This categorization, with respect to Categories 1 and 2, is consistent with that contained in the EEC Labelling Guide (1).

Category 1: Substances Known to be Teratogenic to Man

For Category 1, the accompanying descriptive phrase in the EEC Guide is:

> "there is sufficient evidence to establish a causal association between human exposure to a substance and subsequent non-heritable birth defects in the offspring."

We believe that to be consistent with the word "teratogenic", the above use of "birth defects" must relate to *deleterious, structural malformations in a conceptus* (see definition of teratogen).

It is proposed that before a chemical is classified as a Category 1 teratogen, at least the following criterion should be met: That a clear, unequivocal relationship between human exposure to a chemical during gestation and increased incidence of a specific structural malformation(s) has been established. Evidence for inclusion in this category is provided by formal epidemiological studies. (For details we refer to the section on epidemiology.) The background level of teratogenic manifestations, the difficulty in classification of specific malformations, and multiple confounding factors and sources of bias need to be rigorously addressed in the design of epidemiological studies of reproductive and developmental hazards.

It should be recognized that the influence of these factors will not be adequately controlled in many studies. Such studies will provide suggestive evidence only, inappropriate for classification in Category 1.

Category 2: Substances Which Should Be Regarded as if They Are Teratogenic to Man

For Category 2, the descriptive phrase in the EEC Guide is:

> "There is sufficient evidence to provide a strong presumption that human exposure to the substance may result in non-heritable birth defects in the offspring, generally on the basis of:
> —appropriate animal studies
> —other relevant information."

The comments on the interpretation of "birth defects" discussed under Category 1 teratogens also apply to the descriptive phrase for Category 2.

It is proposed in this document that *"appropriate animal* (i.e., mammalian) *studies"* should meet at least the following criteria:

(a) There should be a relevant route of exposure, i.e., for industrial chemicals, dermal, inhalation, or oral.

(b) There is exposure during pregnancy/organogenesis only.

(c) Structural malformations should be considered indicative of a teratogenic event only if they occur at exposure levels that do not cause overt maternal toxicity (e.g., a reduction in maternal body weight).

(d) The increase in incidence of structural malformations in comparison to that of controls must be statistically significant, and the malformations themselves must be of biological significance.

(e) Good historical background data should exist on the usefulness of the species tested and data should be for the same strain and from the same laboratory. The rat and rabbit are the preferred species.

(f) Adequate group sizes should be used, for example, 20 litters/group for rodents and 12 litters/group for lagomorphs should be available for analysis.

(g) Dose or treatment response should be demonstrable for defects that are not rare occurrences.

(h) Fetal assessment and data evaluation should be carried out using accepted statistical and laboratory methods.

In the evaluation of animal studies against the above criteria, particular consideration should be given to maternal toxicity as a confounding factor. Positive teratogenic findings seen only in the presence of maternal toxicity do not necessarily indicate specific hazard to the conceptus. By contrast, positive findings in the absence of maternal toxicity imply differential susceptibility between mother and conceptus. In any study, calculation of the ratio of a dose that is minimally toxic to the mother (A) to a dose that is minimally toxic to the conceptus (D) (A/D ratio) can be useful in assessing the extent of differential susceptibility.

The occurrence of teratogenic effects in studies meeting the above criteria in one mammalian species may be regarded as sufficient for a Category 2 classification of the substance concerned, unless there are clear reasons to doubt the relevance to humans.

If the above criteria are not met, then insufficient evidence exists to categorize the substance with respect to teratogenicity.

It is recognized that certain chemicals may cause adverse effects on development only at very high dose levels, whether in the presence or absence of maternal toxicity. In such a case careful interpretation is necessary to determine whether or not a hazard exists. When conducting a study or evaluating results of a study, an upper dose limit, e.g., 1000 mg/kg/day, should be considered as a dose which normally is of little practical relevance.

With regard to *"other relevant information,"* in the present state of science in developmental toxicology, the relevance of structure–activity relationships, behavioral data, and *in vivo* and *in vitro* screening procedures is not sufficiently well understood to justify use of such data for classification purposes.

Category 3: Substances Which Cannot Be Classified Based on Available Data

If available data do not meet the criteria described above for Categories 1 and 2, these substances should be considered *nonclassifiable*, i.e., there is inadequate evidence to assign a chemical substance to either Category 1 or 2.

CRITERIA FOR INFERRING CAUSALITY FROM EPIDEMIOLOGICAL STUDIES

A. Quality of Individual Epidemiology Studies

Before the criteria to infer causality are applied to a body of epidemiological studies, it is important that the quality of each study be evaluated as a guide to determine how much weight should be given to the study in the total assessment of the data. The criteria to evaluate individual studies are as follows:

(a) *Design of the study.* Analytic studies, such as case-control and cohort studies, provide tests of causal hypotheses. Proportionate mortality, cross-sectional, and descriptive studies are used to generate hypotheses. Therefore, only analytic studies are appropriate for inferring causality from epidemiologic evidence.

(b) *Definition of the cohort.* All exposed persons in the population being investigated should be identified. Criteria for inclusion or exclusion should be precisely defined. Justifiable exclusion would include short duration of exposure, such as less than 6 months, in the case of a carcinogenicity study.

(c) *Quantification of exposure.* Precise measures of intensity of exposure are difficult to obtain in retrospective studies, but qualitative estimates of exposure, such as "high," "medium," and "low," may be acceptable, provided they are uniformly applied and are consistent throughout the study period.

(d) *Mortality or morbidity ascertainment.* An evaluation should be made of the percentage of persons in the study population whose vital status and/or morbidity experience has been determined. The methodology of validating the diagnoses or causes of death should be assessed.

(e) *Identification of confounding factors.* The study should consider all possible alternative causes of the outcomes, or related factors, such as age, sex, race, and personal habits.

(f) *Statistical power.* Power is determined by the size of the study population and the incidence of the disease under investigation. The statistical power determines the extent of the confidence limits, and thereby provides a measure of the upper boundary of risk. A study with low statistical power could be acceptable, since it can contribute to the total body of evidence.

(g) *Statistical analysis.* The analysis should determine the probability that observed differences could have occurred by chance. Analyses should also be done to measure the strength of the association, establish confidence limits, assess the influence of confounding factors, analyze latency, and so forth.

B. Consideration of the Totality of the Epidemiological Evidence

It is important that all studies meeting the criteria described above should be taken into account, including both negative and positive studies.

Negative studies. Multiple high-quality negative studies cannot prove absence of risk, but the risk may be, at worst, so small that it can, for practical purposes, be disregarded. On the other hand, several studies that appear to be negative because the differences are not statistically significant may indicate the existence of a risk when the data are pooled.

Case reports and clusters. Strongly suggestive anecdotal or clinical observations may indicate a possible causal relation. Although such evidence may appear to demonstrate a risk, these observations should be followed by formal epidemiological studies to verify and quantify the risks, and to determine the role of confounding factors.

C. Criteria for Inferring Causality

Several sets of criteria have been developed. The following criteria are stated in the Preamble to the *IARC Monograph Programme on the Evaluation of the Carcinogenic Risk of Chemicals to Humans.*

(a) There is no identifiable positive bias.
(b) There is no positive confounding.
(c) The association is unlikely to be due to chance alone.
(d) The association is strong.
(e) There is a dose–response relationship.
(f) There is consistency in several independent studies.

The EPA has adopted the IARC criteria shown above (1).

ECETOC has not proposed any criteria to classify epidemiological evidence of carcinogenicity, but states:

> In most cases, excess cancer incidences can confidently be attributed to a selected chemical if this excess is consistent, specific, and clearly relates to particular exposure conditions. (10)

The National Toxicology Program (NTP) 3rd Annual Report on Carcinogens (1983) and EEC have also established categories to classify evidence of carcinogenicity, but have not proposed any criteria to assist in selecting the appropriate category. NTP's categories, however, are closely related to those established by IARC.

The IARC and EPA criteria have two shortcomings: (1) They do not include such important criteria as specificity, temporal relationship, coherence, and biological plausibility; and (2) they are merely stated, without elaboration and without any discussion as to how they should be used. Therefore, they are not sufficiently specific to ensure uniformity in their application.

The most comprehensive criteria that have been developed are those by A. Bradford Hill (20). We recommend that these criteria be adopted.[4] They are as follows:

(*1*) *The Strength of association.* This is usually expressed as some type of risk ratio, such as the death rate of a disease in an exposed population relative to the rate in a nonexposed population. The higher the risk ratio, the less likely it is that the association resulted from some confounding factor, and thus the more likely it is that the association is causal.

(*2*) *Consistency.* A causal hypothesis is supported when positive results are seen repeatedly in several studies done independently by different investigators using different populations. If only one or two studies have been conducted, causation can nevertheless be inferred if the risk ratios are very high.

(*3*) *Specificity.* This criterion refers to an "association limited to specific workers and to particular sites and types of disease and there is no association between the work and other modes of dying" (20). The relation between vinyl chloride monomer and angiosarcoma of the liver is one example. However, causality can occur in the absence of specificity.

(*4*) *Relationship in time.* This criterion requires that the exposure precede the development of the disease by a biologically relevant time period. A temporal relationship may be uncertain in a cross-sectional epidemiological study, but can be demonstrated in a properly conducted cohort study.

(*5*) *Coherence of the evidence.* This criterion is satisfied when the associations found in epidemiological studies do not conflict with what is known of the natural history and biology of the disease.

[4] Criteria 1 through 5 were adopted by the Advisory Committee to the Surgeon General of the US Public Health Service in preparing their report on smoking and health in 1964 (21). The criteria were also used in the Surgeon General's 1982 report on smoking and cancer (22).

(6) *The biological gradient.* To satisfy this criterion, the data should show a dose–response relationship; that is, the disease rate should be found to increase as the level and duration of exposure increase.

(7) *Biological plausibility.* A causal hypothesis is strengthened when there are known biological mechanisms that explain the association between the substance and the disease.

(8) *Experimental verification.* This criterion would be satisfied if removal of the substance or reduction in the level of exposure was eventually followed by a decline in the incidence of the disease.

APPENDIX 1

Participants in the Tripartite Working Party on Chronic Hazard Identification

Cochairmen

Dr. B. BROECKER: Hoechst AG; CEFIC/ECETOC
Dr. E. J. SOWINSKI: Uniroyal, Inc.; CMA/IAG

Carcinogenic Properties

Dr. J. M. FACCINI: Consultant; ECETOC
Mr. J. T. BARR: Air Products; AIHC
Mr. S. BEEKHUIZEN: Bayer AG; CEFIC
Mr. C. S. WEIL: Consultant; AIHC

Mutagenic Properties

Dr. P. GELBKE: BASF; ECETOC
Dr. H. FLEIG: BASF; ECETOC
Dr. D. W. MATHESON: Stauffer Chemical; AIHC
Mr. G. C. GRANVILLE: Shell Canada; CCPA
Dr. W. ROZENBOOM: Shell Internationale Petroleum; CEFIC

Teratogenic Properties

Dr. D. M. FERGUSON: ICI Americas; AIHC
Dr. I. F. CARNEY: ICI plc; CEFIC

Epidemiology

Dr. S. PELL: du Pont; AIHC
Dr. A. MUNN: Monsanto Europe; CEFIC
Dr. J. C. AUBRUN: Rhone-Poulenc; CEFIC
Dr. M. J. REAPE: FMC, Inc.; AIHC

Staff

Dr. H. HAESELER: CEFIC
Mr. G. W. INGLE: CMA
Mr. C. F. EMDE: CMA

CEFIC	European Council of Chemical Manufacturers Federations
CMA	Chemical Manufacturers Association (USA)
ECETOC	European Chemical Industry, Ecology and Toxicology Center
AIHC	American Industrial Health Council
CCPA	Canadian Chemical Producers Association
SOCMA	Synthetic Organic Chemical Manufacturers Association
IAG	International Affairs Group of CMA/SOCMA

REFERENCES

1. "Classification and Labelling of Dangerous Substances" (Directive 83.467/EEC). *Off. J. Eur. Econ. Communities (EEC)*: No. L 257/24; 16 September, 1983.
2. Environmental Protection Agency (EPA) (1984). Proposed guidelines for carcinogenic, mutagenic and reproductive risk. *Fed. Regist.* 49(227), 23 November.
3. U. S. Department of Labor: Occupational Safety and Health Administration (OSHA) (1983). Hazard communication; final rule. *Fed. Regist.* **48**(228), 25 November.
4. OSH Branch Labor Canada (April 1985). *Workplace Hazardous Materials Information System (WHMIS): Report to the Project Steering Committee.* Ottawa, Canada.
5. International Agency for Research on Cancer (IARC) (1984). *Monogr.* 33, April 1984, and Annual Report 1984.
6. National Toxicology Program (NTP) (September 1983). *3rd Annual Report on Carcinogens.* U.S. Department of Health & Human Services.
7. U.S. Department of Health & Human Services (1984). *Report of the NTP Ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation,* 17 August.
8. American Conference of Governmental Industrial Hygienists (ACGIH) (1984). *Threshold Limit Values (TLV) for Chemical Substances and Physical Agents in the Work Environment.* ACGIH, Cincinnati, Ohio.
9. ECETOC (September 1980). *A Contribution to the Strategy for Identification and Control of Occupational Carcinogens,* Monograph No. 2. Brussels.
10. ECETOC (January 1983). *Risk Assessment of Occupational Chemical Carcinogens* Monograph No. 3. Brussels.
11. ECETOC (December 1983). *Identification and Assessment of the Effects of Chemicals on Reproduction and Development,* Monograph No. 5. Brussels.
12. American National Standard for Hazardous Industrial Chemicals (1985). *Precautionary Labeling.* ANSI Z 129.1-1982, 21 September. New York.
13. Office of Science and Technology Policy (1985). Chemical carcinogens: A review of the science and its associated principles. *Fed. Regist.* 50, 14 March.
14. WEISBURGER, J. H., AND WILLIAMS, G. M. (1984). Bioassy of carcinogens: In-vitro and in-vivo tests. In *Chemical Carcinogens* (C. E. Searle, ed.), *Amer. Chem. Soc. Monogr.* **182**(2).
15. EVANS, H. J. (1984). Structure and organization of the human genome. In *Mutations in Man* (G. Obe, ed.), pp. 58–100. Springer-Verlag, New York.
16. United Nations Scientific Committee to Evaluate Energy and Radiation, Sources and Affects of Ionizing Radiation (UNSCEAR) (1977). United Nations, New York.
17. National Research Council; National Academy Press (1983). *Identifying and Estimating the Genetic Impact of Chemical Mutagens.* NRC/NAP, Washington, D.C.
18. World Health Organization (1984). Principles for evaluating health risks to progeny associated with exposure to chemicals during pregnancy. Environmental Health Criteria No. 30. WHO, Geneva.

Appendix II

Comparison of Approaches for Classifying Chemicals as Carcinogens, Mutagens, or Teratogens with EEC, EPA, OSHA, and WHMIS

EEC	EPA	OSHA	WHMIS
		Carcinogens	
Category 1: Known to be carcinogenic to man *Category 2:* Regarded as carcinogenic to man based on animal studies *Category 3:* Cause concern owing to possible carcinogenic effects	*Group A:* Human carcinogen *Group B:* Probable human carcinogen *Group C:* Possible humn carcinogen	No categorization: any chemical in A or B is covered (A) Listed in NTP Annual Report, in IARC monographs 1 or 2, or regulated by OSHA (B) Based on hazard evaluation of other relevant data	No categorization: Human or animal listed under ACGIH; A1a, A1b, A2; also IARC Groups 1, 2A, 2B are covered
		Mutagens	
Category 1: Known to be mutagenic to man *Category 2:* Regarded as mutagenic to man *Category 3:* Cause concern owing to possible mutagenic effects	Sufficient evidence Suggestive evidence Limited evidence	No categorization: Based on hazard evaluation of all relevant data	Not addressed at this time
		Teratogens	
Category 1: Known to be teratogenic to man *Category 2:* Regarded as teratogenic to man	No specific categorization	No categorization: Based on hazard evaluation of all relevant data	No categorization; positive in OECD 414, 415, 416 are covered

19. Commission of the European Communities (DG XI) (1985). *Report of 2nd Meeting of Specialized Experts in the Field of Carcinogenic/Mutagenic/Teratogenic Substances*, 18/19 April, Brussels.
20. Hill, A. B. (1971). *Principles of Medical Statistics*, 9th ed, pp. 309–320. Oxford Univ. Press, New York.
21. Advisory Committee to the Surgeon General of the Public Health Service (1964). *Smoking and Health*, pp. 182–189. U.S. Public Health Service.
22. Report of the Surgeon General (1982). *The Health Consequences of Smoking—Cancer*, pp. 16–20. U.S. Public Health Service.
23. International Commission for Protection against Environmental Mutagens and Carcinogens: Committee 1 (1983). Final Report: Screening strategy for chemicals that are potential germ-cell mutagens in mammals. *Mutat. Res.* 114(2), 117–177.

FUNDAMENTAL AND APPLIED TOXICOLOGY 7, 354–357 (1986)

LETTERS TO THE EDITOR

To the Editor:

Haseman (1986) invokes several well-traveled arguments in defense of long-term rodent carcinogenicity testing in general and the use of maximum tolerated doses (MTDs) in particular which, despite frequent usage in regulatory circles, cannot withstand serious scientific scrutiny.

First, the fact that most human carcinogens have been found to be carcinogens in rodents is not proof that the reverse is true. We all were taught in Philosophy 101 that while all blackbirds are black birds, not all black birds are blackbirds. There may be a strong suggestion that many of them are, particularly if one operates in an area heavily infested with blackbirds, as the NTP does, but that is not scientific proof, nor can it be, by the basic rules of science.

Further, there is very poor agreement among the various sexes, strains, and species as to whether a given substance actually is a rodent carcinogen, even at the MTD (Haseman *et al.*, 1984). Of the 48 NTP bioassays deemed positive by the authors that are discussed in that paper, 5 had only equivocal evidence in one or two sexes of one species (usually the mouse), 13 had stronger evidence, but still only in one sex of one species, 14 had positive evidence of carcinogenicity in both sexes, but only one species, and 3 others had disagreement by one sex of one species with the 3 others. Thus, only 12 of 42 clearly positive studies, 28.5%, had agreement between both sexes in both species, even with inclusion of benign tumors. Twenty-seven of 42 studies, 64%, had disagreement between the mouse and the rat as to rodent carcinogenicity. There was, therefore, also 64% disagreement as to the prediction of human carcinogenicity. This two out of three error rate is more than would be expected from tossing a fair coin. Our efforts should be directed at determining which, if either, is the correct predictor, and not at perpetuating specific test methods.

Second, the existence of a threshold is rejected on the basis of population variability. This ignores any information on the mechanism of carcinogenesis, and is a philosophic argument again. Complete safety, zero risk, or absence of probability of harm cannot be shown by statistical methods, which do not contemplate zero, but only an asymtotic approach to it. Biological methods are required to understand life processes, and statistics should be used as a servant, not a master. There are numerous instances now recognized where nongenetic processes forecast no-effect levels. Space does not permit a longer review of that point here, (see, e.g., Weisburger and Williams, 1985) but the OSTP Policy (1986) in a section not cited by Haseman, says, in part, in Principle 26: "No single mathematical procedure is recognized as the most appropriate for low-dose extrapolation in carcinogenesis. When relevant biological evidence on mechanism of action . . . exists, the models or procedure employed should be consistent with the evidence." The remainder of Sec. V, Principles of Risk Assessment, emphasizes the need for a weight-of-the-evidence approach, with facts, assumptions, and policy decision being identified clearly and their effects being analyzed.

In any event, low dose linearity, which is a prerequisite for the absence of a threshold,

0272-0590/86 $3.00

can be expected only on those relatively rare occasions where the test substance induces the same lesion in the same cells of an organ which exhibits an "appreciable" (more than 10%) incidence of spontaneous tumors (Peto *et al.*, 1985). These authors state that "This means that if this background is appreciable then a low dose-range should exist in which simple proportionality between dose and effect is actually observable, but that if this background incidence is negligible then no general prediction follows of the shape of the dose–response relationship."

We will continue to see opinions, hypotheses, and inference options (NRC, 1983) used in scientific publications and risk assessments, as they must be. They should be identified as such, however, and their effects understood. Further, their value as a scientific hypotheses should be judged.

The necessity for the testability of, and the testing of hypotheses cannot be denied by a scientist. Denial of the very existence of a hypotheses, whether it be the possibility of the existence of a threshold, or the question of the utility of a rodent as a surrogate for humans, is therefore not a scientific posture.

REFERENCES

HASEMAN, J. K. (1986). Letter to the Editor. *Fundam. Appl. Toxicol.* **6**, 186–188.

HASEMAN, J. K., CRAWFORD, D. D., HUFF, J. E., BOORMAN, G. A., AND MCCONNELL, E. E. (1984). Results from 86 two-year carcinogenicity studies conducted by the National Toxicology Program. *J. Toxicol. Environ. Health* **14**, 621–639.

National Research Council Commission Life Sciences (1983). *Risk Assessment in the Federal Government: Managing the Process.* National Academy Press, Washington, D.C.

Office of Science and Technology Policy (1985). Chemical Carcinogenesis: A review of the science and its associated principles. *Fed. Reg.* March 14, 10371–10442.

PETO, R., GRAY, R., BRANTOM, P., AND GRASSO, P. (1982). *Effects on Two Tonnes of Inbred Rats of Chronic Ingestion of Diethyl- or Dimethyl-Nitrosamine: An Unusually Detailed Dose–Response Study.* April, pp. 39–40. British Industrial Biological Research Association, Surrey SM5 4Ds.

WEISBURGER, J. H., AND WILLIAMS, G. M. (1985). Controversies in carcinogenesis: Health risk analyses for carcinogens. In *Safety Evaluation and Regulation of Chemicals* (F. Homberger, ed.), Vol. 2, pp. 150–165. Karger, Basel.

JOHN T. BARR

3926 Lilac Road
Allentown, Pennsylvania 18103

REGULATORY TOXICOLOGY AND PHARMACOLOGY 5, 432–459 (1985)

The Calculation and Use of Carcinogenic Potency: A Review

JOHN T. BARR

Air Products and Chemicals, Inc., Box 538, Allentown, Pennsylvania 18105

Received June 19, 1985

Six methods of estimating carcinogenic potency now in use to some extent by regulatory agencies are examined. It is concluded that none of these methods is adequate for regulatory decision making when used alone, as is usually the case, but that all are useful as a contribution to the whole risk assessment process. The needs for further development and improvement of these methods, where appropriate to human risk, are identified and discussed in view of the growing use of potency estimates in regulatory decision making. © 1985 Academic Press, Inc.

INTRODUCTION

The need for a satisfactory method of estimating the carcinogenic potency of substances has increased with the growing use of risk assessment in reaching regulatory decisions. In most risk estimates, risk is taken as the product of exposure and potency, or, as it is expressed in some instances, dose and hazard (Office of Science and Technology Policy (OSTP), 1985). Pressure to regulate, or at least to evaluate for regulation, substantial lists of materials within fixed schedules increases the need for quick and simple estimates for both potency and exposure factors. This paper describes some of the methods suggested for deriving potency factors, discusses their applications, and evaluates their usefulness.

METHODOLOGY

Several methods have been used or proposed for use. Six of these are discussed, with examples of their use or potential use.

1. Epidemiology

The preferred data on which to estimate risks for humans come from human experience (OSTP, 1985). Epidemiology, therefore, would be the best method for setting potency if it were possible to use it in that way. Two well-known efforts to do so have failed because of lack of adequate past exposure data and other problems. Kuzmack

0273-2300/85 $3.00

and McGaughy (1975) prepared for the EPA an estimate of cancer deaths in the general population from exposure to ambient vinyl chloride near vinyl chloride/polyvinyl chloride (PVC) plants. Their potency value was derived in part from occupational death rates, and their projected ambient exposures were estimated from air modeling studies. Unfortunately, only those occupational sites where deaths from this cause had occurred were used while ignoring all other data, and the releases to the ambient were overestimated significantly. Their estimate of 20 deaths per year, later reduced to 11 (Anderson, 1983), has not been confirmed (Popper *et al.*, 1978). No deaths due to ambient vinyl chloride exposure have been found in the more than 40-year industrial history. That fact alone reduces the value of the potency figure as estimated by the EPA by more than three orders of magnitude, but does not allow setting a more specific figure with any accuracy (Barr, 1982).

Similarly, Occupational Safety and Health Administration (OSHA) has attempted to use estimates of benzene potency derived by Infante and co-workers (White *et al.*, 1982) from occupational epidemiology in its regulatory process for benzene. Here, too, debates over actual historical exposure data and selection of the proper cohort have questioned the value of that estimate as a measure of human hazard (International Agency for Research on Cancer (IARC), 1982).

Epidemiology can, as many have pointed out (Monson, 1980; Doll, 1984), set an upper limit on the value of potency. The accuracy, and therefore the value of that upper limit, depends on the accuracy with which the past exposures can be reconstructed and the statistical power of the study. Regulatory agencies have reported epidemiology-based estimates on occasion (EPA, 1979), but usually subordinate them to values estimated for bioassays, as was done in the case of vinyl chloride (Kuzmack and McGaughy, 1975). In no case can epidemiology, or any other statistical exercise, demonstrate an absence of carcinogenicity (Doll, 1984), and thus epidemiology is unlikely to have a major impact on risk assessments for the low exposures to the weak carcinogens now experienced in the environment.

Epidemiology does permit the estimation of upper limits of potency, as mentioned above, and thus can serve as a broad screening tool. With proper attention to detail this step could assist in placing some risks in proper perspective, for as Doll (1984) pointed out "a time comes in the collection of human data when the existence of any material risk can be postulated only with such heavy exposure that the normal exposures to which man is exposed can be regarded as essentially safe."

Epidemiology studies also can be used to evaluate the accuracy of predictions made from animal studies and to compare the relative potencies of suspect carcinogens.

For example, the potency of vinyl chloride at high occupational exposures can be calculated by using the data for angiosarcoma in the 25,000 or so workers exposed to vinyl chloride in the PVC industry in the United States. At least 10% of these were very heavily exposed as reactor cleaners to an average of at least 1000 ppm. The average time from first exposure to diagnosis has been 25 years for the 35 cases that have been seen so far in this country. Certainly no more, and probably fewer, than another 50 or so cases can be expected from past high exposures. Thus

$$\frac{85}{2500} = P\frac{1000 \times 25}{70}\ \text{ppm}$$

and

$$P = 9.5 \times 10^{-5}\ \text{ppm}^{-1}.$$

Similarly, the data provided by Olsen *et al.* (1984) can be used to calculate a maximum potency for formaldehyde. They report one of the few epidemiological studies which finds a measurable, although not statistically significant, elevation of 1.66 for the rate of nasal cancer in occupationally exposed persons. OSHA (1984) suggests that the most common occupational exposure has been around 1 ppm. National Cancer Institute (Riggan *et al.*, 1983) data show that nasal cancer in this country, although falling sharply over the last three decades, averaged about 575 cases per year in 1970–1979. This would predict a potency of $(1.6 \times 575 \times 70)/230 \times 10^6$, or $P = 2.8 \times 10^{-4}$ ppm^{-1}.

These two rough estimates round off to 10^{-4} and show that vinyl chloride and formaldehyde might be expected to have about the same potency as human carcinogens if both were, in fact, carcinogenic at ambient exposures. Vinyl chloride does not appear to be, because of the negative epidemiological data for vinyl chloride (Popper *et al.*, 1978). This fact sets an upper limit of potency of about 10^{-7} ppm^{-1} at ambient exposures of a fraction of a part per million, as compared to the 10^{-4} value calculated at occupational exposures of over 1000 ppm, and at least suggests that the dose response is not linear at low levels. By contrast, several hundred cases over the last few decades were predicted in the general population by linear extrapolation of animal data (Kuzmack and McGaughy, 1975; Anderson, 1983).

Similarly, the risk estimates prepared by the regulatory agencies from formaldehyde data on animals would suggest human nasal cancer rates many times that actually found in areas with significant occupational exposure (EPA, 1984e; OSHA 1984). Here, too, repeated failure to confirm these estimates by epidemiology suggests that the animal-based estimates must be an exaggeration of the human situation (Hoel *et al.*, 1983).

At the other extreme, very potent substances such as bischloromethyl ether provide abundant confirmation of their carcinogenic power in epidemiological results (Weiss, 1982).

Continued difficulty in obtaining epidemiological support for dubious animal results can be useful and reassuring information for those responsible for risk management. It can be expected that more such analyses of human data can be useful in evaluating the confidence to be placed in risk assessments based on nonhuman data and to help establish regulatory priorities (OSTP, 1985). The Science Advisory Board (1985a) has encouraged the EPA to make these analyses as a part of its analysis of uncertainty in the risk analysis program.

Salsburg (1983) and Doll (1984) remind us that only about half of the agents now classed as human carcinogens by IARC have been, or could have been identified, by the standard test programs now in use. Therefore, epidemiology remains a valuable, if underutilized, tool.

2. *Bioassays*

The most widely developed and accepted method of estimating potency is to calculate a dose–response curve from lifetime animal bioassay data.

Meselson and Russell (1977) have defined potency as

$$K = \ln 2/D_{1/2},$$

where $D_{1/2}$ is the daily animal dose which gives a 50% incidence of cancer after a 2-year exposure. They further state that K may be derived from the results of the Ames test, because the ratio of m/K is approximately 1 in many cases, where m is the dose (in μg) reguired for 100 Salmonella revertants.

Crouch and Wilson (1979, 1983) defined potency as the slope of the linear, no-threshold extrapolation of an animal bioassay result, that is, as β in the equation

$$R = -\ln(1 - p) = \alpha + \beta d$$

where p is the probability of developing cancer at dose d, and α and β are derived constants. They also suggested that because the interspecies variation of potency between rats and mice is small in some examples, the animal-derived data should be applicable to humans also. This is essentially the same relationship used by the Cancer Advisory Group to calculate its unit risks (Anderson, 1983). Their results are slightly different than those of the EPA because of differences in the data bases used and the mathematical extrapolation model.

Bernstein and co-workers (1985) have pointed out, however, that the apparent correlation of potency between rats and mice reported by Crouch and Wilson (1979) is an artifact of the experimental design, and not necessarily because of the similarity of the species, and thus casts doubt on the ability to extrapolate the data even further to humans.

In at least one instance the Office of Toxic Substances of the EPA (1982) has combined the estimates prepared by the two quite different procedures of Meselson and Russell (1977) and Crouch and Wilson (1979) into a single table in order to estimate the risks of workers exposed to potentially harmful chemicals. The difficulties inherent in the underlying assumptions for these proposals, such as the many unvalidated leaps of faith in making both qualitative and quantitative estimates for humans from bacterial and rodent data, are apparent, as are the problems which arise from the combination of the results that are obtained by these two quite different methods.

The National Academy of Science (NAS) (1980) recommended that the EPA develop "indicators of relative pathological activity" of pesticides from bioassay data as an aid to the overwhelming regulatory task which it faced. The committee also recommended that the EPA abandon its attempt to produce numerical estimates of the effect of the use of pesticides on humans except when reliable human epidemiological data are available. If this recommendation has been followed, the results have not been published. Regulatory agencies generally have adopted a computerized method of evaluating bioassay data (Anderson, 1983). This program assumes a linear, no-threshold relationship and calculates the upper 95% confidence limit curve connecting the origin and the lowest one or two data points. The program rejects 0 or negative data, and thus skews the calculated curve toward higher inputed risks than would a program accepting and processing all available data (Mantel, 1985). Crouch *et al.* (1984) and Sielkin (1985) suggest that the result is closer to a 99.9% upper confidence limit. The slope of the resulting curve is taken as the unit risk, which then is converted into a potency index by multiplying the slope, expressed in milligrams per kilogram per day, by the molecular weight of the substance. The potency index is further transformed into an order of magnitude value by extracting the mantissa of the logarithm of the potency index. The agency suggests that this value is not accurate to more than one significant figure (EPA, 1985a).

The agency publishes a table of relative potencies, along with a description of its procedure, in each Health Assessment Document it prepares. At present there are 54 substances on the list, with the log index values ranging from −1 to 8 (EPA, 1985a). The agency has stated, however, that it has not concluded that all of those substances are human carcinogens, but has prepared the potency index nevertheless (EPA, 1984a, 1985a). EPA (1984a) has presented a discussion of the philosophy underlying the evaluation of the animal data and a proposed guideline for the use of such data in estimating human risk.

Others have suggested the use of an index of carcinogenic potency, represented by TD_{50}, the daily dose which produces 50% deaths from tumors at the end of a lifetime feeding study (Ames *et al.*, 1982; Peto, *et al.*, 1984; Sawyer, 1984).

Ames and co-workers (1982) first proposed the calculation of a lower confidence limit TD_{50} from negative bioassays, based on the parameters of the study, much the same as can be done from epidemiology results. Using positive National Cancer Institute (NCI) bioassays, they calculated a range of potencies greater than 10^7 from the results for female mice.

A later paper (Peto *et al.*, 1984) suggested that TD_{50} be based on the standardized results of all bioassays and proposed conventions for addressing the variation in, or absence of, data in the published results. This group has published (Gold *et al.*, 1984) a Herculean review of data from about 3000 long-term studies on 770 substances. Potency values, as the TD_{50}, were calculated for each study, with a statistical confidence limit and other useful information. They confirmed the 10^7 range of potencies, but also found up to a 10^3 range of potencies within the data for single substances (e.g., vinyl chloride) for which a large number of studies have been reported. This same range of estimates is seen (Barr, 1982) when full-scale risk assessments are performed with the results of vinyl chloride bioassays. This variation has been used by OSHA (1980) as a reason for not considering the risk assessment methodolgy valid or useful. It is, rather, a strong reason for the need for careful and sound scientific reasoning in the performance of risk analyses.

Clayson (1983) proposed that potency should be represented by a figure derived from the relationship $7 - \log DE_{50}$, where DE is the 50% tumor incidence dose expressed in micromoles per week per kilogram.

He identified five components of potency: the probability of tumor formation, the dose rate of the carcinogen, the time to tumor, the extent (multiplicity) of tumor formation, and the quality of pathologic diagnosis. The first three issues were stated to be the most important, and were used in his definition. This method produced potencies ranging from 9.2 for aflatoxin to 1.9 for saccharin. He also calculated the potency of 4-aminobiphenyl to range from 6.2 in dogs to 4.4 in mice.

Clayson (1983) separated the carcinogenic process into the initiation and promotion stages. He expressed the potency for initiation as $X\text{-log } ad_m + R$, where d_m is the dose required to initiate a standard number of events, and X and a are constants used to align the results with *in vivo* studies. R is an expression of the repair capacity of the species under study.

Similarly, potency for promotion, or development, was described by $Y\text{-log } b\, d_p\text{-log } c\, d_s$, where Y, b, and c are constants, d_p is the dose required to produce a defined proliferative stimulus, and d_s is the dose required to produce a given degree of selection. Combining the two expressions he derived the combined potency as $f(R\, d_m \cdot d_p \cdot d_s)$. He pointed out that methods of measuring d_p and d_s are not yet available.

Crump (1984) has proposed the use of a "benchmark dose" as an alternative to an observed no-effect dose in determining allowable daily intakes of chronic toxins. These benchmark doses, when compared to each other, could serve as an index of potency for the particular effect under study. He suggests a curve-fitting process to determine if there is an indication of an intersection or discontinuity in the dose–response curve. If one is found, that point is taken as the dose to which an appropriate safety factor is applied to estimate allowable intake levels. The procedure is recommended for occasions when the lowest experimental dose did not appear to be completely without toxic effect, or when the lowest dose appeared to be far below the acutal no-effect dose, and, therefore, would produce an unnecessarily low acceptable dose. It is, in effect, a way of searching mathematically for apparent thresholds for chronic effects, rather than assuming none exist. Crump recommended this procedure for use with chronic health hazards and not specifically for carcinogens.

The American Conference of Governmental and Industrial Hygienists (ACGIH) (1984) has established three categories of potency for animal carcinogens based on the dosage required to "elicit cancer" in the study. High potency ranking requires statistically significant response below 1 mg/m^3 inhalation exposure for 6 or 7 hr/day over a lifetime, or from a single intratracheal dose of not more than 1 mg of particulate per 100 ml of minute volume, or from twice weekly skin paintings for up to 20 weeks from no more than a total dose of 1.5 mg, or within 6 months after a 6-month dosing in food of less than 1 mg/kg/day. Low potency carcinogens are defined as those which result in a response after receiving more than 10 mg/m^3, or 10 mg of particulate, or a total skin dose of greater than 1.5 g over 75 weeks, or greater than 50 mg/day in food. Substances of intermediate potency result in a response at doses between those listed. Substances requiring an inhalation dose greater than 1000 mg/m^3 for the mouse or 2000 mg/m^3 for the rat, or dermal doses greater than 1.5 g for the mouse or 3.0 g for the rat, or lifetime feeding doses greater than 10 g for the mouse or 100 g for the rat are excluded from consideration as carcinogens of any practical significance. These potency classifications are then considered in setting recommended exposure limits for occupational settings, along with the degree of evidence of human carcinogenicity (ACGIH, 1984). To the extent that ACGIH recommendations become incorporated into OSHA standards, this is one of the few cases where the potency concept actually is reflected in setting exposure limits at the federal level. Pragmatic groupings of substances into categories with similar potencies is a very useful device for the workers in occupational hygiene (Crabtree, 1983).

EPA (1985c) has announced a plan to use potency values to establish reportable quantities (RQs) under the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA or "Superfund"). This plan (Combustion Engineering, 1985) takes as the potency factor the reciprocal of the estimated dose that will cause a 10% mortality (ED_{10}). These results were grouped into three classes, and appropriate RQs are being considered for these classes.

This is a different manner of presentation than that used in the potency index (EPA, 1985a), but the two values are related. The ED_{10} can be derived from the slope of the dose–response curve used in calculating the potency index by the expression $0.1 = \text{slope} \cdot X$, where X is the dose in question, or ED_{10}. There is no intercept in this linear equation becaue of the EPA determination of no thresholds for carcinogens. Thus, the slope of the dose–response curve divided by potency factor (the reciprocal of ED_{10}) should equal 0.1. Comparison of the data of the EPA (1985c) and Combustion

Engineering (1985) reveals that many of the data do show a relationship between 0.06 and 0.30. Surprisingly, however, of the 39 substances on both these lists, 9 exhibit a ratio more than one order of magnitude away from the average ratio, some as much as two orders of magnitude in either direction. For example, benzidene gave a ratio of 123, while carbon tetrachloride gave 0.003, a spread of 41,000. Thus, these two potency rankings published within 2 months of each other show an inconsistency range of nearly five orders of magnitude. Such lack of consistency within the same office of the agency is difficult to explain, and casts serious doubts on regulatory decisions made as the result of such procedures.

Risk assessment, and therefore potency values, has been used in attempts to quantify the risk under given circumstances, and thus the reduction in risk resulting from specific controls (OSHA, 1978). Risk assessments also have been used fairly recently in decisions not to regulate certain pollutants (EPA, 1984b) or classes of sources (EPA, 1984c). But, given a decision to regulate, the degree of regulation customarily has been controlled by technology and not by acceptable risk. California (CDHS, 1985) currently is developing a policy for carcinogens in which it is stated that a measure of potency is necessary to distinguish between high and low risks, to determine if a specific exposure is actionable, to help in development of exposure standards, and to help set priorities.

The Consumer Product Safety Commission (1983) ranked a number of consumer-related substances in order of potency, but expressed many caveats as the practical utility of the listing, and apparently have done no more with the results.

The U. S. Department of Agriculture (1985) has developed a revised compound ranking scheme in which separate rankings of hazard and exposure are given rather than a single combined evaluation. This is in effect working backward from a risk assessment to its components.

Potency estimates based on bioassays are limited in accuracy by the design and execution of the assay, as well as by any assumptions used or errors made in the manipulations of the data during the calculation of the estimate. Suggestions have been made that, if the standard bioassay is to be used for risk assessment, adjustments should be made in the number of doses and the relative size of the exposure groups (Hoel and Jennrich, 1979; Portier and Hoel, 1984; NTP, 1984; Gaylor *et al.*, 1985). However, unless some better adjustment is made for interspecies differences than the customary application of an arbitrary weight or surface area factor, the best than can be said for the result of such a risk assessment is that it applies to a very large rodent. Adequate consideration of comparative pharmacokinetics and other factors is needed to obtain significance for other species (Calabrese, 1983).

Prehn and Lawler (1979) report that the rank order of potency of 3-methylcholanthrene among 10 strains of mice reverses when the dosage is reduced 100-fold. They ascribe this effect to the biphasic nature of the immune response, where a weak stimulus induces tumor growth but a strong challenge invokes protective action. The degree of change in these responses can vary with species and strain of rodents. It is recognized that there is very poor concordance between mice and rats in the standard bioassay (von Whittenau and Estes, 1984), and the National Toxicology Program (NTP) (1984) is reconsidering whether the mouse bioassay should be retained.

Further, use of 50 percentile mortality rates, or any other single-point datums, ignores differences between the slopes of the dose–response curves for the substances being compared. Van Ryzen (1980), Park and Snee (1984), and others have recom-

mended that in order to minimize errors arising from the model for this reason, extrapolation of the response curves by some suitable model be carried to a 1 to 10% repsonse level before a safety factor is applied. The data of Lijinsky and Reuber (1984), illustrating variations in potency with dose, support the value of this suggestion, as does the statement of the Science Advisory Board (1985b) that rate of dose is of importance in quantifying the carcinogenicity of chromium (VI).

Monograph No. 3 from European Chemical Industry Ecology and Toxicology Centre (ECETOC) (1982) discusses the difficulty of attempting to express the dose, which is an independent parameter, and intensity and incidence, dependent parameters, into a single figure, and conclude that an accurate potency value cannot be determined for that reason. Also, the OSTP has pointed out (1985) that relative potency calculations based on a standard measure of toxicity "attempt to compress an entire dose–response curve into a single number" and thus "may offer little insight into the relative risk in the low dose region."

There are many respected scientists who question the relevance of the bioassay as now conducted to human risk. For example, Lederberg (1981) has said:

> First of all, understand that one or two or three hundred millions of dollars a year that we're now spending on routine animal tests are almost all worthless from the point of view of standard setting. It may be appropriate for setting alarms, but once you get a bladder cancer in a certain percentage of male mice who have been through two generations of saccharine treatment, what then? Is the male mouse a good model for the male human . . .? I would think the most immediate solution is to redeploy some of our resources; and the resources are not only money, there is the time and effort.

Golberg (1985) has expressed the hope that "'anachronisms such as the carcinogenesis bioassay and the MTD (maximum tolerated dose) will finally be relegated to the Smithsonian Archives as relics of the Dark Ages of Toxicology." IARC (1984) has concluded:

> In the present state of knowledge, it would be difficult to define a predictable relationship between the dose (mg/kg bw per day) of a particular chemical required to produce cancer in test animals and the dose that would produce a similar incidence of cancer in humans. Some data, however, suggest that such a relationship may exist, at least for certain classes of carcinogenic chemicals, although no acceptable method is currently available for quantifying the possible errors that may be involved in such an extrapolation procedure.

A major factor in calculating potency from bioassay studies, as well as from many epidemiology studies, is the lifetime dose received by the individuals or animals. However, upon reflection it is obvious that lifetime dose is irrelevant, and that it is only the dose received before the last irreversible dose-related stage in the initiation–promotion series of steps that has any relevancy to the relative potency of the substance. Once this last irreversible step has occurred, there is no further effect on the malignancy. Current studies make no effort to determine when such a step occurs, and in many cases there currently is no way of determining what the event is or when it occurs. Thus, as Lederberg (1981) states most of our present efforts are worthless because the fundamental concept of latency is ignored.

Nevertheless, the rough estimates which can be prepared solely from crude bioassay data do have a place in regulatory processes. It is appropriate to reduce the level of concern for a substance found not to have a significant potency rating, and then to examine further those with higher values (Barr *et al.*, 1981; Doll, 1984).

3. Animal Skin Painting

The convenience, speed, and economy of animal skin-painting tests, particularly with mice, have led to its widespread use as a screeing test, and has made a particurlarly valuable contribution to our knowldege of promoters and their mechanisms. Pitot (1983) has prepared a concise review of early results in this field. Protocols have been developed which allow comparison of time to tumor appearance after treatment with various substances relative to that for a standard substance, often benzo[*a*]pyrene, thus leading to the calculation of relative potencies for the substance (Holland *et al.*, 1979, 1981).

The first tumorigenic response to skin painting is the appearance of papillomas at the treatment site. Progression to malignant expressions, such as squamous cell carcinomas, occurs in a relatively small fraction of events (Hennings *et al.*, 1983; Furstenberger *et al.*, 1983) and does not occur with "pure" promoters. Initiators are effective in causing this progression. These facts support the two-stage mechanism of promotion and limit the value of much of the skin-painting data as far as its relevance to other routes of exposure. However, Hennings and Yuspa (1985) have questioned many of the shibboleths now current in the skin-painting field, and called for a re-evaluation of many of the present concepts.

Regulatory agencies generally do not accept tumorigenic response in the skin only at the application site as indication of human carcinogenicity (OSHA, 1980). However, some substances can cause carcinogenic response at remote sites after skin application. Tobin *et al.* (1982) examined the 247 chemicals designated by IARC as having some degree of evidence for carcinogenicity, and found skin-painting data for 51 of them. Of the 51, 36 did produce skin tumors. Of those 51, 20 had been studied by necropsy of the animals, 10 of which were positive and 10 of which were not. Of the 10 positive substances, 5 produced lesions at other sites. Of the 10 negative substances, 3 produced lesions at other sites. Thus, 7 of 51 substances supposedly positive by IARC standards gave responses at neither the skin nor other organs. In addition, 15 others did not produce skin tumors. The authors concluded that "animal skin-painting tests, applied according to current protocols, cannot be conclusively relied upon to determine the potential carcinogenicity of applied chemicals."

A potential conflict could arise with many regulatory philosophies if the results of skin-painting tests were adopted as a measure of carcinogenicity. Workers in this field consistently use "subthreshold" or "noncarcinogenic" doses (Hennings *et al.*, 1983; Furstenberger *et al.*, 1983) to prepare animals for tests of promotional capacity. Further, classical studies have shown the reversibility of the initiation step (Pitot, 1983). The concepts of threshold and reversibility, as well as the problem of nonprogression of benign tumors, are in direct conflict with the philosophical positions of many regulatory agencies, and it would cause serious credibility problems for them if potencies from skin-painting data were accepted into existing schemes.

4. In Vitro Data

The ultimate value of *in vitro* data, if they prove useful as indicators of carcinogenicity, would be as a substitute for bioassay data in potency estimation. Several workers have tested this application by calculating potencies, and in some cases, testing the results against bioassays or other indicators of carcinogenicity.

Ames and co-workers have commented (McCann *et al.*, 1975, 1976; Ames and Hooper, 1978) on an apparent relationship between the early results of that test and carcinogenic potency in animals. A spirited discussion of that suggestion occurred in *Nature* (Ashby and Styles, 1978; Ames and McCann, 1981) and elsewhere (Heddle and Bruce, 1977; Hollstein *et al.*, 1979; Dunkel, 1979; Bridges *et al.*, 1981) during which Ames stated that "this kind of correlation is never going to be very precise or likely to hold for every type of chemical" (Ames and McCann, 1981). Rinkus and Legator (1979) joined the discussion and extended the examination of the utility of the test (Rinkus and Legator, 1981). These and many subsequent studies (De Flora *et al.*, 1984; Haworth *et al.*, 1983; McCann *et al.*, 1983) have explored the limits and utility of the Ames test. Legator and Harper (1982) concluded that the tests "cannot be used as a primary screen or to quantitate potency," and an ECETOC committee (1982) stated that these are "good scientific reasons" for not using mutagenicity data in risk assessments. OSTP (1985) concludes that "short-term tests are presently limited in their ability to predict the presence or absence of carcinogenicity and cannot supplant data from long-term animal studies or epidemiological investigations." They do cite short-term tests as useful in screening and in deducing mechansims of action. Most regulatory guidelines accept *in vitro* data as supportive of other positive results.

These positions appear justified when quantitative examinations of correlations between *in vitro* and bioassays studies are examined.

Many studies have reported the failure of short-term tests to identify certain classes of animal carcinogens (Ashby and Styles, 1978; Legator and Harper, 1982). A dozen of the substances on the EPA carcinogen list (EPA, 1985a) failed to give positive results with any test strain in a recent NTP study (Haworth *et al.*, 1983). The negative response included particularly those substances which appear to act through a nongenetic mechanism. It now is generally recognized that such carcinogens must be classified and evaluated separately (Gori, 1980; Weisburger and Williams, 1981b).

However, even for putative genotoxic materials, there are substantial differences in response from various tests, as would be expected under the mulitfactoral and multistage concept of carcinogenesis. This raises the question of which, if any, of the various positive and negative short-term responses should be used to estimate potency. Some responses are decreased by the presence of mammalian metabolic enzymes, while others are enhanced, raising another confounding issue. Nevertheless, scientists, have endeavored to test the correlation of various tests and animal bioassays.

Rosenkranz (1977) found a different order of activity for halogenated olefins and their derivatives between *Escherichia coli* and *Salmonella*, and concluded that "bacterial mutagens cannot be scored reliably in the standard *Salmonella* assay."

Casto (1981) reported on the lowest effective concentration for viral transformation enhancement of 136 chemicals from a broad range of classes. He found a 94% agreement with the current carcinogenic classification. Fifty chemicals were tested by three other tests also, and the DNA repair test was found to be 50% accurate, while the DNA fragmentation test was 72% correct, and the chemical transformation test rated 92%. Potencies, defined as the lowest effective concentration, ranged over five orders of magnitude.

Margolin *et al.* (1981) attempted to adjust for the acute cytotoxic effects by developing a dose–response curve adjusted for this toxicity. Their key parameter β in a single-hit model was taken to represent mutagenic potency. They explained the large uncertainty in the value of β by the presence of an excess of histidine in the test solution, and thus

a failure to restrict the test to the first generation only of the bacteria. They found that the test as normally run was not accurate to more than one order of magnitude.

Parodi and co-workers (1982) found some correspondence between carcinogenicity and one or another of six short-term parameters in 95 cases. The correlation index was about 0.4, giving poor quantitative predictability, and the range of calculated potencies covered about seven orders of magnitude.

Anderson (1982) suggested a ranking of test methods by their predictive value, and defined PV_{10} as the proportion of carcinogens among the chemicals found to be positive when 10% of the group submitted for testing are found to be carcinogens by bioassay or short-term test. The data of Purchase *et al.* (1978) were evaluated and the Ames and cell transformation tests were assigned PV_{10} values of 63 and 77, respectively, but it was pointed out that any positive result in any one of the short-term tests was accepted. Using IARC and NCI bioassay test results she assigned PV_{10} values of 35 and 38 to mouse and rat studies.

After consideration of the characteristics of these tests, she concluded that for potency "a fixed quantitative relationship is improbable for several reasons." The reason included the need to consider dose–response slopes, and the variability of test protocols and conditions, including the *S*-9 preparation, and stated that "false positive and false negative results belie a quantitative relationship between *in vitro* and *in vivo* results."

In discussing the problems and advantages of short-term tests in assessing carcinogenic risk, Ramel (1983) concluded that while qualitative correspondence could be expected between short-term tests and carcinogenicity, quantitative correspondence could not. Some of the reasons he gave were that only the initiating and not the promoting steps are related to mutagenicity, and because of the repair system, cancer is not an inevitable result of a mutation. He points out the need for a better understanding of the overall mechanism before reliable test systems could be developed. This position is supported by the observed reversibility of some of the supposedly mutagenic endpoints in humans (Hansteen *et al.*, 1978; Pitot, 1983; Shubik, 1984; Clive, 1985).

Tates (1982), speaking at the same symposium, reviewed the kinetics of several widely used *in vitro* tests in several types of cells and said that "appreciable differences were observed between the kinetics of EMS-induced mutations in microorganisms and in mammalian cells." His conclusion was that "one should be cautious in quantitatively extrapolating potencies from lower organisms to higher eukaryotes, in particular man. Fuchs (1984) found that there is no direct correlation between the binding spectrum and the mutation spectrum of *N*-acetoxy-2-acetylaminofluorene.

Singh and Gupta (1983) determined the minimum concentrations of 13 anticancer drugs which were effective in the multiple genetic loci test, using five loci, and the sister-chromatid exchange test in Chinese hampster ovary cells. Potencies ranged over five orders of magnitude, and all drugs were positive in all tests, but displayed "interesting genetic lower-specific differences in their responses." A plot was provided of the dose–response curves, which showed wide variation in the slopes of the curves.

DeFlora and co-workers (1984) compared 135 compounds in the Ames test and the DNA repair test. They found a 71% agreement between the two tests, and a 64.5% accuracy for the Ames test and a 72.4% accuracy for the repair test, as evaluated by their carcinogenicity in rodents. The potencies, defined as the number of responses divided by the quantity of the test material, varied over seven to nine orders of magnitude.

Osterman-Golkor (1984) determined the dose–response curves for more than a dozen substances over a wide dose range while measuring the mutation rates in *E. coli*. Many separate doses were spread over two or more order of magnitudes using 10^9 cells per treatment, so that very reliable dose–response curves were generated. Each of these showed a linear relationship on a log–log plot in the lower part of the range studied, and each substance displayed a dose–response curve which yielded less than 0.1 mutant per 10^8 survivors in or near the experimental dose range. Several of the substances gave quadratic dose–response curves in the upper dose ranges.

Garrett *et al.* (1984) grouped many of the short-term tests into categories evaluating similar results and tested them against 24 substances listed as human carcinogens by IARC. This led to identification of the most appropriate tests for each target organ. As expected, a considerable degree of discordance was found for certain test pairs, and it was stated that negative results in certain tests often is the correct response for a substance which may be positive in a variety of short-term test systems. They pointed out the need for further validation of many of the tests with known carcinogens and noncarcinogens.

Russell and co-workers (1984) compared the findings from mammalian germ cell mutation tests, which are presumed to predict heritable risks, with results from other genotoxicity assays. In only 1 of 25 such comparisons, that of heritable translocation versus unscheduled DNA synthesis in the testis, did a significant correlation appear. Ma *et al.* (1984) reported a 67% congruity between the Ames test in 41 substances in the *Tradescantia* micronucleous test.

Dolara and Caderni (1984) compared the test results for a group of carcinogens and noncarcinogens and concluded that, despite the apparent degree of agreement, "a causal link between the two phenomena of mutagenicity and carcinogenicity" has not yet been established. They find that, depending on the criteria used to determine a positive response, correct results are found for 58–81% of carcinogens and 34–66% of noncarcinogens. Their conclusion is that "given the difficulty of extrapolating from animal carcinogenicity studies to human carcinogenicity, it would be unwise to expect a nonproblematic link between mutagenicity and carcinogenicity."

Shelby and Stasiewicz (1984) analyzed the results of several *in vitro* and *in vivo* short-term tests on materials found to be negative in two species in bioassays. They found high rates of false positive results in all test methods, and very poor concordance of any of the test methods with the bioassay or each other.

Cragg *et al.* (1985) attempted to correlate results from the *Salmonella* assay with skin-painting data for several hydrocarbon fractions, but were unable to obtain positive *Salmonella* results for even the fractions positive in the skin test. Blackburn *et al.* (1984) had seen quite good correlation in a similar trial and commented that modifications to the Ames test are needed before reliable results can be obtained for petroleum-based materials. With such modifications, the Ames test appeared as a reasonable predictor for skin carcinogenicity for that group of substances.

The difficulties experienced by these two groups are a reflection of the greater overall problem that none of the short-term *in vitro* tests can depict the whole animal. Butterworth (1985) has been quoted as stating that "I can think of nothing further from the human condition than the Ames test." He goes on to recommend *in vivo/in vitro* DNA repair tests that do incorporate whole animal response.

As more is learned about the mechanisms of carcinogenics it may be possible to find suitable tests which do reflect similar mechanistic processes. For example, Kilkenny *et al.* (1985) report "an excellent positive correlation" betweeen the potency of a

variety of animal skin initiators and their capacity as inducers of altered differentiation of cultured mouse keratinocytes *in vitro*, and Tofilon *et al.* (1985) reported that the *in vivo* tumor response to chemotherapeutic agents was predicted by their capacity to induce sister-chromatid exchange *in vitro*. Related effects such as these are more likely to be predictable than are nonrelated events represented by the apparently random choices of test pairs utilized in some potency comparisons.

An often unrecognized problem pertinent to the use of short-term test data to estimate carcinogenic potency is the no-response level. It generally is accepted that mutagenicity data are relevant only to genotoxic carcinogens, which are presumed by many to have no demonstrable threshold. That is correct insofar as statistics go; it is impossible to prove a negative by statistics. One must look to biological data to evaluate the question.

Thus, when mutagenic test results are taken as the paradigm for carcinogenicity, it must be recognized that, in every case where tests have been run over an adequate range of doses, a no-response dose has been observed. This has been stated explicitly by some authors (Casto, 1981; Singh and Gupta, 1983; Sivak, 1979; Sivak and Tu, 1980; Garrett, 1984; Clive, 1985) and is apparent in the data presented by others (Haworth *et al.*, 1983; Hussain, 1984).

These tests utilize several millions of cells per plate, are done in replicate, and, in the case of the Haworth (1983) study, were reproduced in several laboratories. The statistical analysis therefore is very robust. It should be noted that the plotting used in the Haworth report is incorrect. The line was drawn from the lowest dose to the lower left corner of the diagram, regardless of the break in the x axis, and of the fact that these were log plots, in which there is no "zero corner." The line should extend past the last data point only as an extrapolation of the data from higher dose levels, not as an attempted interpolation to a nonexistent zero.

The paradigm of mutagenicity for genotoxic carcinogenicity becomes a paradox when faced with an obvious no-effect dose level. Potency no longer is an invariant and independent variable, but exists only under certain conditions of exposure as it seems also to do for animals and humans. Ekwall (1983) points out that any attempt to correlate *in vitro* cytotoxicity and LD_{50} data must consider organ-specific toxicity, the pharmacokinetics of the intact animal, and the biases and limitations of the test methods that do not reflect real life.

There is a strong temptation to use the readily available short-term data to estimate potencies. In view of the fact that such data are of no relevance for nongenotoxic carcinogens and genotoxic promoters and of very dubious quantitative value for genotoxic initiators, even when they can be identified as such, that temptation should be resisted except for making the crudest of preliminary screening decisions.

Another reason for caution is the lack of past agreement on what criteria constitutes a positive or negative result. Current studies often report a stated statistic, or give enough data for the reader to draw a conclusion as to the outcome, but earlier studies often only stated the authors opinion of positive or negative. Also, many workers have used a twofold role as the criterion for a positive result which may have been unduly strict.

5. *Acute Toxicity*

There is a recognized relationship between chronic toxicity and the carcinogenic action of some nongenotoxic carcinogens. Several instances are known where cancer

does not develop until after chronic toxic effects are exhibited. Examples include formaldehyde (Swenberg *et al.*, 1980), PBB's (NTP, 1983), halogenated methanes (Reitz *et al.*, 1978), and other substances (Squire, 1984).

Wilson and co-workers (Zeise *et al.*, 1984) have extended their studies on estimation of potency with a proposal that acute toxicity, and more particularly LD_{50}, may be used to estimate carcinogenic potency. A simplified form of their definition is

$$\beta = D/(LD_{50})^C + k,$$

where β is potency, C and D are empirically derived constants, and k contains an expression of the variance of the estimate. Values of C and D were calculated for Osborne–Mendel (OM) and Fischer 344 rats and B6C3F1 mice from NCI bioassay results. Regression analyses of the logarithms of the potency and LD_{50} values were then done. Correlation coefficients were found to range from -0.93 for the OM rat to -0.59 for the mouse. More detailed examples were given of the calculations for saccharin, tetrachlorodibenzo-*p*-dioxin, formaldehyde, and vinyl chloride.

The method is recommended by the authors for making decisions on interim standards and on the need for further study. The authors state that there are, under this proposal, no noncarcinogens, only those whose potency has not yet been measured, and therefore is smaller than some upper bound.

There are two methods that can be used to test this hypothesis. One is to compare the calculated potencies of animal carcinogens with the reported LD_{50} values to see if there is some correlation. The other is to compare the LD_{50} values of several common substances with that of some known carcinogens to see if the supposed carcinogenicity of these as yet unrecognized carcinogens is adequate to allow detection by epidemiology. The authors apply the latter test and find that sugar and salt each would predict an annual cancer mortality of about one million persons. Alcohol and caffeine would add another few hundred thousand each, and aspirin about 25,000. They suggest that this proposal should be tempered with the results of short-term tests, and also state that inorganic chemicals and physical agents (asbestos) are not expected to fall under the scheme.

The first test method also fails to support the hypothesis. Acute oral data are available from NIOSH (1983) for 42 of the substances for which the EPA (1985a) has published potencies based on bioassay data. Regression analysis of the EPA potencies against the rodent LD_{50} data from NIOSH (1983) yielded a correlation coefficient of 0.089 (Turetzky, 1983, personal communication). Recalculation using only the 14 chloroaliphatic substances on the EPA carcinogen list (EPA, 1985a) gave a line with slope of 0, indicating no relationship between the variables.

There may well be some general association between acute toxicity and carcinogenicity. Shimkin *et al.* (1966) first suggested such a relationship based on the development of lung tumors in strain A mice when treated with alkylating agents. It is recognized that a variety of physical and physiological effects either enhance or produce a carcinogenic response. Substances which cause this result are grouped loosely into the nongenetic or epigenetic categories (Gori, 1980; Weisburger and Williams, 1981b). Three of the four substances cited by Zeise *et al.* (1984) as having some correlation between toxicity and carcinogenicity are thought to be nongenotoxic in their mechanisms. The fundamental questions are whether this relationship is generic or specific and whether it can be quantified, particulary between species.

Recent reports suggest that there may be some relationship between acute and

chronic toxicity in fish. For example, Kenaga (1982) reviewed the literature reports on 84 substances and found that 86% of the LC_{50} acute values were less than two orders of magnitude from the chronic no-effect dose. The acute/chronic ratio was 25 or less in 93% of the cases. Suter *et al.* (1983) found that acute toxicities can be extrapolated between either species or orders of fish with good reliability, and confirmed the findings of Kenaga (1982) of a value of 0.8 for the regression coefficient for the 96-hr LC_{50} and the maximum acceptable chronic exposures.

Janardan *et al.* (1984) found reasonable (order-of-magnitude) correlations between LC_{50} for fish and LD_{50} for rats for substances that act through nonspecific toxicity mechanisms, but not for those with specific toxicity, such as organophosphate pesticides. Thus, these may be a very serious chain from acute toxicity to chronic, and from some species to others, but the generality of the relationship is not proven.

Albert has supported the use of acute toxicity in the first stage of an all encompassing test program (Albert, 1983a). All substances for which there are significant human exposures would be subjected to genotoxicity tests in order of descending acute toxicity. Those which passed that test would then be tested for promotional ability. The genotoxic substances found in these two tests would then be subjected to a simplified "supra-MTD" bioassay. Substances which gave positive results in that step would then go to a full-scale bioassay. The purpose of this proposed program is to reduce testing costs and time by concentrating on those which are most likely to be frankly carcinogenic. This proposal received little support from scientists who were asked to review the idea and who pointed out the lack of support for a quantitative relationship between mutagenic potency and carcinogenicity (Marshall, 1982).

T. D. Jones and co-workers (1983) propose that tumorigenesis should be studied in light of cellular population kinetics; they then use cell irritation, proliferation, and differentiation data to estimate tumorigenic potency (cell potentiation) for several simple compounds. While emphasizing that cell-specific potentiation factors must be considered, they utilize acute exposures in their models. Working back from the potency calculation they compare "practical thresholds" for several simple substances such as ozone, carbon monoxide, and sulfur dioxide to the allowable exposures under current recommendations. They conclude that current exposure limits are adequate to protect from carcinogenesis for most of these substances. They also suggest that this approach should be useful for planning research, ranking priorities, and guiding experimental designs.

Parodi and co-workers (1983b) included LD_{50} as one of the variables tested in their study of the relationship between carcinogenicity and *in vitro* results. They found a 0.43 correlation coefficient between the log of the LD_{50} and the log of the potency index in the Ames test with 76 test substances.

One additional relationship between acute and chronic toxicities must be considered also. It is a fundamental tenet of classical toxicity that there in a dose below which no effect is seen (ACGIH, 1984; Weil, 1984). If carcinogenicity is related to acute toxicity, either directly or through secondary or chronic effects, it must then observe the same no-response phenomena as does toxicity. Proponents of the acute toxicity–carcinogenicity relationship are not willing to accept this point. Rather, they propose to denote all substances as human carcinogens (Albert, 1983a; Zeise, 1984) and reduce risk by regulating or testing the substances in order of descending toxicity. It is doubtful that a hypothesis that puts carbon monoxide, hydrogen cyanide, and hydrogen sulfide near the top of the priority list really is helpful to the regulators trying to allocate their resources wisely.

Finally, the shortcomings of the LD_{50} test itself are well known and suggest that extensive extrapolation of conclusions based on such data are subject to substantial error (Davan, 1983; Ekwall, 1983; Malmfors and Tieling, 1983; Paset, 1983; Rowan, 1983).

6. *Structure–Activity Relationship*

Reliance on structure–activity relationships (SAR) is an intuitive act for chemists, who use the principle of homology in their everyday thinking. It is thus natural to begin to try to expand those relationships beyond physical properties and chemical activity. It is not surprising then that seekers of Ehrlich's "magic bullet" or Huxley's "cunningly contained torpedo" (Golberg, 1983) should turn to SAR for guidance, and that others should feel comfortable in predicting toxic properties from structural features. The potential value of such a system, if it should be successful, has generated considerable interest and research and at least one major symposium (Golberg, 1983).

Many procarcinogens are thought to be activated by the cytochrome *P*-450 group of enzymes to epoxides, which can then bind to DNA and result in cancer in animals. These products are seen frequently as the metabolic intermediates and give positive *in vitro* responses. This is particularly true of the chlorinated olefin family and in many polycyclic aromatics. A number of workers have been encouraged to examine on theoretical and experimental bases the tendency of substances to form epoxides and the stability of those products. A high degree of correlations has been found between some of these parameters and the carcinogenicity of the precursors in rodents (Henschler, 1977; Van Duuren, 1977; Van Duuren *et al.*, 1983; R. B. Jones and Mackrodt, 1983a,b; Loew *et al.*, 1984; Politzer and Lawrence, 1984). However, studies at Vanderbilt have reported (Liebler and Gungerich, 1983) that the epoxides do not appear to be in the direct pathway to tumor formation, but are simply coproducts. Furthermore, the adduct of the expoxide from vinyl chloride with guanine does not cause miscoding in a replication fidelity assay (Barbin *et al.*, 1985). Therefore, the interesting SAR correlation found in these cases may be only coincidental.

Vance and Levin (1984) have analyzed a series of nitroaromatics for activity in the Ames test and concluded that ring size, ring position of the nitro group, conformation of the nitroso group to the plane of the ring, and ability to stabilize the ultimate electrophile through resonance are important in determining the activity. They suggest that this work now allows the estimation of the mutagenic potency of nitroaromatics. Hall *et al.*, (1984) have measured the addicitivity contribution of several functional groups to acute toxicity in fish of a series of substituted benzenes and found a correlation coefficient of 0.95. If acute and chronic toxicities are related, this relationship could be useful.

Rosenkranz and co-workers (1984) have concluded that no satisfactory SAR has been devised from a pattern recognition basis and stated that there is no evidence that biological activity is determined by the functional group or the organic moieties present. An insufficient data base is given as one reason for the lack of a suitable prediction system. They propose that a molecular fragment evaluation, rather than pattern recognition, is more successful in identifying both activating and inactivating portions of the molecule, and thus the prediction of the overall risk posed by that molecule. The result of their analysis is an estimate of the probability of the carcinogenicity of the substance, as determined by argument with animal bioassay results. A table of

results shows good agreement for substances thought to act by genotoxic mechanisms, but very poor accuracy for nongenotoxicants.

Nohara *et al.* (1985) worked out the structural relationship necessary in a series of nitro- and aminobiphenyls in order to achieve mutagenicity in the Ames test. They concluded that "the positional effects of substituents on mutagenicity remain to be explained theoretically."

Mellinger (1985) reported that polycyclic amino structures always were predictive for mutagenicity in complex hydrocarbons mixtures and that molecular weight was important in initiation and promotion. Their data base included *Salmonella* tests, mammalian cell transformation, and mouse skin painting.

The ultimate shortcut is, of course, to use a computer to predict carcinogenicity, or other toxicological properties, directly from a formula on paper, with no experimentation needed. Several workers have developed procedures for just that purpose (Enslein and Craig, 1982; Enslein *et al.*, 1983; McCann *et al.*, 1983; Rosenkranz *et al.*, 1984; Loew *et al.*, 1984; Helmes *et al.*, 1985).

At least one on-line electronic data base offers while-you-wait computation of LD_{50} values (Rekker, 1982). In other cases (Kaufman, 1983) a program will design a structural module with the desired toxicological features. These procedures rely on multiple regression analyses of specific structural groupings, atoms, or bonds, together with physical properties, against recorded toxicity data for a specfied group of exemplary substances (Chignell, 1983).

There appears to be a general consensus that, so long as relationships are confined to a single family of compounds, or to a single species, that SAR may be helpful and reasonably accurate for qualitative purposes. Attempts to quantify the relationship (QSAR) or to extrapolate to other species or significantly difficult functional groups are less successful. One major problem is an acknowledged lack of an adequate data base. Golberg (1983) pointed out that "a concerted effort is needed on the part of all of us to produce the necessary information. . . . Then, and only then, will the application of QSAR assume its rightful place." A review committee assembled by the EPA (EPA, 1985b) has concluded that QSAR cannot be "recommended as the sole basis of a risk assessment of chemical mixtures."

More than just large numerical data bases are needed, though, because biological variations must be accounted for. The QSAR working party of the IUPAC Commission on Medicinal Chemistry (1981) defines the requirements as a set of biological test data as obtained from well-defined interactions between chemical substances belonging to congeneric series of structures and an active site in a biological system (Rekker, 1982). That need does not seem to have been met. Trosko (1985) states that the major problem has been the use of "bad" *in vitro* data in attempts to correlate with *in vivo* results, and calls for a critical reevaluation of the entire *in vivo/in vitro* program.

Careful control of all three of the variables listed by the IUPAC commission will be difficult. Metabolism occurs in a variety of organs at rates dependent on dosage in many cases, and minor changes in substance structure can make major changes in one or more of the metabolic stages leading to the ultimate carcinogen, even within the same species. Wishnok (1981) has concluded that application of QSAR "appears to require extensive additional physiological and biochemical information about the species . . . the extrapolation of relative potency data from one species to another should only be attempted with the understanding that such evaluation may be strongly dependent on the original experimental dosing scheme." Dudney *et al.* (1982) also emphasize the variability of results between different systems.

The Office of Toxic Substances relies heavily on SAR in screening the premanufacturing notices reviewed under Sec. 5 of the Toxic Substances Act and is engaged in a 2-year study to verify the validity of many of the assumptions supporting that reliance (EPA, 1984d). NTP was invited to participate in the study, but the Board of Scientific Counselors has declined to do so, citing flaws in study design and the inavailability of much of the needed data. They have offered to assist in redesigning the proposal (NTP, 1985). The FDA has had in use for several years a "catalogue of structural moieties" which it uses in setting priorities for testing in the Food Animal Evaluation Branch (Lorentzen, 1978). The guideline points out the need for metabolic data by the target tissue in order to make accurate predictions.

Acceptance of QSAR has not been as great in Europe even as in this country. ECETOC (1982) concluded that SAR has very limited reliability, and IARC (Bartsch and Tomatis, 1983) concludes that the relationship is "not presently established." However, German authorities do use potency values based at least partly on structure in setting occupational exposure limits for presumed carcinogens (Ministerium fur Arbeit, 1982).

DISCUSSION

There are relatively few substances for which the relationship between toxicity and carcinogenicity is such that most or all of the rodents will have developed malignant tumors at the end of a lifetime bioassay, even when treated at the maximum tolerated dose. In fact, only a minority of bioassays produce positive results in more than one of the four species/sex combinations found in each study (Purchase, 1980; Haseman *et al.*, 1984; DiCarlo and Fung, 1984; Shelby and Stasiewicz, 1984; Meyers *et al.*, 1985). There is little controversy regarding the need to control exposure to those substances which do show positive epidemiological response or which produce high yields in an appropriate animal study. Disagreement over the proper regulatory stance develops when the expression of carcinogenicity is not clear, and especially when it is necessary to resort to detailed statistical analyses to determine if there actually has been a response. As shown in the references cited immediately above, a majority portion of the substances tested by the NCI and NTP or considered by IARC to be human carcinogens fail to give a clear signal in a bioassay. Therefore, if the qualitative determination of the carcinogenicity of a substance is equivocal, efforts to place a quantitative value on its carcinogenicity must be even more difficult.

This difficulty is magnified greatly each time that the data source is found to be another step away from the human condition. By the time the path has led from whole animals to bacteria or individual cells, then to nonchronic effects, and finally to pure structural considerations, as described above, much of the scientific foundation has been moved out of the process and has been replaced with the inflexible assumptions and fixed guidelines which seem to be necessary to make the process coherent, or at least workable.

Much of this apparent difficulty comes from the attempt to use a single datum as a basis for a regulatory decision. A positive bioassay is a significant piece of information, not to be ignored in safety and health decision making. A positive short-term test also is a useful datum and can serve as an indicator of the need for more study. None of the various pieces of information taken singly, or often not even when taken together, is an adequate basis for a sound decision on how best regulate a particular substance in order to protect human health and the environment adequately.

The Food Safety Council (1982) proposed a decision-tree approach for evaluating the safety of food components which utilizes all available data on a substance and the compound's metabolic fate before reaching a final decision. Park and Snee (1984) expanded on this procedure and discussed its utility in other applications.

Squire (1981) proposed a ranking system for setting priorities for further testing depending on a quantitative assessment of the value of each component of the decision tree. Weisburger and Williams (1981a) have suggested a more limited range of tests to assist in decision making.

Only slight signs of acceptance within regulatory circles for this broader approach are evident. California is using a modified Squire ranking system for safety evaluation of pesticides (Wang, 1984) and is in the processs (CDHS, 1985) of developing a policy which will rely in part on potency as a guide to regulatory action. EPA has claimed a weight-of-the-evidence process for evaluating carcinogens for many years (Anderson, 1983), but a recent proposed revision of the 1974 interim guidelines (EPA, 1984a) actually left little choice other than to perform a linear, no-threshold extrapolation of the upper 95% limit risk if any sex of any species showed a positive response at any organ.

Clayson (1983) has been nearly alone in attempting to utilize knowledge of the mechanism of action in the determination of potency, and, therefore, risk. Other methods treat cancer as a single disease which arises by a single mechanism. This is a conscious decision (EPA, 1984a). This biologically unsupportable position is invoked whenever there is less than complete knowledge, and thus is likely to prevail in most regulatory decisions. A better understanding of the mechanisms involved is likely to be of much greater value to society than is the performance of a few dozen more lifetime studies on suspect substances that obviously cannot be having a measurable effect on human life spans.

Research workers generally have had less interest in the need for greater precision and validation of the test systems discussed above than in developing a greater diversity of tests. Neither science nor the public is being served by this expansion in the number of data bases of uncertain value. The existence of a hundred or so short-term tests claiming significance for "mutagenicity," and, therefore, perhaps carcinogenicity, is a prime example. It may well be more profitable to determine the meaning of mutagenicity, and its potential relationship to human carcinogenicity, than to develop more test systems at this time (Clive, 1985).

A major concern is the overcompartmentalization of the risk assessment arena and the overdependence on specialists who are not familiar with other segments of the field, and who thus have inadequate appreciation of the overall significance of the whole body of data. This concern was discussed extensively at the Deer Creek Workshop (Hughes *et al.*, 1983) but continues to be a factor, particularly in regulatory considerations. In particular, statistics often is not the servant of the scientist, but too frequently is used to assert conclusions which do not comport to real life experience. We must find a means to use our scientific tools more wisely than we do at times.

Overinterpretation of the potency estimate as now prepared can be a serious problem. The EPA (1985a) has recognized that their animal-derived upper limit estimates warrant no more than single-figure significance, yet results using three or four significant figures are published regularly. The public has been led to believe that these estimates are correct, and thus they demand greater protection from the overstated risk than is possible. Failure to achieve this expectation leads to dissastifaction and unrealistic legislative or judicial attempts to achieve unrealistic goals (Ruckelshaus, 1985).

Public understanding could be improved greatly by the use of more realistic esti-

mates. The most probable value, rather than upper limit estimates, should be calculated and presented with appropriate confidence ranges. It is not useful to the general public to present a risk estimate with the accompanying statement that "we don't really mean it," as is done with all the explanatory caveats placed alongside the upper 95% estimates. FDA (1981) has described these estimates as "a conclusion with reasonable certainty of what will not occur." Understandably, the public has not expressed great confidence in agency-derived risk assessments.

The National Research Council (1983), through its Committee on the Institutional Means for Assessment of Risks to the Public, has examined the process of risk assessment in the federal agencies and developed a lengthy list of recommendations. A major component of the assessment process is recognized as the inference options which are involved in decisions on how to apply the data, or indeed which data to apply. The principle finding of the committee was that risk managment and risk assessment must be kept separate to minimize improper comingling of science and political requirements and that the inference options be discussed thoroughly when used. They then recommended a board of risk assessment to review the data base and the results of the analysis of that data. They recommended an independent board of scientific stature to serve as a continuing locus of discussion about ways to improve scientific and procedural aspects of risk assessment. Only EPA has attempted to implement this last recommendation. It has increased the utilization of the Science Advisory Board as a review body for its health assessment documents and its risk assessment guidelines.

The estimate of potency which eventually is established sets the allowable exposure for a particular risk. The magnitude of the potency estimate can have significant influence on whether (further) regulation is required, particularly for a ubiquitous ambient exposure. Reduction of exposures beyond current concentrations is becoming more and more difficult as those concentrations are decreased, and thus accuracy and credibility of potency estimates become more important as time progresses and we continue to expand the sensitivity of our test systems to find responses at lower and lower strengths.

A risk assessment is the culmination of all knowledge concerning the effect of a substance on human health. It, therefore, should utilize all available relevant information. The current procedures for estimating carcinogenic potency from select, limited, and often irrelevant data do not do so. Whether they are simple "pencil and ruler" exercises favored by FDA (Lorentzen, 1984) or the more complicated computer-based programs used by EPA, all appear to place too complete a dependence on the single experiment and use too few of the other biological data that are available.

CONCLUSIONS

A completely acceptable method of estimating relative or absolute potency values relevant to humans has not yet become available. The nearest approximation is the upper limit on risk which can be estimated from epidemiological data. If no cases actually are found, the real risk will be below this upper limit and could be 0. However, statistics do not permit a demonstration of 0 risk.

This, however, has not prevented, and should not prevent, estimates being made from other data, so long as appropriate caveats are introduced, and the results are used properly. The further the source of the data departs from the human condition, the stronger the caveat that is needed and the broader the confidence limits that must be placed on the estimate.

No generic procedure can be prescribed for preparing potency estimates. A case-by-case evaluation utilizing all of the available data is required. There should be a clear definition of all assumptions and policy decisions which were introduced, along with a discussion of the consequences of these decisions. The result should be presented as a most likely value, with appropriate confidence or sensitivity analysis limits.

Risk asssessment is an essential element in risk management. The risk manager must be presented with the best scientific support that is available in order for proper decisions to be made.

Public acceptance of the estimates, and the risk management actions taken, can be expected only if the results are shown to be within a reasonable range compatible with real life experience. Acceptance will be improved greatly if there has been independent peer review of the data base and the final results. There is a serious need for a concerted scientific effort to improve the accuracy of, and to demonstrate the validity of, the tests now being used to supply this data base.

We should not abandon the tests currently available because they have not provided complete answers or because their results have been misused by some. We should continue to try to understand the significance of the data and search for its proper application. Much remains to be done to reach this goal. One important and recognized need is to obtain better validation of the many *in vivo* and *in vitro* tests by expanding their coverage to a wider range of noncarcinogens and nongenetic carcinogens. Another need discussed earlier is a method of correcting for latency, or, more accurately, of determining time to tumor, so that the acutal effective doses can be estimated. Most important, however, is the need to learn more about the shape of the human dose–response curve at very low doses and how it varies with different mechanisms of carcinogenesis, including both promoters and initiators.

REFERENCES

ALBERT, R. A. (1983a). *Testing Strategy for Complex Mixtures of Carcinogens.* Presented at Workshop on Methods for Estimating, Predicting, and Interpreting Health Effects of Simultaneous or Sequential Exposure to Several Chemicals: Univ. of Surrey, Guildford, England, August 15–19.

ALBERT, R. A. (1983b). Comparative carcinogenic potencies of particulates for diesel engine exhausts, coke oven emissions, roofing tar aerosols and cigarette smoke. *Environ. Health Perspect.* **47**, 339.

American Conference of Governmental Industrial Hygienists (ACGIH) (1984). *Threshold Limit Values for Chemical Substances in the Work Environment.* Cincinnati, Ohio.

AMES, B. N., GOLD, L. S., SAWYER, C. B., AND HAVENDER, W. (1982). Carcinogenic potency. *Environ. Mutagen. Carcinogen. Proc. Int. Conf. 3rd.* p. 663.

AMES, B. N., AND HOOPER, K. (1978). Does carcinogenic potency correlate with mutagenic potency in the Ames assay?. *Nature (London)* 274, 19.

AMES, B. N., AND MCCANN, J. (1981). Validation of the *Salmonella* test: A reply to Rinkus and Legator. *Cancer Res.* **41**, 4192.

ANDERSON, D. (1982). The predictability of bioassays. In *Mutagens in Our Environment* (M. Sorsa and H. Vainio, eds.). Liss, New York.

ANDERSON, E. (1983). *Quantitative approaches in use to assess cancer risk. Risk Anal.* **3**, 277.

ANDERSON, M. W., BOROUJERDI, M., IOANNOU, Y. M., KUNG, H.-C., AND WILSON, A. G. E. (1983). Modifying factors affecting toxic responses: Example—Inhibitors of carcinogenesis. In *Structure–Activity Correlation as a Predictive Tool in Toxicology* (L. Golberg, ed.). Hemisphere, Washington, D. C.

ASHBY, J., AND STYLES, J. A. (1978). Does carcinogenic potency correlate with mutagenic potency in the Ames assay?. *Nature (London)* **271**, 452.

BARBIN, A., LAID, R. J., AND BARTSCH, H. (1985). Lack of miscoding properties of 7-(2-oxoethyl)guanine, the major vinyl chloride–DNA adduct. *Cancer Res.* **45**, 2440.

BARR, J. T. (1982). *Risk Assessment for Vinyl Chloride in Perspective.* Paper presented at the 75th Annual Meeting of the Air Pollution Control Association, New Orleans, La., June 20–25, paper 82-9.-2.

BARR, J. T., HUGHES, D. H., AND BARNARD, R. C. (1981). The use of risk assessment in regulatory decision making: Time for a review. *Regul. Toxicol. Pharmacol.* **1**, 264.

BARTSCH, H., AND TOMATIS, L. (1983). Comparison between carcinogenicity and mutagenicity based on chemicals evaluated in the IARC monographs. *Environ. Health Perspect.* **47**, 305.

BERNSTEIN, L., GOLD, L. S., AMES, B. N., PIKE, M. C., AND HOEL, D. G. (1985). Some tautologous aspects of the comparison of carcinogenic potency in rats and mice. *Fundam. Appl. Toxicol.* **5**, 79.

BLACKBURN, G. R., DEITCH, R. A., SCHREINER, C. A., MEHLMAN, M. A., AND MACKERER, C. R. (1984). Estimation of the dermal carcinogenic activity of petroleum fractions using a modified Ames assay. *Cell Biol. Toxicol.* **1**, 67.

BRIDGES, B. A., ZEIGER, E., AND MCGREGOR, D. B. (1981). Summary report on the performance of bacterial mutation assays. In *Evaluation of Short-Term Tests for Carcinogens: Report of the International Collaborative Program* (F. J. de Serres and J. Ashby, eds.), pp. 49–67. Elsevier, New York.

BRUSICK, D. (1983). Mutagenicity and carcinogenicity correlations between bacteria and rodents. *Ann. N. Y. Acad. Sci.* **407**, 164–176.

BUTTERWORTH, B. E. (1984). as quoted in *Chemical Week*, June 27, p. 84.

CALABRESE, E. J. (1983). *Principles of Animal Extrapolation* Wiley, New York.

California Department of Health Services (CDHS) (1984). *A Policy for Chemical Carcinogens: Guidelines for Risk Assessment and Their Scientific Rationale.* Draft, September 25, Sacramento.

California Department of Health Services (1985). *A Policy for Chemical Carcinogens: Guidelines for Risk Assessment and Other Scientific Rationale.* Sacramento.

CARNES, B. A., DORNFERD, S. S., AND PEAK, M. J. (1985). A quantitative comparison of a percentile rule with a 2-fold rule for assessing mutagenicity in the Ames assay. *Mutat. Res.* **147**, 15.

CASTO, B. C. (1981). *Potency Ranking of Chemicals Based on Enhancement of Viral Transformation*, EPA-600/1-81-045, July.

CHIGNELL, C. F. (1983). Overview of molecular parameters that relate to biological activity in toxicology. In *Structure–Activity Correlation as a Predictive Tool in Toxicology* (L. Golberg, ed.). Hemisphere, Washington, D. C.

CLAYSON, D. B. (1983). Trans-species and trans-tissue extrapolation of carcinogenicity assays. In *Organ and Species Specificty in Chemical Carcinogenesis* (R. Langenbach, S. Nesnow, and J. Rice, eds.). EPA-600/9-83-008, June, PB83-220137.

CLAYSON, D. B., KREWSKI, D., AND MUNRO, I. C. (1983). The power and interpretation of the carcinogenicity bioassay. *Regul. Toxicol. Pharmacol.* **3**, 329.

CLIVE, D. (1985). Mutagenicity in drug development: Interpretation and significance of test results. *Regul. Toxicol. Pharmacol.* **5**, 79.

Combustion Engineering, Inc. (1985). *Technical Background Document to Support Rulemaking Pursuant to CERCLA*, Section 102, Vol. 3. Under Contract No. 68-03-3182 to the Environmental Protection Agency, Newbury Park, Calif., March. See also ANDERSON, E. L., CHU, M., DOURSON, M., AND DE ROSA, C. (1985). *Methodology for Ranking the Degree of Hazard Associated With Exposure to Carcinogens and Other Toxic Chemicals*, February, PB85-167906.

Consumer Product Safety Commission (1983). *Carcinogen Potency Ranking Discussion Document*, draft, April.

CORN, J. K., AND CORN, M. (1984). The history and accomplishments of the Occupational Safety and Health Adminstration in reducing cancer risks. In *Reducing the Carcinogenic Risks of Industry* (P. F. Deisler, Jr., ed.). Dekker, New York.

CRABTREE, H. C. (1983). Relative risk assessment: An aid to planning. *Ann. Occup. Hyg.* **27**, 213.

CRAGG, S. T., CONAWAY, C. C., AND MACGREGOR, J. A. (1985). Lack of concordance of the *Salmonella* microsome assay with the mouse dermal carcinogenesis bioassay for complex petroleum hydrocarbon mixtures. *Fundam. Appl. Toxicol.* **5**, 382.

CROUCH, E., AND WILSON, R. (1979). Interspecies comparison of carcinogenic potency. *J. Toxicol. Environ. Health* **5**, 1095.

CROUCH, E. A. C., WILSON, R., AND ZEISE, L. (1983). The risks of drinking water. *Water Resour. Res.* **19**, 1359.

CROUCH, E., WILSON, R., AND ZEISE, L. (1984). Comments on *EPA Draft Revised Interim Guidelines for the Health Assessment of Suspect Carcinogens*, see 49FR46294, 15 October.

CRUMP, K. S. (1984). A new method for determining allowable daily intakes. *Fundam. Appl. Toxicol.* **4**, 854.

DAVAN, P. (1984). Complete program for acute toxicity testing—Not only LD_{50} determination. *Acta Pharmacol. Toxicol.* **52**(Suppl. 2), 31.

DEFLORA, S., ZANACCHI, P., CAMOIRANO, A., BENNICELLI, C., AND BADOLATI, G. S. (1984). Genotoxic

activity and potency of 135 compounds in the Ames reversion test and in a bacterial DNA-repair test. *Mutat. Res.* 133, 16.

DiCarlo, F. J., and Fung, V. A. (1984). Summary of carcinogenicity data generated by the National Cancer Institute/National Toxicology Program. *Drug Metabol. Rev.* 15, 1251.

Doll, R. (1984). Occupational cancer: Problems in interpreting human evidence. *Amer. Occup. Hyg.* 28, 291.

Dolara, P., and Caderni, G. (1984). The application of mutagenicity tests to the prediction of carcinogenic activity of chemicals and drugs. *Pharmacol. Res. Commun.* 16, 421.

Dudney, C. S., Walsh, P. J., Jones, T. D., Calle, E. E., and Griffin, G. D. (1982). *On the Use of Relative Toxicity for Risk Estimation.* Conf. 820122-3.DE 82 010200, p. 14.

Dunkel, V. C. (1979). Collaborative studies on the *Salmonella*/microsome mutagenicity assay. *J. Assoc. Off. Anal. Chem.* 62, 874.

Ekwall, B. (1983). Correlations between cytotoxicity *in vitro* and LD_{50}. *Acta Pharmacol. Toxicol.* 52(Suppl. 2), 80.

Enslein, K., and Craig, P. N. (1982). Carcinogenesis: A predictive structure–Activity model. *J. Toxicol. Environ. Health* 10, 521.

Enslein, K., Lander, T. R., Tomb, M. G., and Lander, W. G. (1983). Mutagenicity (Ames): A structure–activity model. *Teratogen. Carcinogen. Mutagen.* 3, 503.

Environmental Protection Agency (EPA) (1979). *CAG Risk Assessment for Exposure to Ethylene Dibromide.* 21 April.

Environmental Protection Agency, Assessment Division, Office of Toxic Substances (1982). *Risk Analysis in Support of the Proposed Exemption of Site-Limited Intermediate and Low-Volume Chemicals.* May 21, see also 47FR33896, Aug. 4.

Environmental Protection Agency (1984a). Proposed guidlines for carcinogenic risk assessment and request for comments, *Fed. Regist.* 49, 46294, November 23.

Environmental Protection Agency (1984b). Regulation of radionucleides. *Fed. Regist.* 49, 43906, October 31.

Environmental Protection Agency (1984c). Regulation of benzene, 40CFR61. *Fed. Regist.* 49, 23478 and 23558.

Environmental Protection Agency, Office of Pesticide and Toxic Strategy (1984d). *Proposed Design for a Retrospective Study of PMN Hazard Prediction,* September 24.

Environmental Protection Agency (1984e). *Formaldehyde: Determination of Significant Risk: Advance Notice of Proposed Rulemaking and Notice,* 49FR21870, May 23.

Environmental Protection Agency (1985a). *Health Assessment Document for Dichloromethane.* EPA 600/8-82-004F, February.

Environmental Protection Agency (1985b). *Proposed Guidelines for the Health Risk Assessment of Chemical Mixtures, and Request for Comments.* 50FR1170, January 9.

Environmental Protection Agency (1985c). *Notification Requirements: Reportable Quantity Adjustments,* 50FR13457, April 4.

European Chemical Industry Ecology and Toxicology Centre (ECETOC) (1982). *Risk Assessment of Occupational Chemical Carcinogens.* Monograph 3, January. Brussels.

Food and Drug Adminstration (FDA) (1981). Removal of stay of regulation for the listing of lead acetate as a color additive in cosmetics that color the hair on the scalp, 46FR15500, March 6.

Food Safety Council (1982). *A Proposed Food Safety Evaluation Process.* The Nutrition Foundation, Washington, D. C.

Fuchs, R. P. P. (1984). DNA binding spectrum of the carcinogen *N*-acetoxy-*N*-2-acetylaminofluorene significantly differs from the mutation spectrum. *J. Mol. Biol.* 177, 173.

Furstenberger, G., Sorg, B., and Marks, F. (1983). Tumor promotion by phorbol esters in skin: Evidence for a memory effect. *Science (Washington, D. C.)* 220, 89.

Garrett, N. E., Stack, H. F., Gross, M. R., and Waters, M. D. (1984). An analysis of the spectra of genetic activity produced by known or suspected human carcinogens. *Mutat. Res.* 134, 89.

Gaylor, D. W., Chen, J. J., and Kodell, R. L. (1985). Experimental design of bioassays for screening and low dose extrapolation. *Risk Anal.* 5, 9.

Gehring, P. J., Watanabe, P. G., and Park, C. N. (1978). Resolution of dose–response toxicity data for chemicals requiring metabolic activation: Example—Vinyl chloride. *Toxicol. Appl. Pharmacol.* 44, 581.

Gehring, P. J., Watanabe, P. G., and Young, J. D. (1977). The relevance of dose-dependent pharmacokinetics in the assessment of carcinogenic hazard of chemicals. In *Origins of Human Cancer* Book A, *Cold Spring Harbor Conferences on Cell Proliferations* (H. H. Hiatt, J. D. Watson, and J. A. Weinsten, eds.). Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

GOLBERG, L. (1983). Structure–activity correlation as a predictive tool in toxicology (L. Golberg, ed.). Hemisphere, Washington, D. C.

GOLBERG, L. (1985). as quoted by *Food Chemical News*, p. 7, March 18.

GOLD, L. S., SAWYER, C. B., MAGAW, R., BACKMAN, G. M., DEVECIANA, M., LEVINSON, R., HOOPER, N. K., HAVENDER, W. R., BERNSTEIN, L., PETO, R., PIKE, M. C., AND AMES, B. N. (1984). A carcinogenic potency database of the standardized results of animal bioassays. *Environ. Health Perspect.* **58**, 9.

GORI, G. B. (1980). The Regulation of Carcinogenic Hazards. *Science (Washington, D. C.)* **208**, 256.

HALL, L. N., KIER, L. B., AND PHIPPS, G. (1984). Structure-activity relationship studies on the toxicities of benzene derivatives. 1. An additivity model. *Environ. Toxicol. Chem.* **3**, 355.

HANSTEEN, I. L., HILLESTAD, L., THIIS-EVENSEN, E., AND HELDAAS, S. S. (1978). Effects of vinyl chloride in man: A cytogenetic follow-up study. *Mutat. Res.* **78**, 271.

HARPER, B. L., AND MORRIS, D. L. (1984). Implications of multiple mechanisms of carcinogenesis for short-term testing. *Teratogen. Carcinogen. Mutagen.* **4**, 483.

HASEMAN, J. K., CRAWFORD, D. D., HUFF, J. E., BOORMAN, G. A., AND MCCONNELL, E. E. (1984). Results for 86 two-year carcinogenicity studies conducted by the National Toxicology Program. *J. Toxicol. Environ. Health* **14**, 621.

HAWORTH, S., LAWLER, T., MORTELMANS, K., SPEIK, W., AND ZIEGER, E. (1983). *Salmonella* mutagenicity tests for 250 chemicals. *Environ. Mutagen. Suppl.* **1**, 3.

HEDDLE, J. A., AND BRUCE, W. R. (1977). Comparison of tests for mutagenicity and carcinogenicity using assays for sperm abnormalities, formation of micronuclei and mutations in *Salmonella*. In *Origins of Human Cancer*, Book C, *Cold Spring Harbor Conferences on Cell Proliferations* (H. H. Hiatt, J. D. Watson, and J. A. Weinsten, eds.), p. 1549. Cold Spring Harbor Laboratories, Cold Spring Harbor, N. Y.

HELMES, C. T., SIGMAN, C. C., AND PAPA, P. A. (1985). Predicting carcinogenicity from structure. *CHEMTECH* **15**, 48.

HENNINGS, H., SHORES, R., WENK, M. L., SPANGLER, E. F., TARONE, R., AND YUSPA, S. H. (1983). Malignant conversion of mouse skin tumors is increased by tumour initiators and unaffected by tumour promoters. *Nature (London)* **304**, 67.

HENNINGS, H., AND YUPSA, A. H. (1985). Two-stage tumor promotion in mouse skin: An alternative interpretation. *J. Natl. Cancer Inst.* **74**, 735.

HENSCHLER, D. (1977). Metabolism and mutagenicity of halogenated olefins—A comparison of structure and activity. *Environ. Health Perspect.* **21**, 61.

HOEL, D. G., AND JENNRICH, R. I. (1979). Optimal design for dose response experiments in cancer research. *Biometrika* **66**, 307.

HOEL, D. G., KAPLAN, N. L., AND ANDERSON, M. W. (1983). Implication of nonlinear kinetics on risk estimation in carcinogenesis. *Science (Washington, D. C.)* **219**, 1032.

HOLLAND, J. M., GOSSLE, D. G., AND WILLIAMS, N. J. (1979). Epidermal carcinogenicity of bis(2,3-epoxycyclopentyl)ether, 2,2-bis(*p*-glycidyloxyphenyl)propane, and *m*-phenylenediamine in male and female C3H and C57BL/6 mice. *Cancer Res.* **39**, 1718.

HOLLAND, J. M., WOLF, D. A., AND CLARK, B. R. (1981). Relative potency estimation for synthetic petroleum skin carcinogens. *Environ. Health Perspect.* **38**, 149.

HOLLSTEIN, M., MCCANN, J., ANGELOSANTO, F. A., AND NICHOLS, W. W. (1979). Short-term tests for carcinogens and mutagens. *Mutat. Res.* **65**, 133.

HUGHES, D. H., BRUCE, R. D., HART, R. W., FISHBEIN, L., GAYLOR, D. W., SMITH, J. M., AND CARLTON, W. W. (1983). A report on the Workshop on Biological and Statistical Implications of the ED01 Study and Related Data Bases. *Fundam. Appl. Toxicol.* **3**, 129.

HUSSAIN, S. (1984). Dose–response relationship for mutations induced in *E. coli* by some model compounds. *Hereditas* **101**, 57.

International Agency for Research on Cancer (IARC) (1982). Some aspects of quantitative cancer risk estimation. *IARC Monogr.* **29** (Annex).

International Agency for Research on Cancer (1984). Preamble to Monograph No. 34, "Polynuclear Aromatic Compounds, Part 3". Lyon, France, June.

International Commission for Protection against Environmental Mutagens and Carcinogens (1984). Report of ICPEMC Task Group 5 on the differentiation between genotoxic and nongenotoxic carcinogens. *Mutat. Res.* **133**, 1.

IUPAC Commission on Medicinal Chemistry, (1981). QSAR Working Party. *Chem. Int.* **1981**, 4.

JANARDAN, S. K., OLSEN, C. S., AND SCHAEFFER, D. J. (1984). Quantitative comparisons of acute toxicity of organic chemicals to rat and fish. *Ecotoxicol. Environ. Saf.* **8**, 531.

JONES, R. B., AND MACKRODT, W. C. (1983a). Structure–genetoxicity relationship for aliphatic epoxides. *Biochem. Pharmacol.* **32**, 2359.

JONES, R. B., AND MACKRODT, W. C. (1983b). Structure–mutagenicity relationships for chlorinated ethylenes: A model based on the stability of the metabolically-derived epoxides. *Biochem. Pharmacol.* **31**, 3710.

JONES, T. D., GRIFFIN, G. D., AND WALSH, P. J. (1983). A unifying concept for carcinogenic risk assessment. *J. Theor. Biol.* **105**, 35.

KAUFMAN, J. J. (1983). Quantum chemical and theoretical prediction of toxicity. In *Structure–Activity Correlation as a Predictive Tool in Toxicity* (L. Golberg, ed.). Hemisphere, Washington, D. C.

KENAGA, E. E. (1982). Predictability of chronic toxicity for acute toxicity of chemicals in fish and aquatic invertebrates. *Environ. Toxicol. Chem.* **1**, 347.

KILKENNY, A. E., MORGAN, D., SPANGLER, E. F., AND YUSPA, S. H. (1985). Correlation of initiating potency of skin carcinogens with potency to induce resistance to terminal differentiation in cultured mouse keratinocytes. *Cancer Res.* **45**, 2219.

KUZMACK, A. M., AND MCGAUGHY, R. E. (1975). *Quantitative Risk Assessment for Community Exposure to Vinyl Chloride.* Dec. 5. U. S. EPA, Washington, D. C.

LEDERBERG, J. (1981). *Toxic Substances and the News Media: Scientific and Value Issues.* Presented at the Rockefeller University World Environment Center, January 12.

LEGATOR, M., AND HARPER, B. L. (1982). Animal–Human studies: An integrated approach to identifying hazardous chemicals in the workplace. In *Mutagens in Our Environment* (M. Sorsa and H. Vainio, eds.). Liss, New York.

LIEBLER, D. C., AND GUNGERICH, F. P. (1983). Olefin oxidation by cytochrome *P*-450: Evidence for group migration in catalytic intermediates formed with vinylidene chloride and *Trans*-1 phenyl-1-butene. *Biochemistry* **22**, 5482.

LIJINSKY, W. (1981). Structure–Activity relationships among *N*-nitroso compounds. In *N-Nitroso Compounds* (R. A. Scanlan and S. R. Tannenbaum, eds.), ACS Symposium Series 174. Amer. Chem. Soc., Washington, D. C.

LIJINSKY, W. (1984). *Testimony before the FDA in the FD&C Blue No. 2 Dye Matter.* as quoted in *Food Chemical News*, May 21, pp. 37-8.

LIJINSKY, W., AND REUBER, M. D. (1984). Carcinogenesis in rats by NDMA and other nitrosos at low doses. *Cancer Lett.* **22**, 83.

LIJINSKY, W., KNUTSEN, G. L., AND KOVATCH, R. M. (1985). Carcinogenic effect of nitrosoalkylureas and nitrosoalkylcarbamates in Syrian hamsters. *Cancer Res.* **45**, 542.

LOEW, G. H., REBAGLIATI, M., AND POULSEN, M. (1984). Metabolism and relative carcinogenic potency of chloroethanes: A quantum chemical structure–activity study. *Canad. Biochem. Biophys.* **7**, 109.

LORENTZEN, R. J. (1978). A catalogue of structural moieties required for the activity of chemical carcinogens: A guide for predicting potential new carcinogens as the basis of molecular structure. Bureau of Foods, Food Animal Evaluation Branch.

LORENTZEN, R. J. (1984). FDA procedures for carcinogenic risk assessment. *Food Technol.* **38**, 108.

MA, T. H., HARRIS, M. M., ANDERSON, J. A., AHMED, I., MOHAMMAD, K., BAVE, J. L., AND LIN, G. (1984). Tradescantia-micronucleous (Trad-MCN) tests on 140 health-related agents. *Mutat. Res.* **138**, 157.

MALMFORS, T., AND TEILING, A. (1983). LD_{50}—Its value for the pharmaceutical industry in safety evaluation of doses. *Acta Pharmacol. Toxicol.* **52**(Suppl. 2), 229.

MANTEL, N. (1985). Quantitative approaches for cancer risk assessment. *Risk Anal.* **5**, 5.

MARGOLIN, B. H., KAPLAN, N., AND ZEIGER, E. (1981). Statistical analysis of the Ames *Salmonella*/microsome test. *Proc. Natl. Acad. Sci. USA* **78**, 3779.

MARSHALL, E. (1982). EPA's high-risk carcinogen policy. *Science (Washington, D. C.)* **218**, 975.

MCCANN, J., AND AMES, B. N. (1976). Detection of carcinogen as mutagen in the *Salmonella* microsome test: Assay of 300 chemicals: Discussion. *Proc. Natl. Acad. Sci. USA* **73**, 950.

MCCANN, J., CHAI, E., YAMASAKI, E., AND AMES, B. N. (1975). Detection of carcinogens as mutagens in the *Salmonella* microsome test: Assay of 300 chemicals. *Proc. Natl. Acad. Sci. USA* **72**, 5135.

MCCANN, J., HORN, L., LITTON, G., KALDOR, J., MAGAW, R., BERNSTEIN, L., AND PIKE, M. (1983). Short-term tests for carcinogens and mutagens: A data base designed for cooperative quantitative analysis. In *Structure–Activity Correlation as a Predictive Tool in Toxicology* (L. Golberg, ed.). Hemisphere, Washington, D. C.

MELLINGER, P. J. (1985). Complex mixtures—Relative potency matrix. In *Pacific Northwest Laboratory Animal Report for 1984 to the DOE Office of Energy Research*, Part I, *Biomedical Sciences*, PNL-5500 Pt. I, UC-48, February. p. 89.

MESELSON, M., AND RUSSELL, K. (1977). Carcinogenic and mutagenic potency. In *Origins of Human Cancer*, Book C, *Cold Spring Harbor Conferences on Cell Proliferations* (H. H. Hiatt, J. D. Watson, and J. A. Weinsten, eds.), Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

MEYERS, D. B., YOUNG, S. S., AND GRIES, C. L. (1985). Design of cancer assays for pharmaceutical agents. *J. Natl. Cancer Inst.* **74**, 1151.

MILLER, R. E., AND GUNGERICH, F. P. (1982). Oxidation of trichloroethylene by lower microsomal cytochrome *P*-450: Evidence for chlorine migration in a trasition state not involving trichloroethylene oxide. *Biochemistry* **21**, 1090.

MONSON, R. R. (1980). *Occupational Epidemiology.* CRC Press, Boca Raton, Fla.

Ministerium fur Arbeit und Soziales (1982). Verordung uber gefahrliche Arbietstoffe, Arbeitsstoffverordung, Anhang II, Nr. 1 vom 11.02.1982. Bundegestzblatt I, Seite 144 vom 17.02.1982. West Germany.

National Academy of Science (NAS) Committee on Prototype Explicit Analyses for Pesticides (1980). *Regulating Pesticides.* Natl. Res. Council, Washington, D. C.

National Institute for Occupational Safety and Health (NIOSH) (1983). *Registry of Toxic Effects of Chemical Substances.* DHHS (NIOSH) Pub. No. 83-107. U. S. Govt. Printing Office, Washington, D. C.

National Research Council (NRC), Committee on the Institutional Means for Assessment of Risks to Public Health (1983). *Risk Assessment in the Federal Government: Managing the Process.* Natl. Academy Press, Washington, D. C.

National Toxicology Program (NTP) (1983). *Carcinogenesis Studies of Polybrominated Biphenyl Mixtures (Firemaster FF-1).* Technical Report No. 244, NTP-81-32, June.

National Toxicology Program, Board of Scientific Counselors (1984). *Report of the Ad Hoc Panel on Chemical Carcinogenics Testing and Evaluation.* August 17.

NEAL, R. A., AND GIBSON, J. E. (1984). The uses of toxicology and epidemiology in identifying and assessing carcinogenic risks. In *Reducing the Carcinogenic Risks of Industry* (P. F. Deisler, Jr., ed.). Dekker, New York.

NOHARA, M., HIRAYAMA, T., FUJIOKA, Y., OZASA, S., IBUKI, E., AND FUKUI, S. (1985). Relationship between mutagenic potency in *Salmonella typhimurium* and chemical structure of amino- and nitro-substituted biphenyls. *Mutat. Res.* **149**, 9.

NORDONE, R. M. (1983). The LD50 test and *in vitro* toxicology strategies. *Acta Pharmacol. Toxicol. (Copenhagen* **52**(Suppl. 2), 65.

NTP Panel Refuses to Participate in Study, Agrees to Cooperate on Resigning Alternatives (1985). *Chem. Regul. Reporter* **9**, 118.

Occupational Safety and Health Administration (OSHA) (1978). Occupational exposure standards for benzene. *Fed. Regist.* 43, 5918. Overturned by the Supreme Court, in *Industrial Union Department*, v. *American Petroleum Institute*, 448, U. S. 607, Washington, 1980.

Occupational Safety and Health Administration (1980). Indentification, classification, and regulation of potential occupational carcinogens. *Fed. Regist.* **45**, 5002.

Occupational Safety and Health Administration, Office of Risk Assessment (1984). *Preliminary Assessment of the Health Effects of Formaldehyde,* November 5. Exhibit 43 in Docket H-225, 50FR1547.

Office of Science and Technology Policy (OSTP) (1985). Chemical carcinogens: Review of the science and its associated principles. *Fed. Regist.* **50**, 21594.

OLSEN, J. H., PLOUGH JENSEN, S., HINCK, M., FAURBO, K., BREUM, N. O., AND MILLER JENSEN, O. (1984). Occupational formaldehyde exposure and increased nasal cancer risk in man. *Int. J. Cancer* **34**, 639.

OSTERMAN-GOLKAR, O. (1984). Reaction kinetics in water of chloroethylene oxide, chloroacetaldehyde, and chloroacetone. *Hereditas (Lund. Swed.)* **101**(1), 65.

PARK, C. N., AND SNEE, R. N. (1984). Quantitative risk assessment: state-of-the-art for carcinogens. *Fundam. Appl. Toxicol.* **3**, 320.

PARODI, S., TANINGHER, M., BALBI, C., AND SANTI, L. (1982). Predictive ability of the autoradiographic repair assay in rat liver cells compared with the Ames test. *J. Toxicol. Environ. Health* **10**, 531.

PARODI, S., TANINGHER, M., AND SANTI, L. (1983). Induction of preneoplastic nodules: Quantitative predictivity of carcinogenicity. *Anticancer Res.* **3**, 393.

PARODI, S., TANINGHER, M., RUSSO, P., PALA, M., VECCHIO, D., FASSINA, G., AND SANTI, L. (1983). Quantitative predictivity of the transformation *in vitro* assay compared with the Ames test. *J. Toxicol. Environ. Health* **12**, 483.

PASET, G. (1983). The LD50 test. *Acta Parmacol. Toxicol.* **52**(Suppl. 2), 6.

PETO, R., PIKE, M. C., BERNSTEIN, L., GOLD, L. S., AND AMES, B. N. (1984). The TD_{50}: A proposed general convention for the numerical description of the carcinogenic potency of chemicals in chronic-exposure animal experiments. *Environ. Health Perspect.* **58**, 1.

PITOT, H. C. (1983). Contributions to our understanding of the natural history of neoplastic development in lower animals to the cause and control of human cancer. *Cancer Surveys* **2**, 519.

POLITZER, P., AND LAWRENCE, P. R. (1984). Relationship between the electrostatic potential, epoxide hydrase

inhibition and carcinogenicity for some hydrocarbons and halogenated hydrocarbon epoxides. *Carcinogenesis* **5**, 845.

POPPER, H., THOMAS, L. B., TELLES, N. C., FALK, H., AND SELIKOFF, I. J. (1978). Development of hepatic angiosarcoma in man induced by vinyl chloride, thoratrast, and arsenic. *Amer. J. Pathol.* **92**, 349.

PORTIER, C. J., AND HOEL, D. G. (1984). Design of animal carcinogenicity studies for goodness-of-fit of multistage models. *Fundam. Appl. Toxicol.* **4**, 949.

PREHN, L. M., AND LAWLER, E. M. (1979). Rank order of sarcoma susceptibility among mouse strains reverses with low concentrations of carcinogens. *Science (Washington, D. C.)* **204**, 309.

PURCHASE, I. F. (1980). Interspecies comparisons of carcinogenicity. *Brit. J. Cancer* **41**, 454.

PURCHASE, I. F. H., LONGSTAFF, E., ASHBY, J., STYLES, J., ANDERSON, D. LEFEVRE, P., AND WESTWOOD, F. (1978). An evaluation of 6 short-term tests for detecting organic chemical carcinogens. *Brit J. Cancer* **37**, 876.

RAMEL, C. (1983). Advantages of and problems with short-term mutagenicity tests for the assessment of mutagenic and carcinogenic risk. *Environ. Health Perspect.* **47**, 153.

REITZ, R. H., GEHRING, P. J., AND PARK, C. N. (1978). Carcinogenic risk estimation for chloroform: An alternative to EPA's procedures. *Food Cosmet. Toxicol* **16**, 511.

REKKER, R. F. (1982). LD_{50} values by computer? *Chem. Eng. News* **60**, 59.

RIGGAN, W. B., BRUGGEN, J. V., ACQUAVELLA, J. F., BEAUBIER, J., AND MASON, T. J. (1983). *U.S. Cancer Mortality Rates and Trends, 1950–1979*, EPA-600/1-83-015a-c, September. U. S. Govt. Printing Office, Washington, D. C.

RINKUS, S. J., AND LEGATOR, M. S. (1979). Chemical characterization of 465 known or suspected carcinogens and their correlation with mutagenic activity in the *Salmonella tymphimurium* system. *Cancer Res.* **39**, 3289.

RINKUS, S. J., AND LEGATOR, M. S. (1981). *Salmonella* Revisited: A reply to Ames and McCann. *Cancer Res.* **41**, 4196.

REYNOLDS, E. S., AND MOSLEM, M. T. (1977). Damage to hepatic cellular membranes by chlorinated olefins with emphasis on synergism and antagonism. *Environ. Health Perspect.* **21**, 137.

ROSENKRANZ, H. S. (1977). Mutagenicity of halongenated olefins and their derivatives. *Environ. Health Perspect.* **21**, 79.

ROSENKRANZ, H. S., KLOPMAN, G., CHANKONG, V., PET-EDWARDS, J., AND HAIMES Y. Y. (1984). Prediction of environmental carcinogens: A strategy for the mid-1980's. *Environ. Mutagen* **6**, 231.

ROWAN, A. (1983). Shortcomings of LD_{50}-values and acute toxicity testing in animals. *Acta Pharmacol. Toxicol.* **52**(Suppl. 2), 52.

RUCKELSHAUS, W. D. (1985). Risk, science and democracy. *Issues Sci. Technol.* Spring, p. 19.

RUSSELL, L. B., AARON, C. S., DE SERRES, F., GENEROSO, W. M., KANNAN, K. L., SHELBY, M., SPRINGER, J., AND VOYTEK, P. (1984). Evaluation of mutagenicity assays for the purpose of genetic risk assessment. *Mutat. Res.* **134**, 143.

RUSSELL, W. L., HUNSICKER, P. R. RAYMER, G. D., STEELE, M. H., STELZNER, K. F., AND THOMPSON, H. M. (1982). Dose response curve for ethylnitrosourea-induced specific locus mutations in mouse spermatogenia. *Proc. Natl. Acad. Sci. USA* 79, 3589.

SALSBURG, D. (1983). The lifetime feeding study in mice and rats—An examination of its validity as a bioassay for human carcinogens *Fundam. Appl. Toxicol.* **3**, 63.

SAWYER, C., PETO, R., BERNSTEIN, L., AND PIKE, M. C. (1984). Calculations of carcinogenic potency from long-term animal carcinogenicity experiments. *Biometrics* **40**, 27.

Science Advisory Board (1985a). *Closure Letter to W. D. Ruckelshaus on the Health Assessment Document for 1,2-Dichloroethane*, January 4.

Science Advisory Board (1985b). *Closure Letter to W. D. Ruckelshaus on the Health Assessment Document for Chromium*, January 4.

SHELBY, M. D., AND STASIEWICZ, S. (1984). Chemicals showing no evidence of carcinogenicity in long-term two-species rodent studies: The need for short-term test data. *Environ. Mutagen.* **6**, 871.

SHIMKIN, M. D., WEISBURGER, J. N., WEISBURGER, E. K., GUBANEFF, N., AND SWARTZEFF, V. (1966). Bioassays of 29 alkylating chemicals by the pulmonary-tumor response in strain A mice. *J. Natl. Cancer Inst.* **36**, 915.

SHUBIK, P. (1984). Carcinogenicity evaluation and regulatory decisions. *Regul. Toxicol. Pharmacol.* **4**, 322.

SIELKIN, R. L., JR. (1985). *Letter to R. McGaughy, U. S. EPA, Containing Comments on the Proposed Guidelines for Carcinogen Risk Assessment at 49FR46294*, January 22.

SINGH, B., AND GUPTA, R. S. (1983). Mutagenic responses of 13 anticancer drugs on mutation induction at multiple genetic loci and sister chromatid exchanges in Chinese hamster ovary cells. *Cancer Res.* **43**, 577.

SIVAK, A. (1979). Overview and status of *in vitro* transformation. *J. Assoc. Off. Anal. Chem.* **62**, 889.

SIVAK, A., AND TU, A. S. (1980). Factors influencing neoplastic transformation by chemical carcinogenesis in BALB/c-3T3 Cells. In *The Predictive Value of Short-Term Screening Tests in Carcinogenicity Evaluation* (G. M. Williams, *et al.*, eds.). Elsevier/North Holland Biomedical Press, New York.

SLAGA, T. J. (1983). Overview of tumor promotion in animals. *Environ. Health Perspect.* **50**, 1983.

SQUIRE, R. A. (1981). Ranking animal carcinogens: A proposed regulatory approach. *Science (Washington D. C.)* **214**, 877 and 2081.

SQUIRE, R. A. (1984). Carcinogenicity testing and safety assessment. *Fundam. Appl. Toxicol.* **4**, S326.

SUGIMURA, T., AND TAKAYOMA, S. (1983). Biological actions of nitroamines in short-term tests on *Salmonella*, cultured mammalian cells and cultured human tracheal tissues: Possible basis for regulating control. *Environ. Health Perspect.* **47**, 171.

SUTER, G. W., II, VAUGHAN, D. S., AND GARDNER, R. H. (1983). Risk assessment by analysis of extrapolation error: A demonstration for effects of pollutants on fish. *Environ. Toxicol. Chem.* **2**, 369.

SWENBERG, J. A., KERNS, W. D., MITCHELL, R. I., GRALLA, E. J., AND PAVKOV, K. L. (1980). Induction of squamous cell carcinoma of the rat nasal cavity by inhalation exposure to formaldehyde vapor. *Cancer Res.* **40**, 3398.

TATES, A. D. (1982). Comparative chemical mutagenesis on the basis of molecular dosimitry. In *Mutagens in Our Environment* (M. Sorsa and H. Vainio, eds.). Liss, New York.

THEISS, J. C. (1983). The ranking of chemicals for carcinogenic potency. *Regul. Toxicol. Pharmacol.* **3**, 320.

TOBIN, P. S., KORNHAUSER, A., AND SCHEUPLEIN, R. J. (1982). An evaluation of skin painting studies as determinants of tumorigenesis potential following skin contact with carcinogens. *Regul. Toxicol Pharmacol.* **2**, 22.

TOFILON, P. J., BASIC, I., AND MILAS, L. (1985). Prediction of *in vivo* tumor response to chemotherapeutic agents by the *in vivo* sister chromatid exchange assay. *Cancer Res.* **45**, 2025.

TROSKO, J. E. (1984). A new paradigm is needed for toxicology evaluation. *Environ. Mutagen* **6**, 767.

U. S. Department of Agriculture (1985). Compound evaluation system, review draft. As described in *Pestic. Toxic News*, April 10, p. 15.

VANCE W. A., AND LEVIN, D. E. (1984). Structural features of nitroaromatics that determine mutagenic activity in *Salmonella tymphimurium*. *Environ. Mutagen.* **6**, 797.

VAN DUUREN, B. L. (1977). Chemical structure, reactivity, and carcinogenicity of halohydrocarbons. *Environ. Health Perspect.* **21**, 17.

VAN DUUREN, B. L., KLINE, S. A., MELCHIONNE, S., AND SEIDMAN, I. (1983). Chemical structure and carcinogenicity relationships of some chloroalkene oxides and their parent olefins. *Cancer Res.* **43**, 159.

VAN RYZIN, J. (1980). Quantitative risk assessment. *J. Occup. Med.* **22**, 321.

VON WHITTENAU, M. S., AND ESTES, P. C. (1984). The redundancy of the mouse carcinogenicity bioassays. *Fundam. Appl. Toxicol.* **3**, 631.

WANG, G. M. (1984). Evaluation of pesticides which pose carcinogenicity potential in animal testing. 1. Developing a tumor evaluation system. *Regul. Toxicol. Pharmacol.* **4**, 355.

WEIL, C. S. (1984). Some questions and opinions on issues in toxicology and risk assessment. *Amer. Ind. Hyg. Assoc. J.* **45**, 663.

WEISBURGER, J. H., AND WILLIAMS, G. M. (1981a). Carcinogen testing: Current problems and new approaches. *Science (Washington, D. C.)* **214**, 401.

WEISBURGER, J. H., AND WILLIAMS, G. M. (1981b). The decision-point approach for systematic carcinogen testing. *Food Cosmet. Toxicol.* **19**, 561.

WEISS, W. (1982). Epidemic curve of respiratory cancer due to chloromethyl ethers. *J. Natl. Cancer Inst.* **69**, 1265.

WHITE, M. C., INFANTE, P. F., AND CHU, K. C. (1982). A quantitative estimate of leukemia mortality associated with occupational exposure to benzene. *Risk Anal.* **2** 195. See also comments on this paper, and response by the authors, in Letter to the Editor section, *Risk Anal.* **4**, 1–11. (1984).

WILSON, R. (1980). Risk benefit analyses for toxic chemicals. *Exotoxicol. Environ. Saf.* **4**, 370.

WISHNOK, J. S. (1981). Effect of structure on the carcinogenic behavior of nitrosoamines. In *N-Nitroso Compounds* (R. A. Scanlan and S. R. Tannenbaum, eds.), ACS Symposium Series 174. Amer. Chem. Soc., Washington, D. C.

WOLD, S., HELBERG, S., AND DUNN, W. J., III (1983). Computer methods for the assessment of toxicity. *Acta Pharmacol. Toxicol.* **52**(Suppl. 2), 158.

ZEISE, L., WILSON, R., AND CROUCH, E. (1984). Use of acute toxicity to estimate carcinogenic risk. *Risk Anal.* **4**, 187.