Environmental Health Perspectives
Vol. 96, pp. 5-9, 1991

Carcinogenicity Evaluations and Ongoing Studies: The IARC Databases

by Harri Vainio,* Michel Coleman,† and Julian Wilbourn*

Many thousands of chemicals are produced industrially and many more occur naturally. Information on the toxicology of these chemicals is often minimal or absent. The International Agency for Research on Cancer (IARC) has published evaluations of the carcinogenic risk to humans of over 700 chemicals, groups of chemicals, and complex mixtures as a regular series of monographs. A database has been created containing summaries of all the relevant epidemiological, animal carcinogenicity, and other relevant biological data for each chemical or mixture evaluated. Additional databases have been created for ongoing epidemiological studies of cancer in humans and for long-term carcinogenicity studies in rodents, as well as a database containing information on genotoxic and related effects of chemicals. Some of these databases have been published in print form. IARC now plans to publish them electronically, together with other databases, in the form of a CDROM (compact disk, read-only memory). The objective will be to make the entire IARC database of cancer information as widely available as possible in an integrated format conducive to efficient and combined exploitation of all the component databases.

Introduction: Chemicals, Carcinogens, and Hazards to Humans

Estimates vary as to the number of chemicals that are produced in significant amounts. In 1980, the European Core Inventory (ECDIN) listed about 34,000 chemical substances believed to be on the European Community market (*1*). Later, Shulze and Mücke (*2*) estimated that there were about 100,000 chemicals on the market within the European Community. Recently, it has been estimated on the basis of the U.S. Toxic Substance Control Act inventory that people in the United States are exposed to about 66,000 chemical substances (*3*), and some 300 to 700 new industrial chemicals are introduced annually into economic use (*4*). The fact that even minimal toxicity data are available for only a minority of these chemicals is due to the rapid development of new industrial activities during a period in which health problems related to chemicals received little attention. Several thousand chemicals on the market have been tested to some extent for carcinogenicity in long-term animal studies; however, the number of chemicals tested adequately and for which there are publications in the open literature is much lower. It should be noted that a small proportion of all the chemicals accounts for most of the production: for instance, 380 primary derivatives of petroleum products account for more than half of the production tonnage of all chemicals (*5*).

The International Agency for Research on Cancer (IARC) has been involved in preparing and disseminating scientific information relevant to cancer control for more than 25 years. During this period, extensive collections of data on carcinogenicity and related effects of chemicals and specific exposures have been made available to the scientific community as a series of over 50 volumes of *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans* (*6*). During the preparation and periodical updating of these carcinogenicity evaluations, several ancillary databases have been developed and maintained alongside the *IARC Monographs* themselves, mainly for internal use. In this brief review, we describe the nature and content of these databases and discuss some of the perspectives for their future development.

The IARC Databases

The *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*

The objective of the *IARC Monographs* program is to publish critical reviews of data on carcinogenicity for chemicals and complex mixtures to which humans are known to be exposed and on specific cultural or occupational exposures; to evaluate these data in terms of human risk with the help of international working groups of experts in chemical carcinogenesis and related fields; and, in some cases, to indicate where additional research efforts are needed. By the end of 1990, 50 volumes of the *IARC Monographs* (*6*) had been published, and three more were in preparation, containing detailed evaluations of carcinogenicity

*Unit of Carcinogen Identification and Evaluation, International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon Cedex 08, France.
†Unit of Descriptive Epidemiology, International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon Cedex 08, France.
Address reprint requests to H. Vainio, Unit of Carcinogen Identification and Evaluation, International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon Cedex 08, France.

on 732 chemicals, groups of chemicals, complex mixtures and cultural/occupational exposure circumstances.

Until 1987, evaluations of carcinogenicity were made separately on the human and experimental data: no attempt was made to produce an overall evaluation of carcinogenicity to humans. At a special meeting in 1987, re-evaluations of available new data were made on 189 agents, mixtures, or exposure circumstances for which there were at least some data available on cancer in humans, taking account of all available evidence. At the same time, overall evaluations of carcinogenicity to humans were made for all 628 agents/mixtures/exposures (comprising more than 700 chemicals, groups of chemicals, mixtures, or cultural and occupational exposures) evaluated so far in the program. Overall evaluations of carcinogenicity to humans have been made by each working group from volume 43 onwards.

This evaluation process has resulted in categorization of the carcinogenicity of each chemical, group of chemicals or complex mixture or exposure circumstance into one of five categories: group 1, the agent, mixture, or exposure circumstance is carcinogenic to humans (a causal relationship has been established between exposure and human cancer); group 2A, the agent, mixture or exposure circumstance is probably carcinogenic to humans (a positive association has been observed between exposure and human cancer for which a causal interpretation is credible, but chance, bias, or confounding could not be ruled out with reasonable confidence; there is also sufficient evidence of carcinogenicity in experimental animals. Exceptionally, this category may be used when there is sufficient evidence of carcinogenicity in animals strengthened by supporting evidence from other relevant data); group 2B, the agent, mixture or exposure circumstance is possibly carcinogenic to humans (there is sufficient evidence of carcinogenicity in experimental animals, but no adequate data on cancer in exposed humans; in some instances, an agent for which there is inadequate or no data in humans, but with limited evidence of carcinogenicity in experimental animals, may be placed into this category when there is supporting evidence from other relevant data); group 3, the agent, mixture or exposure circumstance is not classifiable as to its carcinogenicity to humans (this group applies when no other category is used); group 4, the agent, mixture or exposure circumstance is probably not carcinogenic to humans (there is evidence suggesting lack of carcinogenicity in humans together with evidence suggesting lack of carcinogenicity in experimental animals; in some instances, an agent for which there is inadequate evidence or no data on carcinogenicity in humans but evidence suggesting lack of carcinogenicity in experimental animals, consistently and strongly supported by a broad range of other relevant data, may be placed in this category). Table 1 gives the distribution of all the agents evaluated up to 1990 (*IARC Monographs* volumes 1–53) into these five categories.

The categorizations given in Table 1 are clearly related to exposure and not to a specific target organ. This is because the primary goal of the *IARC Monographs* program is to reach an evaluation of carcinogenicity *per se*, and the working procedures used do not embody provision to classify exposures as to their carcinogenicity for specified target organs. However, for an agent to be classified in group 1, there must be at least one target organ for which sufficient evidence of carcinogenicity in humans is judged to exist (Table 2). Similarly, for all agents judged to have sufficient evidence of carcinogenicity in experimental animals, the judgment is usually based on evidence from one or more species. Target organs may differ from one species to another for many exposures causally related to human cancers; however, there is almost always at least one organ in common between humans and at least one animal species, despite many inherent physiological differences between species (*7*).

Table 1. Overall evaluations of carcinogenicity to humans in the *IARC Monographs* program (volumes 1–53): distribution of agents and exposures by category.[a]

Carcinogenic to humans (group 1)	55
Probably carcinogenic to humans (group 2A)	45
Possibly carcinogenic to humans (group 2B)	191
Cannot be classified as to its carcinogenicity to humans (group 3)	440
Probably not carcinogenic to humans (group 4)[b]	1
Total	732

[a]The evaluations in the *IARC Monographs* are qualitative, reflecting the strength of the evidence as to the carcinogenicity of the agent concerned derived from studies in humans and in experimental animals and from other relevant data. This means that the assessment is of the strength of evidence as to whether a substance is carcinogenic, not as to the degree, if any, to which it is carcinogenic.

[b]Criteria for selection of agents evaluated in the *IARC Monographs* is based mainly on a suspicion of carcinogenicity as published in the scientific literature.

IARC Monographs Compound Database

A computerized database has been created for all chemicals, groups of chemicals, complex mixtures, or exposures evaluated in volumes 1–53 of the *IARC Monographs*. For each agent, the database contains the Chemical Abstracts Service Registry Number (CAS number), the chemical name, monographs reference (volume, year, page) and supplement 7 reference (page). In addition, the following data are presented in indexed format and can be searched using boolean logic: *a*) overall evaluation of carcinogenicity to humans; *b*) degree of evidence of carcinogenicity in humans and in animals (according to the criteria given in the Preamble to the *IARC Monographs*); *c*) chemical class—main molecular structural group(s); *d*) use class—main industrial use or exposure circumstance; *e*) for each report/experiment considered in the evaluation, the species, route of exposure, target organ, certainty (degree of evidence evaluation for the organ concerned), tumor type and bibliographical citation; *f*) short-term test and related data: phylogenetic level, genetic and related end point, result; and *g*) teratogenicity, embryotoxicity, and fetotoxicity data.

IARC/EPA Genetic Activity Profiles

The data used in the *IARC Monographs* on genetic and related effects have been prepared together with the U.S. Environmental Protection Agency (EPA) in the form of genetic activity profiles (*8,9*) and stored in a separate database. The test results have been verified by the *IARC Monographs* working groups. The complete data record for each chemical agent includes the name and CAS number, test code, end point, test results with activation *in vitro*, highest ineffective dose or lowest effective dose, and reference. The graphic display of data was modified from the method of Garrett et al. (*10*) for IARC purposes. By the end of 1990, the IARC/EPA data set contained data on 299 agents considered in supplement 6 to the *IARC Monographs* and volumes 46–50 of the series.

Table 2. Established human carcinogens and their target organs.

Carcinogen	Target organ (suspected target organ) in humans
Agents	
Aflatoxins	Liver (lung)
4-Aminobiphenyl	Bladder
Arsenic and arsenic compounds[a]	Lung, skin
Asbestos	Lung, pleura, peritoneum (gastrointestinal tract, larynx)
Azathioprine	Lymphatic system, mesenchyma, hepatobiliary system, skin
Benzene	Hematopoietic system
Benzidine	Bladder
N,N-bis(2-chloroethyl)-2-naphthylamine (chlornaphazine)	Bladder
Bis(chloromethyl)ether and chloromethyl methyl ether (technical grade)	Lung
1,4-Butanediol dimethanesulfonate (myleran)	Hematopoietic system
Chlorambucil	Hematopoietic system
1-(2-Chloroethyl)-3-(4-methylcyclohexyl)-1-nitrosourea (methyl-CCNU)	Hematopoietic
Chromium(VI) compounds	Lung (nasal cavity)
Cyclosporin	Lymphatic system
Cyclophosphamide	Bladder, hematopoietic system
Diethylstilbestrol	Cervix/vagina, breast, testis (uterus)
Erionite	Pleura, peritoneum
Melphalan	Hematopoetic system
8-Methoxypsoralen (methoxsalen) plus ultraviolet radiation	Skin
MOPP and other combined chemotherapy including alkylating agents	Hematopoietic system
Mustard gas (sulfur mustard)	Pharynx, larynx, lung
2-Naphthylamine	Bladder (liver)
Nickel compounds	Nasal cavity, lung
Estrogen replacement therapy	Uterus (breast)
Estrogens, nonsteroidal[a]	Cervix/vagina, breast, testis (uterus)
Estrogens, steroidal[a]	Uterus (breast)
Oral contraceptives, combined[b]	Liver
Oral contraceptives, sequential	Uterus
Radon and its decay products	Lung
Talc containing asbestiform fibres	Lung
Thiotepa	Hematopoietic system
Treosulphan	Hematopoietic system
Vinyl chloride	Liver, blood vessels (brain, lung, lymphatic system)
Mixtures	
Alcoholic beverages	Pharynx, esophagus, liver, larynx, oral cavity (breast)
Analgesic mixture containing phenacetin	Bladder, kidney
Betel quid with tobacco	Oral cavity, pharynx, larynx, esophagus
Coal-tar pitches	Skin, lung, bladder (larynx, oral cavity)
Coal tars	Skin, lung (bladder)
Mineral oils, untreated and mildly treated	Skin (lung, bladder, gastrointestinal tract)
Shale oils	Skin (gastrointestinal tract)
Soots	Skin, lung
Tobacco products, smokeless	Oral cavity, pharynx, esophagus
Tobacco smoke	Lung, bladder, oral cavity, pharnyx, larynx, esophagus, pancreas, kidney
Exposure circumstances	
Aluminum production	Lung, bladder (lymphatic system)
Auramine manufacture	Bladder (prostate)
Boot and shoe manufacture and repair	Nasal cavity, hematopoietic system (pharnyx, lung, liver, gastrointestinal tract, bladder)
Coal gasification	Skin, lung, bladder
Coke production	Skin, lung, kidney
Furniture and cabinet making	Nasal cavity
Iron and steel founding	Lung (gastrointestinal tract, genitourinary system, hematopoietic system)
Isopropyl alcohol manufacture (strong-acid process)	Nasal cavity (larynx)
Magenta manufacture	Bladder
Painter (occupational exposure as a)	Lung
Rubber industry	Bladder, hematopoietic system (lung, gastrointestinal tract, skin, lymphatic system)
Underground hematite mining with exposure to radon	Lung

[a]This evaluation applies to the group of chemicals as a whole and not necessarily to all individual chemicals within the group.
[b]There is also conclusive evidence that these agents have a protective effect against cancers of the ovary and endometrium.

Cross Index of Synonyms and Trade Names

In 1982, a *Cross Index of Synonyms and Trade Names* was published as supplement 3 to the *IARC Monographs* (*6*) to provide an easy reference source for the chemicals considered in volumes 1–26 of the series. The cross index was subsequently updated to include volumes 1–36 (supplement 5) and volumes 1–46 (supplement 8). A computerized database of this cross index has been created, containing CAS numbers and names, synonyms and trade names for all chemicals or complex mixtures evaluated in the *IARC Monographs*.

The common name of the chemical or complex mixture, as used in the *IARC Monographs*, is given with a citation of the monographs reference in which the substance was first evaluated, as well as cross-references to supplement 6 (summary tables, graphics, and references on genetic and related effects) and to supplement 7 (updated summary evaluations of carcinogenicity to humans, with references). All synonyms and trade names then refer to the specific common name. The cross index is updated at approximately 4-year intervals.

Directory of On-Going Research in Cancer Epidemiology

The *Directory of On-Going Research* is an annual compilation of brief, structured abstracts of current research projects in the field of cancer epidemiology obtained by a survey of several thousand investigators each year. The 1991 edition (*11*), the 15th in the series, contains descriptions of about 1150 studies being carried out in more than 80 countries. Eight separate indexes provide access to the projects. In most indexes, studies indexed to a given entry are subclassified by cancer site. The full mailing address (and, where available, telephone, telex, and telefax numbers) of the 900 or so principal investigators are provided to facilitate contacts between research workers.

The *Directory of On-Going Research* contains the names and addresses of the directors of more than 20 banks of biological materials capable of exploitation in epidemiological research, together with details of the nature and size of the collection of biological materials held and the time period over which they were collected. It also provides the names and addresses of the directors of some 240 population-based cancer registries ready to collaborate in epidemiological research.

The printed volume is currently distributed to more than 2000 scientists, libraries, research units, and government agencies. It is widely used as a source of contacts for collaborative research projects, scientific meetings, and advice from colleagues with similar interests.

Directory of Agents Being Tested for Carcinogenicity

Due to the long duration and high costs involved in testing chemicals for carcinogenicity in animals, an international questionnaire survey of institutes undertaking such long-term carcinogenicity testing was initiated in 1973. The objectives were to avoid unnecessary duplication of research, to increase communication among scientists, and to make available a census of both the research facilities involved and the chemicals being tested.

Survey results are arranged alphabetically by country, city, and institute. For every institute, the chemicals or complex mixtures are listed in alphabetical order. The data are reported in a six-column format: *a*) name (with CAS number and name, synonyms and trade names); *b*) use category; *c*) species, strain, and number of animals in treated and control groups; *d*) purity of substance being tested, exposure route, and dose levels; *e*) starting date and stage of experiment; and *f*) principal investigator(s).

The following indexes can be used to identify experiments: epidemiological studies index (cross-reference to human cancer studies of the same chemical listed in the *Directory of On-Going Research in Cancer Epidemiology*); published studies index; institute index; Chemical Abstracts Service Registry Numbers index; cross index of names (synonyms and trade names) (now includes all studies that were published, unpublished, discontinued, etc.); and categories of use index.

The *Directory of Agents Being Tested for Carcinogenicity* is updated at approximately 2-year intervals. Fourteen volumes have been published to date. *Directory of Agents* no. 14 (*12*) gives information on 922 chemicals or agents being tested for carcinogenicity from 80 institutes in 20 countries; a total of 298 published reports on 242 chemicals or agents are listed.

Electronic Publication: The Way Forward

The combined utility of the information in the IARC databases described here may be even wider than appears from a consideration of each database alone. The whole may be greater than the sum of the parts.

The optimal exploitation of this large body of information, however, would require creation of a single electronic database containing all the individual databases. It should now be possible to integrate all this material into a single, comprehensive and readily accessible set of authoritative information on published experimental and human studies, with expert evaluations of carcinogenicity to humans (the *IARC Monographs* and ancillary databases), together with up-to-date listings of all current research in animals (*Directory of Agents Being Tested for Carcinogenicity*) and in humans (*Directory of On-Going Research in Cancer Epidemiology*). Such an information resource could, for example, be used to streamline the review and evaluation of current knowledge and to help identify the most effective research strategies to extend this knowledge.

Many other uses could be described for such an integrated data set, but the ability to disseminate the entire IARC database of cancer research information to a much wider audience than is currently possible, particularly in developing countries, could be a major additional advantage of managing the data in this way. It could also permit regular updating and correction.

This possibility is under active consideration. It will require development of the data in two broad domains. First, the existng databases will need to become structurally aligned: although they are conceptually coherent, they have been developed in different ways for conventional publication in print and are structurally diverse. Second, this conceptual coherence will need to be reflected in a search and retrieval program that enables the full range of information to be exploited without the need for computing skills. Finally, it will be necessary to make this material available in a widely accessible format at an affordable cost.

Initial Experiments

Initial steps in partial electronic publication of two of the databases have already been made: the *Directory of On-Going Research in Cancer Epidemiology* and the genetic activity profiles of compounds evaluated in the *IARC Monographs* program. Seven of the eight indexes of the 1989/90 and 1991 directories were published on a single IBM-compatible diskette and distributed with every copy of the printed *Directory of On-Going Research* (*11,13*). The search software (PROSE, for project search) was developed specially for the *Directory of On-Going Research* to provide simultaneous searching of several indexes.

This facilitates identification of projects meeting a precise description, such as case–control studies of bladder cancer and sweetners being carried out in North America or Europe with information on tobacco use, a search involving four indexes. It would be inefficient to search for such studies using the various indexes in the *Directory of On-Going Research*. Although this initial experiment in providing electronic access to an IARC database is considered successful, there are several limitations.

First, it was necessary to restrict the data to a single diskettes for both financial and practical reasons. It would be tedious to have to load a large number of diskettes into a computer to install the software and inconvenient to dedicate several megabytes of hard disk space to a database likely to be consulted only occasionally. Further, to make the software accessible to the widest possible audience, the oldest type of diskette (360 kilobytes, double density) was used. Even after compression of the files to fit on a single diskette, it was only possible to include the indexes. A search thus produces the serial numbers of the project abstracts in the printed directory, which is then consulted directly. Despite these limitations, electronic access to the *Directory of On-Going Research* indexes has elicited a positive reaction from users.

The results of short-term tests for all chemicals evaluated in the *IARC Monographs* programs have also been published in electronic form, in collaboration with the EPA. These tests include DNA damage, gene mutation, sister chromatid exchange and chromosomal aberrations. The biological systems examined range from bacteria to various mammalian species, including humans, both *in vitro* and *in vivo*. Results are presented in both numerical and graphical form. The electronic version is published as three separate diskettes for use on IBM-compatible personal computers.

Electronic Options

There are several ways in which the information described here could be made available in electronic form. These include diskettes, on-line systems, and CDROM (compact disk, read-only memory).

Use of diskettes to distribute the indexes, as for the *Directory of On-Going Research*, enables more powerful searching than is possible with print publication, but it has several drawbacks. To examine the result of a search, the user must have all the printed texts available and cannot review or refine the results of a search before examining the original source documents. In addition, a large number of diskettes would be required, which complicates the task of distribution, computer storage and regular updating.

Making the databases available on existing on-line information retrieval systems such as MEDLINE (U.S. National Library of Medicine) or EUROCODE (European Organization for Research and Treatment of Cancer) or creating a new on-line host system is another possibility. It would solve the problems of database size and of updating the database and the search software because all these aspects would be managed at the on-line source, requiring no action by the user. From the viewpoint of the IARC, however, this option has three drawbacks. First, and most important, on-line systems are typically available only in developed countries with the appropriate telecommunications infrastructure. Second, on-line access is expensive, as the cost relates to time used on the host mainframe computer and on the telecommunications system used to establish contact with it, rather than to the information supplied. Third, access is sometimes difficult because of problems in telecommunications, such as long waiting times during peak periods and incompatibility of various communications interfaces.

The preferred solution would appear to be CDROM: this is the least expensive mass data storage medium developed so far. A 12-cm diameter disk can hold 650 megabytes of data, which is the equivalent of 250 books or 1500 floppy diskettes. CDROM readers are inexpensive and widely available and can be attached to most personal computers by a standard connection. Software for rapid search of large text databases on CDROM is also available. CDROM disks conform to a widely accepted international standard (ISO-9660 or the High Sierra Standard) for information exchange, which means that problems of compatibility between disks and CDROM readers are unlikely. Finally, provision of all the text, indexes, and software on a single CDROM, for use with personal computer technology, which is widely available even in developing countries, represents an attractive way of making the information available to the widest possible audience. It is likely that purchase of the CDROM and hardware needed to use it would cost less than the purchase and despatch of all the books involved (some of which are out of print).

REFERENCES

1. Commission to issue 4000 page document for use in chemical inventory reporting. Chem. Reg. Rep. 5: 385 (1981).
2. Shulze, H., and Mücke, W. Priority setting for industrial chemicals of health and environmental relevance. Chemosphere 15: 771–780 (1986).
3. Helmes, C. T., Sigman, C. C., and Papa, P. A. Predicting carcinogenicity from structure. Chem. Technol. 15: 48–53 (1985).
4. Fishbein, L. Critical elements in priority selections and ranking systems for risk assessments of chemicals. In: Advances in Modern Environmental Toxicology: Safety Evaluation: Toxicology, Methods, Concepts and Risk Assessment, Vol. X (M. A. Mehlman, Ed.), Princeton Scientific Publications, Princeton, NJ, 1987, pp. 1–50.
5. Blair, E. H., and Bowman, C. M. Control of existing chemicals. Am. Chem. Soc. Ser. 213: 67–79 (1983).
6. IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Volumes 1–50, Supplements 1–8. International Agency for Research on Cancer, Lyon, 1972–1990.
7. Wilbourn, J. D., Haroun, L., Heseltine, E., Kaldor, J., Partensky, C., and Vainio, H. Response of experimental animals to human carcinogens: an analysis based upon the *IARC Monographs* programme. Carcinogenesis 7: 1853–1863 (1986).
8. Waters, M. D., Stack, H. F., Brady, A. L., Lohman, P. H. M., Haroun, L., and Vainio, H. Activity profiles for genetic and related effects. In: IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Genetic and Related Effects: An Updating of Selected *IARC Monographs* from Volumes 1 to 42, Supplement 6. International Agency for Research on Cancer, Lyon, 1987, pp. 687–696.
9. Waters, M. D., Stack, H. F., Brady, A. L., Lohman, P. H. M., Haroun, L., and Vainio, H. Use of computerized data listings and activity profiles of genetic and related effects in the review of 195 compounds. Mutat. Res. 205: 295–312 (1988).
10. Garrett, N. E., Stack, H. F., Gross, M. R., and Waters, M. D. An analysis of the spectra of genetic activity produced by known or suspected human carcinogens. Mutat. Res. 134: 89–111 (1984).
11. Coleman, M., and Wahrendorf, J., Eds. Directory of On-Going Research in Cancer Epidemiology 1991. IARC Scientific Publications 110. International Agency for Research on Cancer, Lyon, 1991.
12. IARC. Directory of Agents Being Tested for Carcinogenicity, No. 14. International Agency for Research on Cancer, Lyon, 1990.
13. Coleman, M., and Wahrendorf, J., Eds. Directory of On-Going Research in Cancer Epidemiology 1989/90. IARC Scientific Publications 101. International Agency for Research on Cancer, Lyon, 1989.