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
the International Affairs Group of CMA/SOCMA
and
Canadian Chemical Producers Association

August 21, 1985

August 21, 1985

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
the International Affairs Group of CMA/SOCMA and CCPA*

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 documents offering consistent guidance for identifying and classifying carcinogens, mutagens, and teratogens. These documents 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 relevant 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. They may serve as a consistent and harmonized basis upon which appropriate health protection and regulatory measures may be developed.

The set of guidelines developed are based on existing relevant documents which exist worldwide and which vary in detail and scope. Appendix II is a list of relevant reference documents and Appendix III 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
- Criteria for Inferring Causality from Epidemiological Studies.

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

---

* See Appendix I for List of Participants

CRITERIA FOR CLASSIFYING CHEMICALS AS CARCINOGENS

I. INTRODUCTION

The purpose of this document is to provide guidance on the classification of carcinogens in order that available human and animal data on proven and suspect carcinogens are appropriately interpreted as to their relevance to humans.

No attempt has been made to review in this document the current state of the science of carcinogenicity. This has been done recently by OSTP (1). It is assumed, however, that all available scientific data relevant to the carcinogenicity of a substance has been assembled before beginning an assessment of that 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 (2). 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 indisputably carcinogenic to humans under expected conditions of exposure, it will 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, but these properties should be considered when making risk management decisions.

II. GENERAL REMARKS ON CLASSIFYING CARCINOGENS

The majority of classifications currently employed EEC, (3) EPA, (4) IARC, (5) ECETOC (6) envisage three broad categories of carcinogens. 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.

Table I is a summary of the comparison of the proposed classification system by the Joint Work Group on Chronic Hazards to 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.

III. CRITERIA FOR CLASSIFICATION

Category 1

Proven Human Carcinogenic Substance

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

Category 2

Suspected Human Carcinogenic Substance

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

- suggestive epidemiological data not sufficient to satisfy the criteria for establishing causality, as described in the paper on epidemiology

and

- proven evidence from animal studies carried out under conditions which are relevant to human exposure.

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 below.

Category 3

Proven Animal Carcinogenic Substance 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 that case, it may be assumed that human exposure to the substance might result in the development of cancer.

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 man:

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% (2,6). 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.

b) The induction of an excess of malignant tumors at more than one site.

c) The existence of a clear dose-response relationship in the number and the 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 an 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,
ii) only an excess of tumors such as hepatic nodules in rats or mice, or only pulmonary tumors in mice is induced,
iii) only a small excess of malignant tumors is induced in an organ which has a high spontaneous incidence,

iv) an excess of malignant tumors is induced only by an inappropriate route of administration (see comments below),

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 physico-chemical properties, or

vi) tumors are induced in experimental animals only at exposure levels which produce chronic injury in the tissue or organ in which tumors later appear, with no tumors developing at exposure levels at or above that which is likely to occur in man but which do not produce local injury.

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. 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 series of chemicals which are very closely related in structure.

Category 4

Suspected Animal Chemical Carcinogens Substance 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 suggests that this incidence could have occurred by chance, particularly if this is only towards the end of the animal's natural lifespan.

b) An increased incidence of malignant tumors in only one species or strain with no increase in incidence in other species or strains.

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

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

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

It should be noted that results from some routes of administration e.g. intragastric intubation, and subcutaneous, intravenous and intraperitonal treatment are typically not reliable indicators of carcinogenicity. The gavage and particularly the subcutaneous route can give false positive findings and much caution should be exercised in interpreting the data especially when only local tumor formation is induced.

Category 5

Non-Classifiable Substances 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:

a) The only positive evidence of carcinogenicity is from an animal study.

   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 urothelium induced by the presence of calculi. Some groups have suggested dosages above which no practical significance should be attached (7).

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) 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 inadequately performed, had improper controls, 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 that of man.

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" (4). A more extensive discussion of this subject is provided in the attached epidemiology paper.

Category 7

No Evidence

In this case no relevant data are available.

TABLE I

COMPARISON OF VARIOUS CLASSIFICATION SCHEMES FOR CARCINOGENS

This Document	EEC[3]	EPA[4]	IARC[5]
1 Proven Human Carcinogenic Substance	1	A	1
2 Suspected Human Carcinogenic Substance	2	B1	2A
3 Proven Animal Carcinogenic Substance with potential Relevancy to Humans	2	B2	2B
4 Suspect Animal Carcinogenic Substance with Possible Relevance to Human	3	C	3
5 Nonclassifiable Substances with Regard to Carcinogenicity	-	D	4
6 Negative Evidence	-	E	-
7 No evidence	-	-	-

REFERENCES

1) "Chemical Carcinogens; A Review of the Science and its Associated Principles," Office of Science and Technology Policy, Washington, D.C.; F.R. Vol. 50, 14 March, 1985.

2) "Report of the NTP Ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation," U.S. Dept. of Health & Human Services; 17, August, 1984.

3) "Classification and Labelling Dangerous Substances" (Directive 83/467/EEC) Official Journal of the European Economic Communities (EEC); No. L 257/24; 16 September, 1983.

4) "Proposed Guidelines for Carcinogenic Risk Assessment," Environmental Protection Agency (EPA), F.R. Vol. 49, No. 227, 23 November, 1984.

5) "Polynuclear Aromatic Hydrocarbons," Part 2; International Agency for Research on Cancer (IARC), Volume 33 of Monograph Series; Lyon, France; April 1984.

6) "Risk Assessment of Occupational Chemical Carcinogens" (ECETOC), Monograph No. 3, Brussels; January, 1982.

7) "Threshold Limit Values (TLV's) for Chemical Substances and Physical Agents in the Work Environment," American Conference of Governmental Industrial Hygienists (ACGIH); Cincinnati, Ohio, 1984.

CRITERIA FOR CLASSIFYING CHEMICALS AS MUTAGENS

I. INTRODUCTION

The classification of chemicals according to their mutagenic properties into various categories as described below, 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.

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 in 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 one percent of all newborns bear chromosome abnormalities of some type (1,2). This high background incidence of genetic damage and problems in the identification and classification of mutagenic effects in man makes the epidemiologic investigation of heritable effects caused by one specific agent very difficult (3). It should be noted that direct evidence for the occurrence of a causal relationship linking chemical exposure and 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. Animals studies and supporting *in vitro* studies are currently used to predict genetic hazard associated with chemical exposure.

These 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.

II. GENERAL REMARKS ON CLASSIFYING

1. In the classification of mutagens only the potential hazard for man 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 that a response is only observed under conditions 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 a criterion for classification purposes.

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

III. CLASSIFICATION CRITERIA

EEC, USA and Canadian documents for classifying mutagenicity data were reviewed. Similarities between the systems were identified and a single system was developed for classifying substances according to relative strength of evidence for potential human germ cell mutagenicity.

These categories show concordance with the EEC classification scheme as indicated in Table I and also share elements with the EPA categories of sufficient, suggestive and limited evidence for mutagens (4,5).

TABLE I

This Document	EEC Scheme	EPA Scheme
Category 1	Category 1	-
2	2	Sufficient Evidence
3	3	Suggestive Evidence
4	-	Limited Evidence
5	-	-

Five categories have been developed in the following classification new classification scheme, and criteria are listed according to their degree of relative weight for use in evaluating the available data. Although, multiple criteria within some categories are weighted for emphasis, evidence satisfying any criterion is sufficient to classify a chemical according to that category.

Category 1

Evidence for Substances with Human Germ Cell Mutagenicity

There is sufficient evidence to establish a causal connection between human exposure to chemical and heritable genetic effects.* This evidence can only be developed by

*At this time, no causal relationships have been described.

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.

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** with the genetic material of mammalian germ cells in vivo, plus clearly positive results in at least two valid studies assessing either gene mutation or chromosome aberrations or one valid positive finding of gene mutation and one valid positive finding of chromosome aberrations. At least one of the positive findings*** must be from human (epidemiological) studies or from tests with other mammals in vivo.

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 two valid mutagenicity studies assessing either gene mutation or chromosome aberrations or one valid positive finding of gene mutation and one valid positive finding of chromosome aberrations. At least one of the positive results must be from human (epidemiological) studies*** with humans or from tests with other mammals in vivo.

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

*** See attached paper for Inferring Causality from Epidemiological Studies.

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,

3) When there are positive tests which would otherwise satisfy the criteria for Category III but with evidence from appropriate tests that the chemical does not interact with the genetic material of mammalian germ cells in vivo.

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 mammalian germ cells in vivo.

REFERENCES

1) Evans, H. J. "Structure and Organization of the Human Genome," Mutations in Man edited by G. Obe; Springer-Verlag, New York; pp. 58-100, 1984.

2) United Nations Scientific Committee to Evaluate Energy and Radiation, Sources and Affects of Ionizing Radiation (UNSCEAR); United Nations, New York; 1977.

3) "Identifying and Estimating the Genetic Impact of Chemical Mutagens," National Research Council; National Academy Press; Washington, D.C.; pp. 296, 1983.

4) "Classification and Labelling of Dangerous Substances" (Directive 83.467/EEC) Official Journal of the European Economic Communities (EEC); No. L 257/24; 16 September, 1983.

5) "Proposed Guidelines for Carcinogenic, Mutagenic and Reproductive Risk," Environmental Protection Agency (EPA); FR Vol. 49, No. 227, 23 November, 1984.

CRITERIA FOR CLASSIFYING CHEMICALS AS TERATOGENS

I. INTRODUCTION

There is a need to develop criteria for classifying those chemicals that cause adverse effects on the reproductive process. As a first step, this document offers guidance for classifying chemicals on there 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 reproducably 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 latter 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.

In this case, the following three definitions have been agreed by many specialists in the fields of toxicology and occupational medicine and are not believed to conflict with those derived from 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 exposue during pregnancy. These effects may be manifested in the embryonic or foetal periods, or postnatally.

One specific manifestation of developmental toxicity is the production of deleterious, structural malformations. 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 judgement 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 (Annex VI d) which refers specifically to 'teratogens' (and requires the phrase 'May cause birth defects' (R 47). Thus the other types of developmental toxicity remain unclassified for the present as morphological change is "the only reproducible assessment currently available" (2).

II. PROPOSED CLASSIFICATION

The following categories of substances are proposed:

Category 1: Substances known to be teratogenic to man (humans).

Category 2: Substances which should be regarded as if they are teratogenic to man (humans).

Category 3: Substances which cannot be classified based on available data.

This categorization, with respect to categories 1 and 2, is consistent with that contained in the EEC Labelling Guide (3).

III. CRITERIA FOR CLASSIFYING TERATOGENS

Category 1: Substances known to be teratogenic to man.

For Category 1, the accompanying description 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 now 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 attached paper on epidemiology.) The high 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 substances 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 interperetation '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) relevant route of exposure, i.e., for industrial chemicals, dermal, inhalation or oral,

b) exposure during pregnancy/organogenesis only,

c) structural malformation 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 mal- formations themselves must be of biological significance,

e) good historical background data should exist on the usefulness of the species tested. Data should be for same strain and from 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 methods and sound scientific practices.

In the evaluation of animal studies against the above criteria, particular consideration should be given to excluding maternal toxicity as a confounding factor. Positive teratogenic findings seen only in the presence of maternal toxicity do not 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 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 man.

If the above criteria are not met, then insufficient evidence exists to categorize the substance in respect of 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. The group proposed that in the conduct or evaluation of a study, an upper

dose limit, e.g., 1000 mg/kg/day should be established since effects seen above such a dose are 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 descrived above for Categories 1 and 2, these data should be considered non-classifiable: i.e. inadequate evidence to assign a chemical substance to either Category 1 or 2.

REFERENCES

1) "Principles for Evaluating Health Risks to Progeny Associated with Exposure to Chemicals During Pregnancy," World Health Organization, Environmental Health Criteria No. 30; Geneva, 1984.

2) Commission of the European Communitites (DG XI) Report of 2nd Meeting of Specialized Experts in the Field of Carcinogenic/ Mutagenic/Teratogenic Substances. Brussels. 18/19 April 1985.

3) "Classification and Labelling of Dangerous Substances," (Directive 83/467/EEC) Official Journal of the European Economic Communities (EEC); No. L 257/24; 16, September, 1983.

CHRONIC HAZARDS ASSESSMENT: CRITERIA FOR INFERRING CAUSALITY FROM EPIDEMIOLOGICAL STUDIES

I. INTRODUCTION

In establishing hazard identification procedures for chronic 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.

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. Criteria that are too general may allow so much room for differing interpretations that they can lead to widely differing conclusions. If there are to be harmonized warnings on 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 chronic 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 or death rate among exposed persons may be due to confounding factors that are unknown or known but could not 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 warnings on labels.

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 above comments relate to epidemiological studies of carcinogenicity, but the principles apply equally to studies of mutagenicity and teratogenicity in humans.

II. CRITERIA TO EVALUATE THE QUALITY OF INDIVIDUAL 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 determining 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 studies, 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.

III. 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.

IV. 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, in its Proposed Guidelines for Carcinogenic Risk Assessment (1984) has adopted the IARC Criteria shown above.

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."(1)

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 on 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.[2] We recommend that these criteria be adopted.* 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..."[2] 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 precedes 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.

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.

---

*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.[3] The criteria were also used in the Surgeon General's 1982 report on smoking and cancer.[4]

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.

REFERENCES

1. ECETOC: "Risk Assessment of Occupational Chemical Carcinogens," Page 13, Monograph No. 3, Brussels; January, 1982.

2. Hill, AB: Principles of Medical Statistics, 9th Edition, Oxford University Press; pp. 309-320, New York, 1971.

3. Advisory Committee to the Surgeon General of the Public Health Service: Smoking and Health, U.S. Public Health Service, U.S. Public Health Service, pp. 182-189, 1964.

4. Report of the Surgeon General: The Health Consequences of Smoking - Cancer, U.S. Public Health Service, pp. 16-20, 1982.

Appendix I

Participants in the Tripartite
Working Party on
Chronic Hazard Identification

Co-Chairmen

Dr. B. Broecker - Hoechst AG
CEFIC/ECETOC

Dr. E. J. Sowinski - UNIROYAL, Inc.
CMA/IAG

Carcinogenic Properties

Dr. J. M. Faccini -Consultant
ECETOC

Mr. S. Beekhuizen - Bayer AG
CEFIC

Mr. J. T. Barr - Air Products
AIHC

Mr. C. S. Weil - Consultant
AIHC

Mutagenic Properties

Dr. P. Gelbke and
Dr. H. Fleig - BASF
ECETOC

Dr. G. C. Granville - Shell Canada
CCPA

Dr. D. W. Matheson - Stauffer
Chemical
AIHC

Dr. W. Rozenboom - Shell Europe
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. J. C. Aubrun - Rhone-Poulenc
CEFIC

Dr. A. Munn - Monsanto
Europe
CEFIC

Dr. M. J. Reape - FMC, Inc.
AIHC

Staff

Dr. H. Haeseler	CEFIC
Mr. G. W. Ingle	CMA
Mr. C. F. Emde	CMA

Footnote

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

Appendix II

REFERENCES

1) "Classification and Labelling of Dangerous Substances" (Directive 83.467/EEC) Official Journal of the European Economic Communities (EEC); No. L 257/24; 16 September, 1983.

2) "Proposed Guidelines for Carcinogenic, Mutagenic and Reproductive Risk," Environmental Protection Agency (EPA); FR Vol. 49, No. 227, 23 November, 1984.

3) "Hazard Communication; Final Rule," U.S. Department of Labor: Occupational Safety and Health Administration (OSHA) FR Vol. 48, No. 228, 25 November, 1983.

4) "Workplace Hazardous Materials Information System" (WHMIS), Report to the Project Steering Committee: OSH Branch Labour Canada; Ottawa, Canada; April, 1985.

5) International Agency for Research on Cancer (IARC); Lyon, France; Monograph 33, April 1984 and Annual Report, 1984.

6) National Toxicology Program (NTP) 3rd Annual Report on Carcinogens; U.S. Department of Health & Human Services; September 1983.

7) "Report of the NTP Ad Hoc Panel on Chemical Carcinogenesis Testing and Evaluation"; U.S. Dept. of Health & Human Services; 17 August, 1984.

8) "Threshold Limit Values (TLV's) for Chemical Substances and Physical Agents in the Work Environment," American Conference of Governmental Industrial Hygienists (ACGIH); Cincinatti, Ohio; 1984.

9) ECETOC "A Contribution to the Strategy for Identification and Control of Occupational Carcinogens" Monograph No. 2, Brussels, September 1980.

10) ECETOC "Risk Assessment of Occupational Chemical Carcinogens" Monograph N. 3, Brussels, January 1983.

11) ECETOC "Identification and Assessment of the Effects of Chemicals on Reproduction and Development" Monograph No. 5, December 1983.

12. American National Standard for Hazardous Industrial Chemicals--Precautionary Labeling: ANSI Z 129.1-1982, New York; 21 September, 1982.

Appendix III

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

Carcinogens:

EEC	EPA	OSHA	WHMIS
Category 1 Known to be carcinogenic to man.	Group A Human 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.
Category 2 Regarded as carcinogenic to man based on animal studies	Group B Probable human carcinogen		
Category 3 Cause concern owing to possible carcinogenic effects.	Group C Possible human carcinogen		

Mutagens:

EEC	EPA	OSHA	WHMIS
Category 1 Known to be mutagenic to man.		No categorization Based on hazard evaluation of all relevant data.	Not addressed at this time.
Category 2 Regarded as mutagenic to man.	Sufficient evidence		
Category 3 Cause concern owing to possible mutagenic effects.	Suggestive evidence Limited evidence		

Teratogens:

EEC	EPA	OSHA	WHMIS
Category 1 Known to be 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.
Category 2 Regarded as teratogenic to man.			

*References 1, 2, 3 and 4 from Appendix II