4
Methods for Detecting and Identifying Carcinogens

Contents

LIST OF TABLES

LIST OF FIGURES

4. Methods for Detecting and Identifying Carcinogens

Beginning with this chapter, the focus of this report shifts from all causes of cancer to only chemicals. This shift does not represent a decision that chemicals, in the workplace or in the general environment, are more important in cancer causation than dietary elements, personal habits, radiation, or certain aspects of human biology. However, it does reflect the major legislative and regulatory emphasis recently placed on chemicals, and the greater ease with which chemical carcinogens can be detected by present-day methods.

Different methods are available, employing different techniques, using different test organisms, and producing different types of information about carcinogenicity. This chapter discusses a variety of those methods, their strengths and weaknesses, some results from each, and the tools they require.

METHODS

There are four major methods for detecting and identifying carcinogens:

1. molecular structure analysis,
2. short-term tests,
3. long-term chronic bioassays in laboratory animals (termed "bioassays" or "animal tests" hereafter), and
4. epidemiology.

The first two methods produce information about potential carcinogenicity; the third provides direct evidence of carcinogenicity in animals; the fourth produces direct evidence about cancer in man. These categories are briefly described in table 23.

Probably no statement made in the last column of table 23 is free from dispute. Results may be, and frequently are, challenged for several reasons: because the test was incorrectly designed or executed (all methods); because the method does not directly measure carcinogenicity (methods 1 and 2); because the test is too sensitive and produces false positives (methods 2 and 3); because the test is too insensitive and produces false negatives (method 4); and because the test does not measure human experience (all methods but 4), etc.

Knowledge about tests and about the validity of test results increases as the tests are more often used, more discussed, and more refined. The state of scientific knowledge plays an important role in decisions to test or not to test a chemical, decisions about which tests are appropriate, and decisions about interpretation of the test results. As will be discussed, "policy statemerits, " sometimes issued as guidelines or standards, detail the methods that an organization will use in making decisions. A certain tension is apparent in all the policies. Tests cost money and take time; bigger and better tests cost more and take more time; compromises are necessary in the design of each test so that a reasonable number of chemicals can be tested.

An equally important issue is the amount of information necessary to decide that a chemical is or is not a carcinogen that requires some control action. The fact that some regulations are based on nonhuman test systems shows that proof that a chemical is a human carcinogen is not demanded. This illustrates that prevention of cancer is seen as so important that it is appropriate to make decisions to restrict exposures before human damage is observed.

Table 23.—General Classification of Tests Available To Determine Properties Related to Carcinogenicity

Method	System	Time required	Basis for test	Result	Conclusion, if result is positive
Molecular structure analysis	"Paper chemistry"	Days	Chemicals with like structures interact similarly with DNA	Structure resembles (positive) or does not resemble (negative) structure of known carcinogen	Chemical may be hazardous. That determination requires further testing.
	Basic laboratory tests	Weeks			
Short-term tests	Bacteria, yeast, cultured cells, intact animals	Generally few weeks (range 1 day to 8 months)	Chemical interaction with DNA can be measured in biological systems	Chemical causes (positive) or does not cause (negative) a response known to be caused by carcinogens	Chemical is a potential carcinogen.
Bioassay	Intact animals (rats, mice)	2 to 5 years	Chemicals that cause tumors in animals may cause tumors in humans	Chemical causes (positive) or does not cause (negative) increased incidence of tumors	Chemical is recognized as a carcinogen in that species and as a potential human carcinogen.
Epidemiology	Humans	Months to lifetimes	Chemicals that cause cancer can be detected in studies of human populations	Chemical is associated (positive) or is not associated (negative) with an increased incidence of cancer	Chemical is recognized as a human carcinogen.

SOURCE Office of Technology Assessment

ANALYSIS OF MOLECULAR STRUCTURE AND OTHER PHYSICAL CONSTANTS

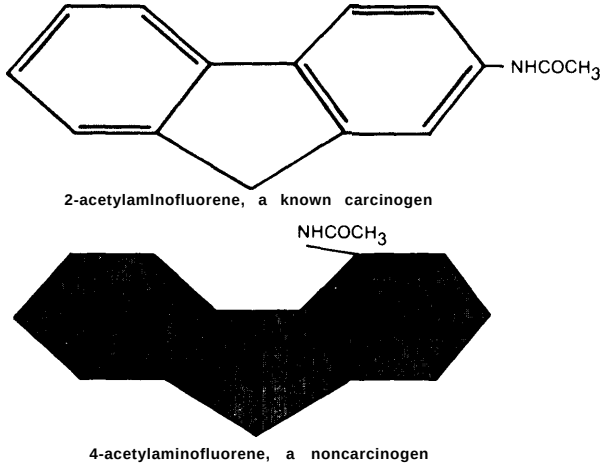
Some information about the likelihood of a chemical being a carcinogen maybe obtained by comparing its structure and chemical and physical characteristics with those of known carcinogens and noncarcinogens. This first stage in an orderly determination of whether or not a chemical is a carcinogen requires the gathering of all available information about it. The information can be obtained from sources as diverse as results from testing a chemical in animals to anecdotal stories about human disease, but the most readily available information is often about the molecular structure and physical and chemical properties of the suspect chemical.

Certain molecular structures have been associated with carcinogenicity, and structural similarity is used in making decisions about which agents are more or less suspect. For instance, 8 of the first 14 carcinogens regulated by the Occupational Safety and Health Administration (OSHA) are aromatic amines. The Environmental Protection Agency (EPA) relies heavily on structural analysis in determining whether or not "new" chemicals, described in premanufacturing notices, may present an unreasonable risk to health or the environment. Chemical and physical properties which are useful in evaluating a chemical's carcinogenic potential include volubility, stability, sensitivity to pH, and chemical reactivity. Often this type of information is generated by the manufacturer of the chemical.

A number of proposals have been made that chemicals be divided up into classes depending on their structural similarities and that testing be done on a number of members in each **class.** Unfortunately, carcinogens are known in several chemical classes, and " . . . the dozen or more known classes of these agents [carcinogens] share no common structural features" (239). **Furthermore,** even within classes, closely related chemicals may differ with respect to carcinogenicity—e.g., 2-acetylaminofluorene (2-AAF) is a well-documented carcinogen; its

chemical relative, 4-acetylaminofluorene (4-AAF), is not a carcinogen (see figure 19).

Figure 19.— Molecular Structures of Two Closely Related Chemicals: One a Carcinogen and One a Noncarcinogen



SOURCE. Off Ice of Technology Assessment

EPA recently based a regulatory decision on molecular structural analysis. Under section 5(e) of the Toxic Substances Control Act (TSCA), EPA prohibited the entry of six new chemical substances into the marketplace unless the manufacturer provided additional information about toxicity. A National Cancer Institute (NCI) bioassay had shown a related chemical to be carcinogenic, and based on that result, EPA decided that more information was needed before manufacture could begin (88). The manufacturer decided not to proceed with the testing and did not market the chemicals.

SHORT-TERM TESTS

Short-term tests are so named because of the relatively short time needed to conduct the experiments. Some studies involving micro-organisms require less than 1 day to complete (87), most require a few days to a few weeks, and the longest, using mice, requires 8 to 9 months (172), These times may be compared to the more than 3 years required to complete a bioassay and the months to years required to complete epidemiologic studies.

A number of reasons account for the growing interest in using short-term tests to predict a chemical's carcinogenic potential:

- shorter time period required for the tests;
- low cost ($100 to a few thousand dollars for each test compared to $400,000 to $1 million for a bioassay);
- evidence that the majority of chemical carcinogens are mutagens and that many mutagens are carcinogens;
- growing opinion that short-term tests can predict which chemicals may be carcinogens.

The third point is important because many short-term tests determine whether or not a chemical causes mutations (mutagenicity) rather than if it causes cancer (carcinogenicity). The postulated relationship between mutagenicity **and carcinogenicity stems from biological properties common to all living organisms.** The genetic information in both germ cells (egg and sperm) and somatic cells (nongerm or "body cells") is composed of deoxyribonucleic acid (DNA), and agents that cause mutations in germ cells are also expected to cause mutations in somatic cells. A germ cell mutation may prevent

the formation of viable offspring, cause a genetic malformation, or produce subtle defects in the progeny, such as minimal depression in intelligence or increased susceptibility to disease.

The consequences of somatic cell mutations are quite different from those in germ cells. Somatic cells do not contribute genetic information to succeeding generations, but as each somatic cell grows and divides, copies of its DNA are passed on to its two "daughter" cells. Some somatic cell mutations result in uncontrolled cellular growth: The normal tightly controlled growth pattern of the somatic cell is broken down, the cell grows and divides more quickly than it should, progeny cells exhibit the same uncontrolled growth, and cancer results.

The hypothesis that assigns genetic changes in somatic cells a role in cancer initiation is referred to as "the somatic mutation theory of cancer." Cairns (42) discusses the origin and development of this theory, which provides intellectual support for associating mutagenicity and related changes measured in short-term tests with potential carcinogenicity.

The short-term tests that depend on mutagenicity can detect only materials that interact with DNA. Some cancers may be caused by other, "epigenetic," pathways that may not involve alterations in genetic information (317). Short-term tests cannot detect such activities. Additionally, short-term tests do not detect promoters that do not interact with DNA. The generally good correlation between mutagenic and carcinogenic activity as well as the bulk of results from basic cancer biological research support the notion that carcinogens generally interact with DNA.

The Ames Test

The most widely used and best-studied short-term test, the "Ames test, " is named for its developer, Bruce Ames, a molecular biologist. The test measures the capacity of a chemical to cause mutations in the bacterium *Salmonella typhimurium*, a favorite tool for laboratory investigations since the 1940's. Salmonella's genetics and biochemistry are well understood: it is quickly and easily grown; it presents few manipulative problems in the laboratory, and test results are easily interpretable and reproducible between laboratories.

Basically, the Ames test involves mixing the chemical under test with a bacterial culture and then manipulating the culture so that only mutated bacteria will grow. The number of mutated bacteria is a measure of the potency of the tested material as a mutagen.

It is well known that some chemicals must be altered before they interact with DNA and that in humans and other mammals these changes are often accomplished by enzymes in the liver. The addition of liver extracts to the Ames test system and to other short-term tests provides a mechanism for these metabolic activation changes to be accomplished. Generally, extracts are prepared from rats, hamsters, or other laboratory animals. The source of the extracts and the amount used in the tests affect results, and careful experiments report these specifics so that others can replicate the tests. Some chemicals are "activated" by bacteria normally present in the intestine rather than by the liver. The addition of extracts of such bacteria to Ames test mixtures has been shown to activate some chemicals to mutagenic forms (338).

As of early 1979, more than 2,600 Ames test results had been published (172). The interested reader is referred to Hollstein et al. (172) and Devoret (87) for more detailed descriptions of the tests, to Ames (11) for a description of the problems of carcinogen identification addressed by short-term tests, to a series of papers in the April 1979 issue of the *Journal of the National Cancer Institute* and to Bartsch et al. (22) about experiments to validate the reliability of short-term tests.

Short-term tests are still in their infancy; development of the Ames test began about 15 years ago (11). The major factor influencing the acceptance or rejection of any short-term test as a method for identifying carcinogens is a demonstration that the test can discriminate between carcinogens and noncarcinogens.

The crux of validation experiments is determining: 1) how frequently carcinogens are correctly identified by short-term tests (sensitivity)

and 2) how frequently noncarcinogens are correctly identified (specificity). Ideally the frequency for both sensitivity and specificity would be 100 percent. If the Ames test worked perfectly, every tested carcinogen would be a mutagen; every tested noncarcinogen would be a nonmutagen in the test.

The difference between the ideal and the measured performance can be expressed in terms of sensitivity. If the test identified 90 of 100 carcinogens as mutagens, it would have a sensitivity of 90 percent. The same observation can be described in terms of its false-negative rate. In the example, 10 carcinogens were falsely negative in the mutagenicity test, so it had a 10-percent false-negative rate.

Similarly for noncarcinogens, the test's success can be expressed as a specificity rate. If it identified 90 of 100 noncarcinogens as nonmutagens, its specificity was 90 percent. Alternatively, the result can be expressed in terms of the false-positive rate, which is 10 percent in the example.

Ames and his associates tested agents that had been classified as carcinogens or noncarcinogens in bioassays. They found that 156 of 174 animal carcinogens (90 percent) were mutagenic, and, equally important, 96 of 108 (88 percent) chemicals classified as noncarcinogens were not mutagenic **(227).** Ames has suggested that some of the "noncarcinogenic" chemicals that were detected as mutagens might have been incorrectly classified as noncarcinogens on the basis of bioassay results (11,227). His suggestion points up a problem inherent in "validating" any test against the results of other tests: There is no guarantee that the results of the tests that are used as standards are completely accurate.

Other researchers have investigated the correlation between Ames test mutagenicity and animal carcinogenicity in efforts to validate the mutagenicity test for predicting carcinogenicity. The good correlation between mutagenicity and carcinogenicity found by McCann et al. (227) was "confirmed in a smaller . . . study by the Imperial Chemical Industries" (34). (In addition, see for instance **13,22,226,309,335).** There is general agreement that the tests are predictive, but some disagreement about whether they are **70-, 80-,** or 90-percent sensitive. A number of factors contribute to the observed differences in sensitivity. For instance, better correlations may reflect testing chemical classes on which the Ames test performs well. Ames has shown that his test does not work well with certain classes of chemicals, e.g., halogenated hydrocarbons and metals, and including those in validation tests decreases the sensitivity of the tests.

Bartsch et al. (22) report on 89 chemicals studied in the Ames test for mutagenicity. The 89 were chosen because sufficient data existed to classify each of them as a carcinogen or noncarcinogen in animal tests. Results from the Bartsch et al. (22) study along with those from an earlier study by McCann et al. (227) are shown in table 24.

It can be seen that **76** percent of the tested carcinogens were mutagenic in the Bartsch experiments as compared to the 90 percent that were reported mutagenic in McCann et al. (227). The sensitivity in th, former study was lower, but comparable to the other report, The comparison of specificity is somewhat deceiving. The 57-percent specificity recorded by Bartsch et al. (22) is much lower than the 88 percent from the earlier study, but results from only seven noncarcinogens were reported by Bartsch et al. (22). McCann et al. **(237)** tested 108 noncarcinogens. The 57 percent is subject to much larger error than the higher estimate of specificity.

In both reports, the predictive value was found to be over 90 percent. The predictive value is calculated by comparing the number of carcinogens identified as mutagens to the total number of both carcinogens and noncarcinogens that were mutagenic. This means that more than 90 percent of the substances detected as mutagens *were* carcinogens.

An important qualifier must be applied to the predictive value of a test. It depends strongIy not only on sensitivity and specificity but also on the proportion of carcinogens in the collection of substances tested for mutagenicity. The proportion of carcinogens in both validation experiments shown in table 24 was well above 50 percent.

Table 24.—Results Obtained in Two Validation Studies of the Ames Test When Known Carcinogens and Noncarcinogens Were Tested

	Calculation[a]	Results from: Bartsch et al. (22)	McCann et al. (227)
Sensitivity	$\frac{C^+ M^+}{C^+ M^+ + C^+ M^-}$	760/o (62/82)	900/0 (156/174)
Specificity	$\frac{C^- M^-}{C^- M^+ + C^- M^-}$	570/0 (4/7)	880/0 (96/108)
Predictive value	$\frac{C^+ M^+}{C^+ M^+ + C^- M^+}$	950/0 (62/65)	920/o (156/168)
Proportion of carcinogens	$\frac{C^+ M^+ + C^+ M^-}{\text{all chemicals tested}}$	920/o (82/89)	620/o (174/282)

[a] C^+ chemicals known to be carcinogens; C^- chemicals known to be noncarcinogens
M^+ chemicals identified as mutagens; M^- chemicals identified as nonmutagens
$C^+ M^+$ carcinogens "correctly" identified as mutagens
$C^- M^-$ noncarcinogens "correctly" identified as nonmutagens
$C^- M^+$ noncarcinogens "incorrectly" identified as mutagens
$C^+ M^-$ carcinogens "incorrectly" identified as nonmutagens

SOURCE: Office of Technology Assessment, adapted from Bartsch et al. (22)

Table 25 shows the expected results from examining two hypothetical collections of chemicals. The first collection of 1,000 chemicals contains 10 carcinogens (1 percent); the second contains 100 carcinogens in 1,000 total chemicals (10 percent). The short-term test in both cases is assumed to be 90-percent sensitive and 90-percent specific. The predictive value in the two collections differs more than sixfold because of the higher contribution of false positives to the total positives in the l-percent carcinogen collection. This example illustrates the important role played in predictive value computations by the percentage of carcinogens included in validation experiments.

McMahon, Cline, and Thompson (233) developed a modification of the Ames test and used it to assay 855 chemicals. To validate their own test system, the authors included 125 chemicals that had been tested previously in the Ames test. They reported "excellent agreement" between results in their tests and those reported by McCann et al. (227), Among the other chemicals tested by McMahon, Cline, and Thompson (233) the 299 chosen from manufacturing or laboratory synthesis provided the largest percentage of mutagens; 60 of 299 (20 percent) were mutagenic. In contrast, chemicals developed as potential agricultural or pharmaceutical products were less often mutagenic; 29 of 361 such compounds (6 percent) were positive. The authors state (233):

> Very few of the chemical mutagens detected in this study had chemical structures uniquely different from known carcinogens. Further study in other test systems will be required to assess the significance of results with the few unique compounds encountered. The results of the study do suggest, however, that as testing continues on more and more compounds it will be found that most of the new mutagenic compounds detected will be related to known carcinogens and mutagens and that new unique chemical structures possessing these properties will be found rarely.

The McMahon, Cline, and Thompson paper (233) illustrates that large numbers of chemicals can be tested quickly. Furthermore, their results, which are in "excellent agreement" with those earlier reported by McCann et al. (227) for chemicals tested in both studies, illustrate that the test is reproducible in different laboratories. Whether the prediction that most

Table 25.—Expected Results of Examining Two Collections of Chemicals for Mutagenicity Using a Short-Term Test That is 90-Percent Sensitive and 90-Percent Specific: One Collection of Chemicals Contains 1 Percent Carcinogens; the Other Contains 10 Percent

	Calculation[a]	Collections of chemicals with 1 o/o carcinogens	Collections of chemicals with 100/0 carcinogens
Proportion of carcinogens in sample of 1,000 chemicals	$\frac{C^+ M^+ + C^+ M^-}{\text{all chemicals tested}}$	1o/o (i.e., 10 carcinogens)	100/0 (i.e., 100 carcinogens)
Carcinogens identified as mutagens	$\frac{C^+ M^+}{C^+ M^+ + C^+ M^-}$	90°/0 (i.e., 9 of the 10 carcinogens)	90% (i.e., 90 of the 100 carcinogens)
Carcinogens identified as nonmutagens—false negatives in the test	$\frac{C^+ M^-}{C^+ M^+ + C^+ M^-}$	10°/0 (i.e., 1 of the 10 carcinogens)	100/0 (i.e., 10 of the 100 carcinogens)
Noncarcinogens identified as non mutagens	$\frac{C^- M^-}{C^- M^- + C^- M^+}$	90% (i.e., 891 of 990 non carcinogens)	90% (i.e., 810 of 900 noncarcinogens)
Noncarcinogens identified as mutagens—false positives in the test	$\frac{C^- M^+}{C^- M^- + C^- M^+}$	100/0 (i.e., 99 of 990 noncarcinogens)	100/0 (i.e., 90 of 900 noncarcinogens)
Summary:			
Carcinogens identified as mutagens		9	90
Noncarcinogens identified as mutagens		99	90
		108	180
Predictive value— carcinogens identified as mutagens/total carcinogens plus noncarcinogens identified as mutagens	$\frac{C^+ M^+}{C^+ M^+ + C^- M^+}$	8.3% (i.e., 9/108)	50°/0 (i.e., 90/180)

[a] C^+ chemicals known to be carcinogens; C^- chemicals known to be noncarcinogens
M^+ Chemicals identified as mutagens, M^- chemicals identified as nonmutagens
$C^+ M^+$ carcinogens "correctly" identified as mutagens
$C^- M^+$ noncarcinogens "Incorrectly" identified as mutagens
$C^+ M^-$ carcinogens "incorrectly" identified as nonmutagens

SOURCE Office of Technology Assessment, adapted from Bartsch et al (22)

mutagens will have structures related to those of known carcinogens awaits further testing.

The International Program for the Evaluation of Short-Term Tests for Carcinogenicity (partially supported by the National Toxicology Program (NTP), the National Institute of Environmental Health Sciences (NIEHS), and EPA) is analyzing the accuracy of about 30 short-term test systems including the Ames test. The program distributed 42 coded carcinogens and noncarcinogens to 66 investigators, and the results of those studies will be published in *1981*. The accuracy of the Ames test in the 12 laboratories which examined it is comparable to the higher accuracy figures (about 80 percent) in the literature (188).

The largest program in genetic toxicology (mutagenicity) is EPA's Genetox Program (357). It is not now engaged in validating short-term tests for predicting carcinogenicity, but it is examining correlations among various short-term tests. Beginning in early 1982, the program expects to publish recommendations for batteries of tests most appropriate for measuring particular mutagenic effects.

How much reliance is to be placed on the results of short-term tests continues to be

discussed. Leon Golberg, former President of the Chemical Industry Institute of Toxicology (CIIT), for instance, compared the results from Ames tests to bioassay results for hair dye components. He found little agreement and cautions against using the Ames test as a substitute for bioassays (142,143). However, a recent paper found good correlation between the mutagenicity and carcinogenicity of phenylenediamine hair dye components **(330).**

In summary, the Ames test is reported to detect known carcinogens as mutagens with a frequency as high as 90 percent. Although a breakthrough in understanding of the correlation between mutagenicity and carcinogenicity is required before more definitive conclusions can be drawn from the Ames test, a positive Ames test result shows that the agent is a mutagen and suggests that it may be a carcinogen.

Other Short-Term Tests

The number of short-term tests has proliferated rapidly. Purchase et al. (301) included only six short-term tests in a 1976 review of the published literature; less than 1 year later, OTA had saccharin tested in 12 short-term tests (282). Two years after that, in the summer of 1979, a review by Hollstein et al. (172) reported that over 100 short-term tests had been described in the scientific literature. The proliferation of tests reflects the great interest in cheaper, faster tests for identifying chemical carcinogens.

Hollstein et al. (172) divided short-term tests into eight classes, according to what they can detect:

1. mutagenesis in bacteria (including Salmonella) and bacterial viruses;
2. mutagenesis in yeast;
3. mutagenesis in cultured (laboratory-grown) mammalian cells;
4. mutagenesis affecting mouse hair color;
5. mutagenesis in fruit flies (Drosophila *melanogaster);*
6. effects on chromosomal mechanics in intact mammals and in mammalian cells in culture;
7. disruption of DNA synthesis and DNA repair mechanisms in bacteria and other organisms; and
8. in vitro transformation of cultured cells.

One of the powerful tools available to biology is the use of cell culture systems, which allows cells obtained from animal or human tissues to be grown and manipulated in the laboratory. Cell cultures can be manipulated to serve as assays for mutagens (#3 above) and for chemicals that interfere with chromosomal mechanics (#6 above), but the most directly applicable use of cultured cells for carcinogen identification involves in vitro transformation (#8 above).

Cells grown in culture exhibit characteristic morphologies and growth patterns. Exposing cultured cells to known tumor-causing viruses or to chemical carcinogens causes changes in morphology and growth characteristics. The changes are collectively called “transformation. ” Transformed cells resemble cells from tumors and have the important property of causing tumors when they are injected into animals, thus demonstrating a direct relationship between transformation and oncogenicity (tumor formation). Transformation of cell cultures is biologically more closely related to oncogenicity than is mutation, and transformation assays may take on major importance in testing programs. The NTP 1979 Annual Plan (271) stated:

> A lifetime bioassay in rodents is the current procedure utilized to determine carcinogenic potential of a chemical. The NTP does not propose alternative methods but acknowledges a need in the longer term, to develop or validate less expensive and more rapid methods that may in some instances supplant the need for lifetime bioassays. Mammalian cell transformations are potential short-term assays that indicate carcinogenic potential of a chemical . . .

And less than a year later, a more optimistic comment appeared in the Department of Health, Education, and Welfare (DHEW) *Health Research Planning* document of December 1979 (130):

> The dimensions of NTP, and the significant demands it places on the funds and personnel of the participating agencies, should diminish by 1985, as the fiscal projections suggest . . . . It is our hope that, by then, better test systems will begin to replace the tedious and costly animal assays now required.

Not everyone is so optimistic as to think a replacement for animal assays will be available in 4 or 5 years. Transformation tests are probably the best bet for the replacement, but they require more development and validation.

Transformation assays are technically more difficult than the Ames test, but not so difficult as to preclude their use on a routine basis. Validation studies are being carried out on a number of in vitro transformation systems to determine how accurately they identify carcinogens and noncarcinogens (271,272).

NTP is conducting additional validation studies of a test that uses whole animals. This test, which has been in limited use for about 30 years, requires about 6 months to complete. Exposure of a particular strain of mice to known carcinogens causes an increase in the frequency of lung tumors (adenomas) and causes earlier appearance of the tumors. The test takes much less time than the standard assay, and NTP (272) has found this test accurately predicts results in bioassays. NTP tested 60 chemicals in this system in 1980 and plans to test another 30 in 1981.

The British publication The *Economist* (97) singled out a transformation assay, using Chinese hamster ovary cells, as having promise for carcinogen identification. Discussion of short-term tests in that publication, which seldom publishes articles about biology, reflects the increasing importance of the tests. A more authoritative source, the NCI National Cancer Advisory Board's Subcommittee on Environmental Carcinogenesis (245) said: " . . . this subcommittee is enthusiastic about the possible future use of in vitro [short-term] tests as part of a screening system for potential carcinogens and believe that their further development and validation deserve high priority."

Use of Short-Term Test Results and Policy Statements About the Tests

How best to utilize short-term tests in carcinogen identification is hotly debated. The majority view is that the tests are most useful as a screen to determine a chemical's potential carcinogenicity. As a new chemical is developed or as an old one comes under suspicion, an inexpensive short-term test or battery of tests can provide information about whether it is or is not likely to be a carcinogenic hazard. If the results of the test are negative, the chemical is considered less likely to be a hazard than a chemical that is positive. In the case of a chemical being commercially developed, a positive result might suggest that the chemical not be produced or that the cost of testing it in a bioassay should be considered in deciding whether or not to produce it. A positive short-term test result on a commercially produced chemical most likely causes more of a problem. The manufacturer is faced with having to begin other tests and to warn his employees and customers of potential hazard.

Opinions differ about the weight to be placed on short-term test results. Peter Hutt, former General Counsel at the Food and Drug Administration (FDA), and now in private law practice says that he advises his clients not to continue the development of a product which is positive in a short-term test. He maintains that, "life is too short" to invest time and effort in a chemical that is more likely than not to be considered a suspect carcinogen. Near the ether end of the spectrum of opinion, Leon Golberg, in reviewing poor correlations between results of Ames testing and bioassays of components of hair dyes, concludes " . . . it is very hard to accept the fact that the Ames test is a predictor of carcinogenic potential" (143).

The OSHA document "Identification, Classification and Regulation of Potential Occupational Carcinogens" (279), accepts the results of short-term tests as supportive evidence for deciding whether a chemical will be classified as a carcinogen or noncarcinogen. TSCA test stand-

80-481 O - 81 - 9

ards (106) and Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) guidelines (102) accept short-term tests as measures for mutagenicity but do not consider them in making decisions about carcinogenicity. However, they mention that test developments are promising.

An important step in making decisions about the use of short-term tests are the ongoing validation studies which compare predictions made from short-term tests to knowledge about carcinogenicity from bioassays or epidemiology. These studies, although limited by the quality and quantity of data about carcinogenicity, are producing valuable information. In addition, studies of molecular mechanisms of mutagenicity and carcinogenicity are important in deciding about the applicability of short-term tests.

The problem of the carcinogens that are not detected (false negatives; lack of sensitivity) and the noncarcinogens that are falsely detected (false positives; lack of specificity) by any one test might be solved with additional short-term tests. The great attractiveness of a battery of short-term tests is that it might correctly identify all carcinogens and noncarcinogens. Unfortunately, no such battery has yet been defined. The composition of the battery will depend on validation studies and acceptance of each component test.

The growing use of short-term tests shows that short-term tests have moved to an important position in toxicology. The speed with which they have been incorporated into Government and private sector programs reflects the importance of the need to which they are addressed.

BIOASSAYS

Chemicals cannot be tested for carcinogenicity in humans because of ethical considerations. A substantial body of experimentally derived knowledge and the preponderance of expert opinion support the conclusion that testing of chemicals in laboratory animals provides reliable information about carcinogenicity. Animal tests employ whole mammal systems, and although they differ one from another, all mammals, including humans, share many biological features (266):

> Effects in animals, properly qualified, are applicable to man. This premise underlies all of experimental biology and medicine, but because it is continually questioned with regard to human cancer, it is desirable to point out that cancer in men and animals is strikingly similar. Virtually every form of human cancer has an experimental counterpart, and every form of multicellular organism is subject to cancer, including insects, fish, and plants. Although there are differences in susceptibility between different animal species, and between individuals of the same strain, carcinogenic chemicals will affect most test species, and there are large bodies of experimental data that indicate that exposures that are carcinogenic to animals are likely to be carcinogenic to man, and vice versa.

In comparison to short-term tests and epidemiology, bioassays have had a longer development period and enjoy greater acceptance than the short-term tests; they are more easily manipulated to produce evidence linking a particular substance to cancer than epidemiology, and they can predict human risks rather than relying on cases of human cancer to demonstrate risk. On the other hand, they take longer and cost much more than short-term tests.

The bioassay's apparent simplicity belies the difficulty of executing such experiments. Briefly, the suspect chemical is administered to a population of laboratory animals. As animals die or are killed during the course of the study, they are examined for the presence of tumors. At the end of the treatment and observation period (generally about 2 years), the surviving animals are killed and examined. A control group of animals is treated exactly the same except that they are not exposed to the suspect substance. The type and number of tumors and

other relevant pathologies present in the exposed animals are compared with those in the control group, and statistically analyzed.

A statistical expression commonly used to describe a positive result is ". . . it has a p value less than 0.05" (5 percent). The p value is the probability that the observed effect might be explained by chance; in this case, the expression means that the probability of the observed carcinogenic effect being due to chance is less than 5 percent. A p value of 0.05 or less is commonly required to decide that a test result was statistically positive.

Finally a conclusion is drawn about whether or not the evidence indicates that the substance caused cancer in the exposed animals. An excellent discussion of experimental design and analysis is available from the International Agency for Research on Cancer (IARC) (187).

The first successful experimental induction of cancer in animals (in 1915) showed that painting rabbits' ears with coal tar produced tumors which morphologically resembled human tumors associated with exposure to the same agent (cited in 342). Most chemicals which are presently known to cause cancer in humans are also carcinogens in animals.

Verification of the predictive power of bioassays would require that the agent be shown to be a human carcinogen. Currently, IARC maintains that convincing evidence for human carcinogenicity is available for only 18 exposures, including 14 chemicals. At the same time, it lists 142 substances for which data are "sufficient" to conclude that they are carcinogenic in animals. It is difficult to demonstrate human carcinogenicity. Once a substance is shown to be an animal carcinogen, regulatory restrictions on and voluntary reductions in the use of the chemical may reduce human exposures, making those demonstrations more difficult. Reductions in exposure in such cases are far more important to human health than the foregone opportunities to verify the predictive powers of bioassays.

Standard Protocols for Bioassays

An important event in bioassay design was the development of NCI's Guidelines for Carcinogenic Bioassay in Small Rodents (33 I). The guidelines describe minimum requirements for the design and conduct of a scientifically valid bioassay and discuss important considerations in undertaking such studies. They are written to provide flexibility in experimental design while setting certain minimal requirements:

1. Each chemical should be tested in at least two species and both sexes. Rats and mice are usually the species of choice.
2. Each bioassay should contain at least 50 animals in each experimental group. When both sexes of two species are used and two treatment levels are administered and a third group is used as controls, a total of 600 animals is needed (see table 26).
3. Exposure to the chemical should start when the animals are 6 weeks old or younger and continue for the greater part of their lifespan. Mice and rats are usually exposed for 24 months.
4. One treatment group should receive the maximum tolerated dose (MTD), which is defined as the highest dose that can be given that would not alter the animals' normal lifespan from effects other than cancer. The other treatment group is treated with a fraction of MTD.
5. The route by which a chemical is administered should be the same or as close as possible to the one by which human exposure occurs.
6. Animals are closely monitored throughout the study for signs of toxic effects and other causes affecting their health.
7. Examination of animals is conducted by or under the direction of a pathologist qualified in laboratory animal pathology.

The guidelines also specify that special procedures (e.g., organ function tests, body burden determinations, absorption and excretion tests) may be needed for evaluating certain chemicals.

Table 26.—Distribution and Number of Animals in a Typical Bioassay Study of Carcinogens

Experimental groups	Species A		Species B	
	Males	Females	Males	Females
Dosage MTD[a] group	50	50	50	50
Dosage MTD/x group	50	50	50	50
Control group	50	50	50	50

[a]Maximum tolerated dose.

SOURCE: National Cancer Institute

A National Research Council report (262) suggested a two-generation bioassay when there is special concern with prenatal and perinatal effects of a substance. In the suggested two-generation experiment, the first generation is exposed from 6 weeks of age through their adult life, including periods of reproductive activity and pregnancy, and the offspring are exposed for their lifespans. The most important two-generation experiments, from the standpoint of public policy, were the three showing saccharin to be a bladder carcinogen in second generation rats (282). The advantage of the two-generation experiment is that it exposes fetuses and very young animals, which may represent the most sensitive stages of life. This procedure costs more and requires more time to complete than the one generation experiment. Probably because of those disadvantages, it has not often been used.

Table 27 compares some specifics of NCI guidelines with those proposed by EPA for testing under FIFRA (102) and TSCA (106). NCI guidelines state that at least two-dose levels be tested; EPA requires three test doses. The highest dose, high-dose level (HDL), for TSCA is defined as being "slightly toxic" and is to be determined in a 90-day-feeding study to precede the oncogenicity experiment. The other two doses are to be fractions of HDL. EPA prefers the term HDL, which is less specifically defined than MTD, because of the controversies which have erupted over defining MTD. Three-dose levels are proposed by EPA because a review of many NCI bioassays revealed that tumor incidence was sometimes higher at a dose of MTD/2 than at MTD because the higher dose killed some animals before tumors might have developed. The third-dose level will also provide additional information about the dose-response curve (102). The text which accompanies the proposed EPA guidelines for carcinogen testing under FIFRA (102) contains additional information about alternative approaches considered and subsequently discarded for the guidelines.

Three to five years are required to complete a bioassay. A subchronic test, to set dose levels, requiring **2 to 6** months, is followed by an average 24 months of exposure to the agent, and sometimes an additional 3 to 6 months observation period. Pathological examination of tissues from the animals and evaluation of the pathological and other data may take an additional year.

The cost of a bioassay has been variously estimated: NCI estimates about **$400,000,** TSCA guidelines (106) estimated $400,000, and EPA (112) later estimated a range from $390,000 to $980,000.

Some changes have been made in protocols since the NCI guidelines were published, but in contrast to the situation in spring 1977, when OTA reviewed carcinogen testing **(282),** apparently much less confusion and contention now exist about what constitutes an adequate bioassay. NCI guidelines are for minimal standards; certainly larger populations of animals can be tested. For very important bioassays, larger numbers of animals or more dosage groups might be specified.

Objections to the Usefulness of Bioassays

Some general aspects of test design are seldom disputed. Examples of such provisions are the requirement of a minimum number of animals in the test groups and that (generally) both sexes

Table 27.—Guidelines for Bioassay in Small Rodents

	NCI (331)[a]	FIFRA (102)	TSCA (106)
Endpoint	Carcinogenicity	Oncogenicity	Oncogenicity
Study plan:			
Animal species	2, rats and mice	2, rats and mice	2, rats and mice
Number of animals at each dose	50 males, 50 females	50 males, 50 females	50 males, 50 females
Dosages	2, MTD MTD/2 or MTD/4 plus no-dose control	3, MTD MTD/2 or MTD/4 MTD/4 or MTD/8 plus no-dose control	3, HDL HDL/2 or HDL/4 HDL/4 or HDL/10 plus no-dose control
Dosing regimen			
Start	At 6 weeks of age	In utero or at 6 weeks	At 6 weeks of age
End	At 24 months of age	Mice, 18-24 mos; rats, 24-30	At 24-30 months of age
Observation period	3-6 months after end of dosing	N.s.	N.s.
Organs and tissues to be examined	All animals: external and histopathologic examination (ea. 30 organs and tissues)	All animals: external examination; some animals: pathologic exam of 30 organs and tissues, other animals, fewer organs and tissues	All animals: external and histopathologic examination (ea. 30 organs and tissues)[b]
Personnel qualifications:			
Study director	N.s.	N.s.	Responsibilities detailed
Pathologist	Board-qualified	N.s.	Board-certified or equivalent
Animal husbandry	N.s.	N.s.	Board-certified vet. or equivalent
Cost estimate	N.s.	N.s.	**$400,000 ± 160,000**

Abbreviations: MTD, Maximum tolerated dose, causes minor acute toxicity
HDL, high dose level, causes some acute toxicity
N s., not specified

a Th, NC I Guidelines specify the indicated minimum requirements, They allow for flexibility in experimental design so long as the minimum requirements are met,
b EPA estimates that the 40,000 microscope slides produced in this examination will require more than 3/4 of a year of a pathologist's time for analysis

SOURCE Off Ice of Technology Assessment

be tested. On the other hand, consensus does not exist about some aspects of experimental design, for instance: How high a dose is to be administered? The policy of the agency that draws up the guidelines is reflected in what it says about the arguable aspects of experimental design. Tomatis (341) has discussed five debated issues about bioassay.

1. **Doses of chemicals to which test animals are exposed are too high and are not predictive for effects at levels of human exposure.**

High doses of suspect carcinogens are necessary to increase tumor incidence to a level that can be detected in the limited number of animals used for tests (180). A chemical causing tumors at a rate of 0.5 percent in the U.S. population would result in over 1 million cancer cases. But an incidence of 0.5 percent would very probably go undetected in the usual test population of 50 male plus 50 female rats or mice. Administration of the chemical at a 10-times higher dose might increase the response to a detectable level given comparable sensitivity between the test animals and humans. High doses are necessary to increase the sensitivity of the tests, but arguments arise over whether or not a dose is so high that it is not predictive of what happens at lower doses.

The biological argument against depending on results at high doses centers on the contention that such doses may alter metabolism or cause local irritations or other toxic response and cause cancer through a pathway that would not exist at lower doses (60,139,278). A solution sometimes offered to the problems raised by high doses is to run bioassays with many more animals tested at lower doses. The most spectacular attempt at this was the National Center for Toxicological Research **(NCTR) ED_{01} study**

which used 24,000 mice and was designed to detect one cancer in 100 animals. Unfortunately, neither it nor (probably) any experiment can eliminate the necessity for testing chemicals at doses significantly higher than those experienced by humans. Even 24,000 is a small number compared to the number of people exposed to many chemicals.

2. **Routes of exposure in animal tests do not correspond to routes of human exposure.**

Chemicals are administered to laboratory animals in either food, water, by inhalation, force-feeding (gavage), skin-painting, or injection. Few objections are raised to administration in food, water, or by inhalation when the chosen route mimics the route of human exposure. More objections are raised to gavage, skin-painting, injection, or ingestion when that is not the normal human exposure route. However, such methods are sometimes necessary and, furthermore, carcinogens appear to be distributed throughout the body regardless of the route of exposure. EPA (106) stipulates that "to the extent possible, route(s) of administration should be comparable to expected or known routes of human exposure." Adherence to such suggestions will reduce the frequency of this objection.

3. **Some animals used for testing are so biologically different from humans that results from them have no value.**

Choice of animals for bioassays represents a compromise. Most current guidelines require or suggest that chemicals be tested in two rodent species, generally rats and mice. The advantages of these species are their small size, reducing the space necessary for housing, short lifespans (2 to 3 years), reducing the time needed for a lifetime study, and a large amount of information about the genetics, breeding, housing, and health of these animals. Rats and mice are cheap to buy, feed, and house, as compared with larger animals.

Primates are sometimes used for certain toxicological testing. They are certainly more like humans than rodents but their supply is limited. They are expensive, live up to 25 years, and require large areas for housing. Despite these difficulties, NCI now maintains about 600 monkeys for carcinogenicity testing at a cost of about $500,000 per year (2). Dogs are thought to be between rodents and monkeys in their apparent likeness to humans, but they are more like primates in costs.

Differences in metabolism, bioaccumulation, and excretion between rodents and humans should be considered in interpreting the significance of animal results for humans. There is no question that further research in the comparative biochemistry and physiology of man and rodents is necessary, but the comparisons will ultimately be limited by restrictions on what can be determined by experimentation in humans. Moreover, metabolic studies have shown that most differences between humans and experimental animals are quantitative rather than qualitative and support the idea that animal results can be used to predict human responses.

4. **Some test animals or organs of test animals are exquisitely sensitive to carcinogens, and such sensitivity invalidates use of results from such animals.**

Griesemer and Cueto (146) have analyzed the results of testing 190 chemicals in the NCI Bioassay Program (see discussion in Expert *Review of Bioassay Results* and app. A). They identified 35 chemicals that were "strongly carcinogenic" in either the rat or the mouse and noncarcinogenic in the other species. Of the 35, 18 were positive in the mouse and negative in the rat, and 17 were positive in the rat **and negative** in the mouse, which indicates that neither animal was much more often the sensitive species. However, 12 chemicals caused mouse liver tumors, no other lesion in the mouse and no lesions in rats. Taken by themselves these results suggest that the mouse liver is a sensitive organ.

Liver tumors are often found in mice but are infrequently found in U.S. citizens, although they occur frequently in human populations in other parts of the world. Should a chemical that causes mouse liver tumors be considered a hazard? An approach to resolving the mouse liver question was to review how predictive such results are for tumors in other animals. Tomatis, Partensky, and Montesano (343) showed posi-

tive correlation between a chemical's being oncogenic in the mouse liver and its being oncogenic either in the liver or at some other site in rats or hamsters. (Griesmer and Cueto (146) presented data only from rats and mice, and some mouse liver carcinogens in their compilation might be positive in animals other than rats.)

Despite the finding of Tomatis, Partensky, and Montesano (343) that mouse liver carcinogens were often positive at other sites or in other animals, IARC (185,187) considers mouse liver and lung tumors as "limited evidence" for carcinogenicit y. However, OSHA **(272)** accepts mouse liver tumors as "indicators of carcinogenicity" if "scientific experience and judgment" are used in interpreting the data.

Crouch and Wilson (72) have analyzed some of the NCI Bioassay Program data. They calculated the potency and the standard error associated with potency of a chemical in causing a tumor at a particular site in either the rat or the mouse. In their analysis of 35 tested chemicals, they found that in 31 cases the potency in both rats and mice agreed within a factor of 10. Their analysis and similar analysis by Ames et al. **(14), which** consider the inherent error in experiments and the sensitivity of experiments, significantly reduces the number of chemicals that are positive in one species and negative in another. However, those analytical methods have not been generally applied to bioassay results.

5. **Finding benign tumors in test animals has no value in defining carcinogenicity.**

Tumors can be divided into two classes, benign and malignant. Benign tumors do not metastasize, the tumor cells remain in contact with each other and do not invade other tissues or organs. Malignant tumors can invade other tissues and metastasize, spreading to distant parts of the body and causing other tumors. Both types of tumors are found in experimental animals.

Hearings on pesticide regulations have been marked by repeated arguments about the importance of benign tumors. Should benign tumors found in experimental animals be taken as evidence that a chemical causes cancer?

The position that a benign tumor may later become malignant, that the line of demarcation between benign and malignant is unclear, and that benign tumors can also be life-threatening has prevailed in regulatory agencies. Therefore, no distinction is made in regulatory decisions between benign and malignant tumors. This is clearly reflected in FIFRA guidelines (102) and TSCA test standards (106) in which the endpoint is oncogenicity (tumor causation) rather than carcinogenicity (which emphasizes malignancy) (see table 27).

Griesemer and Cueto **(146)** reported no difficulty in deciding between benign and malignant tumors in the NCI Bioassay Program and that their conclusions about carcinogenicity were unaffected by including or excluding benign tumors. IARC statements reflect difficulties in the interpretation of benign tumors. As mentioned above, it considers the occurrence of some benign tumors as "limited evidence" for concluding that a chemical is a carcinogen (185). It also states that "preneoplastic lesions" may progress to "frank malignancy." Furthermore, IARC (186,187) considers that the occurrence of both preneoplastic and neoplastic lesions in the same organ strengthens conclusions about carcinogenicit y.

Expert Review of Bioassay Results

In addition to the general objections to bioassay procedures that have been discussed, specific objections may be raised to particular tests. For instance, animals may have been inadvertently exposed to more than one chemical, to pathogenic micro-organisms, to extreme temperature, or to temporary deprivation of food or water, any one of which might influence results. Another frequent item of contention is whether or not a particular pathologic lesion indicates carcinogenicity. Critical reviews can reduce concern that flaws in experimental design or conduct mar the experiment or bring its results into question.

IARC periodically calls together panels of experts to review the worldwide literature about the carcinogenicity of particular chemicals or exposures. The IARC Monograph Program, be-

gun in 1970, had published reviews of bioassays of 422 chemicals and processes by 1979. For 60 of the 422 chemicals and processes, IARC was able to evaluate both human and animal evidence for carcinogenicity. Those 60 are described in the epidemiology section below.

For the remaining 362 chemicals and processes (185),

> . . . there was no information available from studies in humans [but] the IARC was asked repeatedly to consider making an assessment of the carcinogenic risk for humans which was based only on animal data.

In response to those requests, the IARC working group recommended (185):

> . . . that in the absence of adequate data in humans it is reasonable, for practical purposes, to regard chemicals for which there is sufficient evidence of carcinogenicity (i.e. a causal association) in animals *as if they presented a carcinogenic risk* for *humans* (emphasis in original).

An IARC working group defined five categories of evidence--sufficient, limited, inadequate, negative, and no data—against which experimental data are to be compared (see app. A). The working group decided that for 142 chemicals there was sufficient evidence, for about 100 there was limited evidence, and for the remainder there was inadequate evidence for carcinogenicity. The IARC exercise is especially important because it shows that experts from both the private and public sectors, sitting on IARC panels, can consider experiments and results and reach decisions about their meaning.

Griesemer and Cueto (146) analyzed the results of NCI's testing 190 chemicals. IARC (185) and Griesemer and Cueto (146) classifications of chemicals included 33 chemicals in common. Analysis of the classification of those 33 (see app. A) shows that there was good agreement about the more carcinogenic of the chemicals. Such agreement lends credence to the idea that tests carried out in different laboratories and analyzed by different experts can lead to similar conclusions about carcinogenicity.

Policy Considerations About Bioassays

General acceptance of bioassays as predictors of human risk is sometimes obscured by controversies about particular test results. The Federal Government, in response to controversies arising from testing artificial sweeteners or pesticides, has asked a number of expert panels to consider bioassay designs, results, and usefulness. In all cases, the panels endorsed bioassays while reserving judgment about particular tests and attaching caveats to some results. Table 28 is a listing of some Government-affiliated panels and reports plus two trade associations, the American Industrial Health Council (AIHC) and the Food Safety Council, and a union organization which have commented on bioassays.

The Office of Science and Technology Policy (OSTP), in the Executive Office of the President (281) produced a good, brief exposition of methods, a well-crafted list of recommendations about carcinogen testing, and suggestions for a Federal carcinogen policy. About bioassays, the report says,

> . . . it would seem prudent to view a positive test result in a carefully designed and well-conducted mammalian study as evidence of potential human carcinogenicity.

As indicated by the quote, some organizations urge consideration of experimental design and execution before drawing conclusions about carcinogenicity. AIHC (8), in commenting on OSHA's **(278)** proposed cancer policy, endorsed bioassays as predictors of human risk:

> There is agreement further that a substance which is a confirmed oncogen in two mammalian species should be subject to regulation as a probable human carcinogen.

All Federal regulatory agencies accept animal test results as predictors of carcinogenic risk for humans. The Interagency Regulatory Liaison Group (IRLG) was formed in 1977 to aline the policies of the Consumer Product Safety Commission (CPSC), FDA, the Food Safety and Quality Service of the Department of Agriculture, EPA, and OSHA. IRLG stated (180):

Table 28-Some Organizations That Have Considered and Endorsed Animal Tests as Predictors of Human Risk From Chemical Carcinogens

Organization	Publication
The Secretary's (HEW) Committee on Pesticides and their Relationship to Environmental Health	Report (325)
National Research Council	— Evaluating the Safety of Food Chemicals (259) — Safety of Saccharin and Sodium Saccharin in the Human Diet (261) — Pest Control: An Assessment of Present and Alternative Technologies (262) — Principles for Evaluating Chemicals in the Environment (263) — Food Safety Policy (269)
Office of Technology Assessment	— Cancer Testing Technology and Saccharin (282) — Environmental Contaminants in Food (284)
National Cancer Advisory Board, Subcommittee on Environmental Carcinogenesis	— General Criteria for Assessing the Evidence for Carcinogenicity of Chemical Substances (245) — The Relation of Bioassay Data on Chemicals to the Assessment of the Risk of Carcinogens for Humans Under Conditions of Low Exposure (246)
American Industrial Health Council	— AIHC Recommended Alternative to OSHA's Generic Carcinogen Proposal (8)
Food Safety Council	— Proposed System for Food Safety Assessment (125)
Occupational Safety and Health Administration	— Identification, Classification and Regulation of Potential Occupational Carcinogens (279)
Interagency Regulatory Liaison Group	— Scientific Bases for Identification of Potential Carcinogens and Estimates of Risks (180)
Office of Science and Technology Policy	— Identification, Characterization and Control of Potential Human Carcinogens: A Framework for Federal Decision-making (281)
American Federation of Labor, Congress of Industrial Organizations, and United Steelworkers of America	— Post-Hearing Brief on OSHA's Proposed Standard on the Identification, Classification and Regulation of Toxic Substances Posing a Potential Occupational Carcinogenic Risk (7)

SOURCE: Off Ice of Technology Assessment.

> Evidence of the carcinogenic activity of an agent can be obtained from bioassays in experimental animals.

and,

> Positive results, obtained in one species only are considered evidence of carcinogenicity. Positive results in more limited tests (e.g., when the observation period is considerably less than the animal's lifetime), but by experimentally adequate procedures, are acceptable as evidence of carcinogenicity. Negative results, on the other hand, are not considered as evidence of lack of a carcinogenic effect, for operational purposes, unless minimum requirements have been met.

As these quotes show, both AIHC and IRLG accept bioassays as predictors of potential human risk. However, they differ in the weight of evidence each requires. AIHC wants positive test results in two species before making decisions about carcinogenicity, while IRLG will consider making a decision on the basis of positive results in one species. Labor organizations (e.g., American Federation of Labor, Congress of Industrial Organizations, and United Steelworkers of America (7)) and environmental groups also consider positive results in a single animal as sufficient to make a judgment about carcinogenicit y.

How Many Chemicals Are Carcinogens in Animal Tests?

The number of chemicals that are carcinogenic in humans or animals is uncertain. A few estimates have been made, but the bases for the estimates are poor.

OSHA (279) states: "Most substances do not appear to be carcinogenic, " and uses information from two lists of test results to support the statement. OSHA cites the study of Innes et al. (176), which tested 120 industrial chemicals and pesticides. Eleven (less than ten percent) were reported to be carcinogenic, 20 were considered to require further study, and 89 "did not give significant indication of tumorigenicity. " A National Academy of Sciences review of pesticides (262) also drew attention to the low number of positive results reported in the Innes study. The small number was judged to be especially significant because of the large number of biologically active pesticides included in the test. Certain reservations must be attached to these conclusions. The maximum dose tested and the number of animals used in the experiments were smaller than now required, and consequently, the experiments were less conclusive than more recent ones.

A more direct criticism of relying on the Innes et al. (176) document as a measure of how many chemicals are carcinogens is that it was a preliminary report. Based on the complete report (26), Barnard (21) states that 29 percent of the tested compounds were carcinogenic.

The second source cited by OSHA (279) is the seven-volume Public Health Service (PHS) list of chemicals tested for carcinogenicity (298). OSHA (279) reported that about 17 percent of the 7,000 tested chemicals were tumorigenic. However, OSHA concluded that these data "overstated the true proportion of carcinogenic substances in the human environment" because suspicious chemicals are selected for testing.

Despite such reservations, the Task Force on Environmental Cancer, Heart and Lung Disease (339) said:

> Of the upwards of 100,000 known chemicals of potential toxicity, only approximately **6,000** have been tested for carcinogenicity. It is estimated that 10 to **16 percent** of the chemicals so tested provide some evidence of animal carcinogenicity.

To treat the reported 10 to 16 percent of tested chemicals as carcinogens as a reliable number is probably an error because many of the 7,000 tests are clearly inadequate when measured against current testing guidelines. Scientists employed by the General Electric Co. (141a) have also examined the lists from the seven volumes (298). **Using a relaxed** criterion for carcinogenicity " . . . any listing that reported an incidence of tumors in the test animals higher than in the control animals was scored as 'positive, ' " upwards of 80 percent of the listed chemicals were found to be positive. It is important to recognize that this relaxed criterion would not be accepted in any regulatory decision, and it exaggerates the number of positives. Furthermore, biases toward testing chemicals that are likely to be carcinogenic and toward reporting positive results push the percentage of positives upward.

OSHA (279) had a contractor analyze a list of 2,400 suspect carcinogens compiled by the National Institute of Occupational Safety and Health (NIOSH). The contractor estimated that for 570 (24 percent of the total) there were sufficient data to permit initiation of rulemaking to classify them as category I or 11 carcinogens under the proposed OSHA policy.

Neither the Innes study nor the PHS list provides information as reliable as that which is more recently available from NCI, IARC, and NTP. Fifty-two percent of NCI-tested and reported chemicals were carcinogens (146). Either "sufficient" or "limited" data existed to classify about 65 percent of IARC-listed chemicals as carcinogens (186,344). NTP (273) reported that 252 (including the 190 reported by Griesemer and Cueto (146)) tests have been completed in the NCI Bioassay Program. Under conditions of the tests, 42 percent were positive, 9.5 percent were equivocal, 36 percent were negative, and 13 percent were inconclusive.

The biases toward testing likely-to-be-positive chemicals cannot be ignored, and the fact that negative test results are less likely to be published and included in any compilation (279) further complicates analysis of lists of tested chemicals. These factors tend to increase the percentage of positive chemicals, and consequently may inflate the estimates of the percentages of chemicals that are carcinogens.

A definitive answer to questions about how many chemicals are carcinogens would depend on testing every chemical, and that is beyond the capacity of the bioassay system. Tomatis (341) reported that 828 chemicals were under test worldwide in governmental programs in 1975, and that 317 were repeat tests of chemicals for which, in his opinion, adequate data already existed. He did not estimate the number of chemicals under test in private or commercial laboratories.

Finding more and more chemicals to be carcinogenic in bioassays raises important policy questions and may force a decision to rank carcinogens for possible regulation or voluntary reductions. It is not apparent how to deal with a large number of carcinogens without ordering them according to their riskiness.

SELECTION OF CHEMICALS FOR TESTING IN BIOASSAYS AND SHORT-TERM TESTS

Tests develop information to aid in making decisions about chemicals, and the most important step in information development is placing the chemical on test. Limited testing capacity makes it necessary to pick and choose among chemicals. Not all can be tested. The second NTP Annual Report **(272)** underlines this point in discussing bioassays:

> It is unreasonable even to attempt to study all **48,000** chemicals in this way. Current known world capacity permits the initiation of perhaps **300** such chemical tests each year, and the results published this year are from the tests begun 4 or more years ago. This capacity—even if financial resources were not limiting—could be no more than doubled in the next 5-10 years. At this rate, it would take an additional 80 years to study all currently existing chemicals. Further, approximately 500 new chemicals are introduced into commerce each year, and this would result in an additional backlog of some 40,000 untested chemicals.

The same report pinpoints Federal Government testing capacity:

> In carcinogenesis testing the NTP proposes to start **75** new chemicals on the lifetime carcinogenicity bioassays in fiscal year 80. This is the same as the level achieved in fiscal year 79 and is a two-fold increase over the rate of testing prior to the establishment of the program. [In fact, 50 were started in fiscal year 1980; 30 are expected in fiscal year 1981.]

By spring 1980, the Federal Government had centered its selection activities on two programs, the Interagency Testing Committee (ITC), and the NTP Chemical Nomination and Selection Committee. ITC recommends chemicals for testing to the EPA Administrator under jurisdiction of section 4 of TSCA (the ITC list) and NTP selects chemicals for testing by the Federal Government. In addition, NCTR and CIIT have also published criteria for testing substances.

The Interagency Testing Committee

Section 4(a) of TSCA stipulates that EPA issue a rule to require testing if an individual chemical or category of chemicals "may present an unreasonable risk of injury to health or the environment . . . or will be produced in substantial quantities and . . . enter the environment in substantial quantities or there is or may be significant or substantial human exposure. " The vagueness of such terminology, "unreasonable risk, " "substantial quantities, " "significant or substantial human exposure, " requires EPA to generate interpretations within scientific, legal, and policy contexts.

Section **4(c)** of TSCA established ITC to recommend chemical substances or mixtures which should be given priority consideration for testing. ITC is composed of eight members, one

each from EPA, OSHA, the Council on Environmental Quality (CEQ), NIOSH, NIEHS, NCI, the National Science Foundation (NSF), and the Department of Commerce.

ITC is to give priority to those chemical substances which are known or suspected to cause or to contribute to the causation of cancer (carcinogens), mutations (mutagens),- or birth defects (teratogens). The total number of chemical substances or mixtures recommended for testing at any one time is not to exceed 50. TSCA specifies that ITC must update, as necessary, the testing priority list at least every 6 months. The EPA Administrator is directed to respond to a chemical's being placed on the list within 12 months. The Administrator must either initiate rulemaking to require testing or publish reasons for not initiating a testing rule.

ITC's initial selection of chemicals for consideration was made by combining about 20 Government-compiled lists of potentially hazardous substances. Chemicals that did not come under TSCA's purview were eliminated from further consideration, and the remaining substances were ranked against two measures: exposure and biological activity. The bases for determining exposure were explicitly specified by Congress in section 4(e)(i)(A) of TSCA:

- general population exposure (number of people exposed, frequency of exposure, exposure intensity, penetrability);
  quantity released into the environment (quantity released, persistence);
- production volume; and
- occupational exposure.

In general, chemicals which were ranked high on the exposure scale were further ranked against biological activity measurements. TSCA specified the first three factors, and the others were included because of their significance in characterizing biological effects (183):

- carcinogenicity,
- mutagenicity,
- teratogenicity,
- acute toxicity,
- other toxic effects,
- ecological effects, and
- bioaccumulation.

ITC examined the biological activity score and the individual biological and exposure factors, and weighed the biological scores against the exposure scores to select chemicals for detailed reviews. After the reviews, chemicals were recommended to the EPA Administrator for consideration. A detailed description of the scoring system can be found in the initial ITC report to the Administrator of the EPA (183).

As of November 1980, ITC had filed seven reports to the EPA Administrator. Each report has updated and revised the list of chemicals eligible for promulgation of test rules under section 4(a) of TSCA.

Chemical Categories

An important consideration in selecting chemicals for testing is how to deal with chemical categories or classes. Section 26(c) of TSCA specifies that any action authorized or required under the act "may be taken by the Administrator . . . with respect to a category of chemical substances or mixtures [and] shall be deemed to be a reference to each chemical substance or mixture in such category. " In making its recommendations ITC selected some categories of chemicals for testing. EPA must determine whether to evaluate the scientific, economic, and regulatory consequences of every present or potential member of a category of chemicals or to evaluate selected representatives from the category. If the decision is made to test representatives, EPA has to assess whether it could "reasonably" extrapolate test results to chemicals within the group that have not been studied. Some industry representatives have questioned the validity of using categories for chemical testing, particularly the categories recommended by ITC. They express concern that it would be unrealistic to establish a general policy for choosing which chemicals should be tested for every possible category. Charles Holdsworth, of the American Petroleum Institute, recommended that EPA select members of a category once the "metabolic actor" for the group is determined. Another possible approach would be to select chemicals of interest because of exposure or production volume.

Disposition of the ITC List

As of May 1981, 3-1/2 years after completion of the first ITC list, EPA had issued no final rule requiring testing of any ITC-identified chemicals. However, EPA (110) in the summer of 1980 proposed its first health effects test rule on chemicals nominated by ITC. The proposed rule would require testing of chloromethane and representative members of the category of chlorinated benzenes for oncogenicity and other health effects. Support documents describing the reasons for deciding to require testing and an economic impact analysis of the tests were also released (110,112). As of April 1, 1980, draft copies of test rules for four additional chemicals and groups of chemicals from the 21 identified on the first three ITC lists were available from EPA.

The General Accounting Office (141) reported that EPA estimates it will take 5 years to issue a final test rule for an ITC-nominated chemical. That time is required for the agency to evaluate the published information about the chemical. EPA further estimates that an additional 4 years will be required for execution of bioassay and analysis of their results. At a minimum, then, 9 years will pass between the decision to test an ITC-selected chemical and completion of testing. Certainly the long time between the decision to require testing and the production of results is of concern to EPA, and the agency is working to reduce it (141).

There is a suggestion that many ITC-selected chemicals are currently under test in other arms of the executive branch. NTP (273) reported that of 20 single chemicals and 15 classes recommended by ITC (as of April 1980), 16 of the 20 chemicals and representative chemicals from 14 of 15 classes were then under test or scheduled for test by NTP. NTP (273) did not make a comparison between the types of tests recommended by ITC and the types of tests being carried out or planned by NTP. The NTP report draws attention to a problem common to an active field such as toxicology. ITC judgments about whether or not to require a test are based or what it knows about in-progress and completed tests. It is important that NTP and other testing organizations share their latest information and plans with ITC. Accordingly, a liaison representative of NTP attends and participates in ITC meetings and related activities.

Responsibility for Testing Under TSCA

Once a chemical or category is selected for testing, EPA must determine who should bear the responsibility and burden of testing. TSCA requires EPA to indicate whether manufacturers, processors, or both manufacturers and processors bear the responsibility. EPA is presently evaluating exposures that occur at various points in a chemical's life cycle. If EPA finds that the chemical's manufacture may present a risk, only the manufacturers must test. If processing may pose a hazard, only the processors are required to test. However, if distribution in commerce, use, or disposal of the chemical may present a risk, both the manufacturers and the processors are required to test. This determination has substantial economic and legal ramifications since it will establish the universe of firms which must bear the cost of testing. Some of the chemicals for which test rules have been drafted are manufactured by more than one company. EPA is urging that firms cooperate to sponsor single tests of the chemicals rather than have each company sponsor its own test.

National Toxicology Program

NTP was established within DHEW (now Department of Health and Human Services (DHHS)) on November 15, 1978, to further the development and validation of integrated toxicological test methodologies. The NTP Executive Committee is composed of the heads of FDA, OSHA, CPSC, EPA, National Institutes of Health (NIH), NIOSH, NCI, and NIEHS.

NTP's annual plan for fiscal year 1980 (272) describes methods to select chemicals for testing: NTP operates under the principle that industry will test chemicals for health and environmental effects as intended and mandated *by* Congress under legislative authorities. However, some chemicals will not be tested by the private sector, and NTP will select chemicals for

its own testing program from the following categories:

1. chemicals found in the environment that are not closely associated with commercial activities;
2. desirable substitutes for existing **chemicals, particularly therapeutic agents, that might not be developed or tested without Federal involvement;**
3. **chemicals that should be tested to improve scientific understanding of structure-activity relationships and thereby assist in defining groups of commercial chemicals that should be tested by industry;**
4. **certain chemicals tested by industry, or by others, the additional testing of which by the Federal Government is justified to verify the results;**
5. **previously tested chemicals for which other testing is desirable to cross-compare testing methods;**
6. **"old chemicals" with the potential for significant human exposure which are of social importance but which generate too little revenue to support an adequate testing program (some of these may be "grandfathered" under FDA laws);**
7. two or more chemicals together, when combined human exposure occurs (such testing probably cannot be required of industry if the products of different companies are involved); and
8. in special situations, as determined by the executive committee, marketed chemicals which have potential for large scale and/or intense human exposure, even if it may be possible to require industry to perform the testing.

NTP solicits lists of chemicals from NTP research (NCI, NIEHS, FDA, NIOSH) and regulatory (FDA, OSHA, CPSC, EPA) agencies, other Federal agencies, academia, industry, labor, and the public. All of the chemicals suggested for study are funneled to the NTP Chemical Nominations Group.

The Chemical Nominations Group, composed of representatives from EPA, OSHA, FDA, CPSC, NIH, NCI, NIEHS, and NTP, prepares a dossier describing what is known about the physical properties of each chemical, its production volume, its use, exposures to it, and any toxicological information, Each chemical is judged against the chemical selection principles described above and nominations are forwarded to the NTP Executive Committee which makes final decisions about which chemicals to place on test.

A decision by NTP to test a chemical does not mean necessarily that the chemical will be placed on a bioassay program. It may mean that the chemical will be entered into less expensive, short-term tests first, and depending on the results of those tests, subsequent decisions will be made about whether testing should continue.

The National Center for Toxicological Research

NCTR, a research arm of EPA and FDA was established to develop a better understanding of adverse health effects of potentially toxic chemicals. NCTR's research emphasis is on determination of adverse health effects resulting from long-term, low-level exposure to chemical toxicants (food additives, residues of animal drugs, etc.); determination of the basic biological processes involving chemical toxicants in animals in order to enable better extrapolation of toxicological data from laboratory animals to man; and development of improved methodologies and test protocols for evaluation of the safety of chemical toxicants (good laboratory practices, automated data systems, etc.). NCTR chooses substances for testing from the following categories (271):

1. substances that have no clear industrial sponsorship and for which it is determined that further toxicological data are needed. Usually these are either food contaminants, GRAS (generally recognized as safe food additives) compounds, or cosmetic ingredients.
2. substances that can act as model compounds in a continuing toxicological methods development program.

3. substances for which there is a pressing need to acquire toxicological data above and beyond that which may be supplied by industry.
4. studies as a direct regulatory response to consumer complaints.

Chemical Industry Institute of Toxicology

CIIT was established in 1974 as an independent, nonprofit research laboratory financed by annual contributions from the member companies. Membership in CIIT is open to any corporation or other business entity whose activity consists to a substantial extent of the manufacture, processing, or use of chemicals and any formal association of such entities. CIIT is to provide objective study of toxicological problems problems involved in the manufacture, handling, use, and disposal of commodity chemicals.

CIIT has developed a set of criteria to select and rank chemicals into priority lists for study. These criteria are:

. volume of production,
- physical and chemical properties,
- estimated human exposure,
- toxicological suspicion and opinion,
- public interest, and
- significance to society.

To date about 40 chemicals have been selected by CIIT for review and study. CIIT's testing showed that formaldehyde caused nasal cancer in rats. Those results have been used by Federal agencies in considering regulations about the chemical.

TIER TESTING

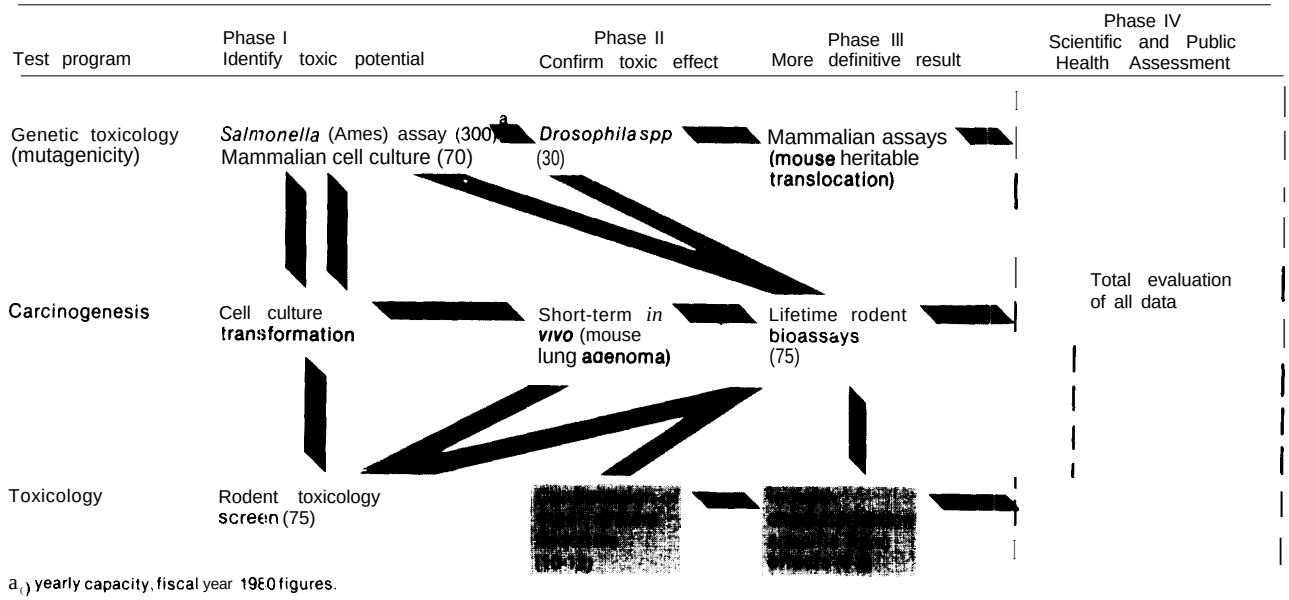
This chapter has so far discussed various test procedures, from quick, low-cost molecular structure analysis through relatively quick, relatively cheap short-term tests to long-term, high-cost bioassays. The fourth category of test, epidemiologic studies, differs from the other three. Detection of a carcinogen because it causes human illness and death can be regarded as a failure in hazard identification, because the other three tests should have or might have predicted the risks before the substance had a chance to inflict harm.

During the last several years, a number of expert committees and panels have discussed an ordered approach to testing—proceeding from quick, cheap tests to longer, more expensive tests. One such "tier testing" plan was developed by an expert group drawn from academic institutions, public-interest organizations, industry and Government agencies (66). A repeated criticism expressed in letters that commented on the plan was the absence of criteria on which to drop a chemical from further testing requirements (66), The point was made that the tier system developed by the Conservation Foundation was actually a sequential test series, since once a chemical entered the test series it would apparently continue on through every test.

Tier testing has no place in regulations of carcinogens under FIFRA and TSCA, since EPA regulation of substances as a carcinogen requires bioassay or human data (102, 106). To talk of a tier testing system under those regulations is academic, but evidently EPA is considering a role for short-term tests for making decisions about carcinogenicity. In the suit brought by the Natural Resources Defense Council against EPA because of its failure to act on the ITC chemicals, Warren Muir of EPA said: " . . . EPA is in the process of considering what kinds of results from short-term tests suggest the need to require long-term tests for the potential for causing cancer . . . " (243).

An approach to tier testing appears in the 1980 NTP annual plan and is described in figure 20. The close interrelationships between genetic toxicology and carcinogenesis test programs are shown by the lines which connect them. The ab-

Figure 20.—Interrelationships of Major Testing Activities of NTP



a) yearly capacity, fiscal year 1980 figures.

SOURCE: Off Ice of Technology Assessment, adapted from NTP (272)

sence of arrowheads on the lines is intentional; according to David Rail, NTP Director (303), there has not yet been enough experience with the scheme to be certain which phase I tests should come first or whether all chemicals should go through all phase I tests.

A critical feature of tier testing is the ability to make decisions about whether or not to continue testing from phase I to II, or from 11 to III. Guidelines are necessary for making the decision that a chemical is sufficiently without risk and that no further testing is necessary. Rail (303) said that development of decisionmaking guidelines is a priority item for NTP in 1980 and 1981. NTP intends to analyze the testing histories of chemicals that have gone through all three phases (albeit not necessarily under NTP aegis) to determine which test results were most predictive of the ultimate decision about the chemical. NTP will take advantage of the fact that the most expensive and time consuming testing, phase III, has been completed on some chemicals which have not otherwise been tested. Such chemicals will be entered into phase I and II testing to provide additional information about which tests are most predictive. Finally, the NTP decision to continue testing a few chemicals that are negative in phase I tests will provide additional information.

EPIDEMIOLOGY

Lilienfeld (210) defined epidemiology as the study of the distribution of disease in human populations and of the factors that influence disease distribution. Epidemiologic techniques are useful for identifying causative agents and conditions **that** predispose for cancer. Studies can determine associations in populations between exposure to carcinogenic agents or between aspects of lifestyle and increased cancer risk.

The earliest association of a factor with cancer was made by Ramazzini in 1713. He found that nuns had a higher rate of breast cancer than other women and, in his Treatie on

the Diseases of Tradesman of 1700, he attributed the increase to celibacy (322). That association has been sharpened to include the observation that women who deliver a child at an early age are less likely to develop breast cancer. In addition to identifying cancer risks, epidemiology may play a positive role by pointing out less hazardous diets or agents which are protective against cancer.

For the purpose of this discussion, epidemiologic studies are divided into three general types: 1) experimental, 2) descriptive, and 3) observational. While several basic strategies exist, there are no rigid study designs within any of these categories. Flexibility is important since, unlike laboratory experiments, epidemiology examines groups of unpredictable people living in dynamic environments. The importance of flexibility is underlined by IRLG Guidelines, which state that epidemiologic study design must "be described and justified in relation to the stated objective of the study" (181).

Experimental Epidemiology

The ideal procedure for investigating cause-and-effect hypotheses is through experimental epidemiology. This type of study requires the deliberate application or withholding of a factor and observing the appearance or lack of appearance of any effect. Given the severity of cancer, ethical considerations preclude the administration of suspected carcinogens to people, though it is possible to test agents thought to aid in prevention (292).

Experimental epidemiology studies are difficult to conduct because of the need to secure the cooperation of a large group of people willing to permit an experimenter to intervene in their lives. The investigator must have reason to believe that the proposed intervention, whether a deliberate application or withholding, will be beneficial, but at the same time, he must be somewhat skeptical of the effects. Once sufficient evidence leads to the conclusion that the intervention is or is not beneficial, the experiment must be terminated.

Descriptive Epidemiology

Descriptive epidemiology studies examine the distribution and extent of disease in populations according to basic characteristics—e.g., age, sex, race, etc. The primary purpose of conducting descriptive epidemiologic studies is to provide clues to the etiology of a disease which may then be investigated more thoroughly through more detailed studies. Descriptive studies have focused on international comparisons and comparisons among smaller geographical regions, such as U.S. counties (29).

The identification of high bladder cancer rates in New Jersey males and excess mortality rates from cancer of the mouth and throat, esophagus, colon, rectum, larynx, and bladder in the industrialized Northeast have suggested that occupational factors might be incriminated and have prompted additional investigations. Blot et al. (29) describe NCI's stepwise approach to search for etiological clues. Examination of age-specific rates of disease occurrence or mortality across time (see ch. 2) is another example of descriptive epidemiology.

Observational Epidemiology

Observational epidemiology depends on data derived from observations of individuals or relatively small groups of people. These studies are analyzed using generally accepted statistical methods to determine if an association exists between a factor and a disease and, if so, the strength of the association. Often the hypothesis to be investigated arises from the results of a descriptive study. NCI has embarked on several observational studies based on findings from their county-correlation studies. For example, high rates of lung cancer were found in the Tidewater Virginia area, and a large study was initiated which found elevated risk for lung cancer in shipbuilders and smokers.

Cohort Studies

Two types of observational epidemiology studies, cohort and case-control studies, differ

80-481 O - 81 - 10

in the selection of the population groups for study.

A cohort study starts with a group of people, a cohort, considered free of the disease under study, and whose disposition regarding the risk factor under consideration is known. Usually the risk factor is an exposure to a suspect carcinogen or a personal attribute or behavior. The group is then studied over time and the health status of the individual members observed. This type of study is sometimes referred to as "prospective" because it looks forward from exposure to development of the disease characteristic (210,225). Cohort studies can be either concurrent or nonconcurrent in design. Concurrent cohort studies depend on events which will occur in the future, while nonconcurrent cohort studies rely on past data or past events,

Case-Control Studies

In a case-control study, individuals with the disease under study (cases) are compared to individuals without the disease (controls) with respect to risk factors which are judged relevant. Some authors label this study design "retrospective" because the presence or absence of the predisposing risk factor is determined for a time in the past (210,225). However, in some cases the presence or absence of the factor and the disease are ascertained simultaneously.

The choice of appropriate controls is rarely without problems. Often, for practical reasons, controls are chosen from hospital records. However, they may not be representative of the population, and they therefore may introduce "selection bias," as discussed by MacMahon and Pugh (218).

In case-control and cohort studies, the groups selected should be comparable in all characteristics except the factor under investigation. In case-control studies, the groups should resemble each other except for the *presence* of the disease, while in cohort studies, the study and comparison groups should be similar except for *exposure* to the suspect factor. Since this rarely is possible in practice, comparability between groups can be improved by either matching individual cases and controls (in case-control studies) or by standard statistical adjustment procedures (in either case-control or cohort studies). Demographic variables, e.g., age, sex, race, socioeconomic status, are most commonly used for adjustment or matching.

There are advantages and disadvantages with both the case-control and cohort studies (see table 29). Case-control studies tend to be less expensive to conduct, require relatively fewer individuals, and many have been especially useful in studying cancer. The great advantage of cohort studies is that they allow observation of all outcomes, not only those originally anticipated. Bias is somewhat reduced in cohort studies since classification into an exposure category cannot be influenced by prior knowledge that the disease exists. In a concurrent cohort study, it is often necessary to wait many years for the manifestation of enough disease cases to conduct an analysis. The cost and time of the study can be reduced if conducted nonconcurrently. Cohort studies tend to require many more subjects than case-control studies and assignment of individuals to the correct cohort for analysis is difficult.

Causal Associations

A pragmatic view of causality is necessary, particularly when studying complex, multifactorial diseases such as cancer. Analysis of the association between exposure and disease in an epidemiologic study depends on tests of statistical significance. However, finding a positive statistically significant association is not sufficient to conclude a causal relationship. Artifactual and indirect associations must be considered. As MacMahon and Pugh (218) state,

> . . . only a minority of statistical associations are causal within the sense of the definition, which requires that change in one party to the association alters the other. "

Policy Considerations About Epidemiology

While short-term tests and bioassays are used to evaluate a chemical's carcinogenic potential in the laboratory, the effect on humans is direct-

Table 29.—Advantages and Disadvantages of Case-Control and Cohort Studies

Type of study	Advantages	Disadvantages
Case-control	Relatively inexpensive	Complete information about past exposures often unavailable
	Smaller number of subjects	Biassed recall
	Relatively quick results	Problems of selecting control group and matching variables
	Suitable for rare diseases	Yields only relative risk
Cohort	Lack of bias in ascertainment of risk factor status	Possible bias in ascertainment of disease
	Yields incidence rates as well as relative risk	Large numbers of subjects required
	Can yield associations with other diseases as by-product	Long follow-up period
		Problem of attrition
		Changes over time in criteria and methods
		Very costly
		Difficulties in assigning people to correct cohort

SOURCE Office of Technology Assessment

ly assessed by epidemiologic techniques. Well-conducted and properly evaluated epidemiology studies which show a positive association are accepted as the most convincing evidence about human risks.

Negative epidemiologic results show that exposure of a certain number of people to a substance at a specified level did not cause cancer. *From* such results, it is possible to calculate that human risk is no higher than what the study could have detected. For instance, a study of 1,000 people which showed no excess cancer would be "more negative" than one of 100 people exposed at the same level. Neither study would show that a risk exists, and neither shows that no risk exists, but the larger study shows a lower probability of risk.

The OSHA Generic Cancer Policy (279) proposed that OSHA, after ascertaining the adequacy of the study design, would interpret negative epidemiologic studies as setting an upper limit for human risk. AIHC (8) wants negative human evidence to be considered along with animal data in making decisions about carcinogenicity. The OSHA position, and that of Federal Government regulatory agencies in general (306), is to use epidemiology to estimate limits of risk, but not to weigh negative human evidence against other positive evidence in deciding whether or not a substance is a carcinogen.

The Regulatory Council (306) considers properly designed and conducted epidemiologic studies, which show a significant statistical relationship between human exposure to a substance and increased cancer risk, to provide "good evidence" that a substance is carcinogenic. The Council mentioned some difficulties in epidemiology, e.g., long latency periods, multiplicity of exposures, and cautioned that often "even large increases (which could involve thousands of people) . . . cannot be detected." For these reasons they cite two caveats in using epidemiological studies:

> The failure of an epidemiological study to detect an association between the occurrence of cancer and exposure to a specific substance should not be taken to indicate necessarily that the substance is not carcinogenic.
>
> Because it is unacceptable to allow exposure to potential carcinogens to continue until human cancer actually occurs, regulatory agencies should not wait for epidemiological evidence before taking action to limit human exposure to chemicals considered to be carcinogenic.

OSTP (281) states that "a positive finding in a well-conducted epidemiologic study can be viewed as strong evidence that a chemical poses

a carcinogenic risk to humans. ” Alternatively, “a negative finding is not nearly so meaningful . . . “ and OSTP emphasizes the importance of examining the sensitivity of a negative study, and suggests that the upper limit of risk that might have gone undetected in the study be calculated and presented.

Carcinogens for Which There Is Human Evidence

Through various means, epidemiologic studies have identified several human carcinogens. The first use of epidemiologic principles to relate environmental contaminants to human cancer is credited to Pott in 1775 (322). Pott, a physician in London, suggested that scrotal cancers, which he observed in men who had worked as chimney sweeps when boys, were caused by exposure to soot. Pott is an example of an astute physician recognizing an unusual cluster of cancer cases. A more recent example is vinyl chloride which was identified as a carcinogen after three cases of a rare liver tumor (hepatic angiosarcoma) were diagnosed in workers in a manufacturing plant (71). In the case of vinyl chloride, evidence for its carcinogenicity in laboratory animals was available in advance of the human evidence. In both cases, control action followed the demonstration of occupational risk. The Danish Chimney Sweeps Guild instructed its members about protective clothing and to practice preventive hygiene soon after Pott’s report was published, and OSHA regulated vinyl chloride.

IARC bases its evaluation of carcinogenic risk to humans on consideration of both epidemiological and experimental animal evidence. IARC considered human evidence bearing on 60 chemicals and industrial processes and classified 18 as human carcinogens (see table 30). Many of the human data considered by IARC are from studies concerning workplace and medical exposure. This does not necessarily reflect the distribution of carcinogens but more likely the higher exposure and relative ease of performing epidemiologic studies on patients and occupational groups. For instance, the availability of medical records facilitates locating people exposed to a drug and provides information about time of exposure and dose level.

IARC also classified 18 additional compounds as *probably* carcinogenic for humans but there was insufficient evidence to establish causal associations. These 18 were further subdivided according to the degree of evidence, high or low, as displayed in table 30. Insufficient evidence was available to decide about the carcinogenicity of 18 chemicals listed in table 30. Finally, because of time limitations, IARC was unable to evaluate six compounds for which human data exist.

Annual Report on Carcinogens

In an effort to provide information on carcinogens which would be useful to regulatory agencies, Congress passed an amendment to the Community Mental Health Act (Public Law 95-622). It requires the Secretary of DHHS to publish an annual report containing a list of substances which are known to be carcinogens or may reasonably be anticipated to be carcinogens and to which a significant number of persons residing in the United States are exposed. The task was assigned to NTP, and the first report (82) includes the 26 exposures which IARC had determined in 1978 to be human carcinogens (344). Candidates for the 1981 and 1982 list also will be drawn from IARC beginning with chemicals and processes judged “probably carcinogenic for humans. ” The initial report did not, as required by the statute, list “ . . . all substances which either are known . . . or . . . may reasonably be anticipated to be carcinogens . . . . “

One limitation enumerated in the first report was that, “Science and society have not arrived at a final consensus on the definition of carcinogen either in human populations or in experimental animals. ” The report, by including chemicals and industrial processes already classified by IARC, has sidestepped dealing with the question of what is a carcinogen. A definition for “carcinogen” remains elusive unless it is given in the context of a particular methodology.

Table 30.—Chemicals and Industrial Processes Evaluated for Human Carcinogenicity by the International Agency for Research on Cancer (IARC)

Chemicals and processes judged carcinogenic *for* humans

	Degree of evidence	
	Humans	Experimental animals
4-aminobiphenyl	Sufficient	Sufficient
Arsenic and certain arsenic compounds	Sufficient	Inadequate
Asbestos	Sufficient	Sufficient
Manufacture of auramine	Sufficient	Nonapplicable
Benzene	Sufficient	Inadequate
Benzidine	Sufficient	Sufficient
N,N-bis (2-chloroethyl)-2-naphthylamine (chlornaphazine)	Sufficient	Limited
Bis(chloromethyl)ether and technical grade chloromethyl methyl ether	Sufficient	Sufficient
Chromium and certain chromium compounds	Sufficient	Sufficient
Diethylstilboestrol (DES)	Sufficient	Sufficient
Underground hematite mining	Sufficient	Nonapplicable
Manufacture of isopropyl alcohol by the strong acid process	Sufficient	Not applicable
Melphalan	Sufficient	Sufficient
Mustard gas	Sufficient	Limited
2-naphthylamine	Sufficient	Sufficient
Nickel refining	Sufficient	Nonapplicable
Soots, tars and mineral oils	Sufficient	Sufficient
Vinyl chloride	Sufficient	Sufficient

Chemicals and processes judged probably carcinogenic for humans

Group A: Chemicals and processes with "higher degrees of evidence."

	Degree of evidence	
	Humans	Experimental animals
Aflatoxins	Limited	Sufficient
Cadmium and certain cadmium compounds	Limited	Sufficient
Chlorambucil	Limited	Sufficient
Cyclophosphamide	Limited	Sufficient
Nickel and certain nickel compounds	Limited	Sufficient
Tris(1-aziridinyl)phosphine sulphide (thiotepa)	Limited	Sufficient

Group B: Chemicals and processes with "lower degrees of evidence."

	Degree of evidence	
	Humans	Experimental animals
Acrylonitrile	Limited	Sufficient
Amitrole (aminotriazole)	Inadequate	Sufficient
Auramine	Limited	Limited
Beryllium and certain beryllium compounds	Limited	Sufficient
Carbon tetrachloride	Inadequate	Sufficient
Dimethylcarbamoyl chloride	Inadequate	Sufficient
Dimethylsulphate	Inadequate	Sufficient
Ethylene oxide	Limited	Inadequate
Iron dextran	Inadequate	Sufficient
Oxymetholone	Limited	No data
Phenacetin	Limited	Limited
Polychlorinated biphenyls	Inadequate	Sufficient

Chemicals and processes that could not be classified as to their *carcinogenicity for* humans

	Degree of evidence	
	Humans	Experimental animals
Chloramphenicol	Inadequate	No data
Chlordane/heptachlor	Inadequate	Limited
Chloroprene	Inadequate	Inadequate

Table 30.—Chemicals and Industrial Processes Evaluated for Human Carcinogenicity by the International Agency for Research on Cancer (IARC) (Continued)

Chemicals and processes that could not be classified as to *their* carcinogenicity *for* humans—continued	Degree of evidence	
	Humans	Experimental animals
Dichlorodiphenyl trichloroethane (DDT)	Inadequate	Limited
Dieldrin	Inadequate	Limited
Epichlorohydrin	Inadequate	Limited
Hematite	Inadequate	Negative
Hexachlorocyclohexane (technical grade HCH/lindane)	Inadequate	Limited
Isoniazid	Inadequate	Limited
Isopropyl oils	Inadequate	Inadequate
Lead and certain lead compounds	Inadequate	Sufficient (for some soluble salts)
Phenobarbitone	Limited	Limited
N-phenyl-2-naphthylamine	Inadequate	Inadequate
Phenytoin	Limited	Limited
Reserpine	Inadequate	Inadequate
Styrene	Inadequate	Limited
Trichloroethylene	Inadequate	Limited
Tris(aziridinyl)-para-benzoquinone (triaziquone)	Inadequate	Limited

Chemicals and processes for which human data are available, but which were not considered by the IARC Working Group

Ortho-and para-dichlorobenzene
Dichlorobenzidine
Phenylbutazone
2,3,7,8-tetrachlorodiberzo-para-dioxin (TCDD, the ''dioxin" of Agent Orange)
Ortho-and para-toluidine
Vinylidene chloride

SOURCE: Office of Technology Assessment, adapted from IARC (185)

SOURCES OF EPIDEMIOLOGICALLY USEFUL DATA

Three major types of information are useful in assessing the carcinogenic risk of a substance: l) health status of exposed and unexposed populations; 2) exposure data; and 3) physical, chemical, and biological properties of the substance. Information related to each of these categories can come from a variety of sources and can be used in different ways. Testing the substance generates information about its potential hazard, but information about its distribution in the environment and any impacts on human health are necessary to describe its human risk

Health Status Information

DHHS is primarily responsible for administering health data collection, storage, and analysis projects. An overview of existing DHHS programs and other departments' health data collection activities can be found in *Selected Topics in Federal Health Statistics (283).*

National Center for Health Statistics

The National Center for Health Statistics (NCHS) located within the DHHS Office of Health, Research, Statistics, and Technology, was established to collect and disseminate data on the health of Americans. Since 1960, it has played **a** major role in the development of national health statistics policy and programs. The NCHS Division of Vital Statistics collects information on natality, mortality, marriage, and divorce from the individual states and regions. (See ch.2 for a discussion of cancer mortality and incidence statistics.) In addition to vital statistics, NCHS conducts several general-purpose surveys that provide statistics about the health status of the U.S. population. Additional information can be obtained from *Data Systems* of *the National Center for Health Statistics (253).*

Health Interview Survey (HIS) .–HIS is the principal source of information on the health of the civilian noninstitutionalized population of the United States. Initiated in 1957, interviews are conducted each week in a probability sample of households to provide data on a range of health measures, including the incidence of illness and accidental injuries, the prevalence of diseases and impairments, the extent of disability, and the use of health care services. Each year, approximately 40,000 households containing about 120,000 persons are sampled.

HIS collects information only about conditions which respondents are willing to report. The basic questionnaires are similar from year to year but supplemental questions may be added. In 1978 and 1979 questions on smoking were added, but these were discontinued in 1980.

The NCHS Study of Costs of Environment-Related Health Effects, mandated by Public Law 95-623, may use HIS as a data source **(179).**

Health and Nutrition Examination Survey (HANES).–HANES, initiated in 1970, is a modification and expansion of the earlier Health Examination Survey (HES). These surveys collect and use data from interviews and physical examinations to estimate the prevalence of chronic diseases, establish physiological standards for various tests, determine the nutritional status of the population, and assess exposure levels to certain environmental substances. The sampling techniques employed provide representative national data. Two surveys, HANES I (1971-75) and HANES II (1976-79) have been conducted. Both surveys examined approximately 20,000 persons.

HANES is the most extensive national assessment of health and nutritional status of the American people. The nutritional component of HANES includes: information on dietary intake; data from hematologic and biochemical tests; body measurements; and chemical examination for various signs of high risks of nutritional deficiency. Preliminary findings from the HANES II pesticide monitoring program have found an apparent rise in tissue levels of DDT and PCBs. The implications of the observed levels are uncertain.

HANES surveys might become valuable sources of information for cancer epidemiology if sufficient resources were available. Because of its representative nature, aggregate data from the survey can be used to represent “normal” or background levels. For example, white cell count levels determined in HANES I were used for comparative purposes in an epidemiologic study of laboratory workers exposed to suspected toxic chemicals. HANES II contains certain information about dietary intake of substances which have been associated with a lower risk of cancer, vitamins A and C, and substances such as fats which are associated with higher risks.

HANES might be linked with other health data systems, such as the National Death Index (see below) to facilitate assessment of whether particular exposure levels or certain nutritional statuses were associated with cancer mortality. NCHS, with its HANES capabilities, has been asked to participate in studies near Love Canal, and to evaluate the health status of certain high-risk industry groups. It was unable to do so because of limited resources.

The NCHS overall monitoring survey budget for fiscal year 1981 is $28 million. This is a $3 million increase over 1980 and includes $1.1 million for a special HANES study which will focus on Hispanics in selected areas of the United States. The study is designed to describe the health and nutritional status of the Mexican-American, Puerto Rican-American and Cuban-American populations. Studies of specific groups are necessary to acquire data in sufficient detail to describe subgroups of the population which differ from the “average.” General national surveys such as HANES I and II produce data about the “average” citizen by sampling groups in proportion to their representation in the total population, and this often results in too small a sample size to be useful for identifiable smaller groups.

As examples of data useful for cancer studies, the HANES Hispanic study will determine:

- name, date of birth, and social security number recorded in machine readable form for subsequent use with the National Death Index;
- history of toxic substance exposure;
- nutritional status including dietary interviews, serum vitamin A levels, prevalence of vitamin C deficiencies; and
- the quantity and frequency of alcohol consumption.

Hospital Discharge Survey .—This survey was established in 1964 to provide representative statistics for the U.S. population discharged from short-term hospitals. The survey collects information on the characteristics of patients, the lengths of stay, diagnosis and surgical operations, and patterns of use of care. Completion of each medical abstract form is estimated to take approximately 5 minutes. Only short-stay hospitals with six or more beds and with an average length of stay of less than 30 days are included in the sample (177).

Vital Statistics Followback Surveys. —NCHS-conducted mortality "followback" surveys have provided information on possible relations between environmental and lifestyle factors and death from cancer. Information is sought about decedents through inquiries addressed to those providing information for the death certificate, such as the medical certifiers, funeral directors, and family members. These surveys are an efficient means to augment the routinely reported information contained in the vital records.

The efficiency of the followback approach for eliciting additional information about deaths is related to the relative rareness of death as an event in the U.S. population. About 1 percent of the population dies annually; and cancer-related deaths are reported for about 20 percent of this 1 percent, or a total of 0.2 percent of the total population. The followback approach permits sampling directly from the file of those death certificates of interest in order to supplement existing information.

Mortality followback surveys were conducted in the United States annually from 1961 through 1968. They have since been discontinued due to inadequate resources, including personnel.

National Death Index (NDI).—Deaths in the United States are registered by the States or other death registration areas (e.g., District of Columbia). The records are transmitted on microfilms or on magnetic tape to NCHS for compilation. Historically, because there was no integration of records for the country as a whole, no mechanism had existed at the national level to determine if a person had died. In 1981, after several years of planning and preparation, the NDI will be placed into operation to serve that purpose. The NDI will code deaths that occurred in 1979 and each year thereafter. Although there has been discussion of coding deaths that occurred before 1979, no plans are now in place to do so.

NDI, administered by NCHS, is designed to provide medical and health researchers with probable fact of death, the death certificate number, and the location of the death certificate, when supplied with a minimum set of identifiers (generally the person's name and social security number or date of birth). The researcher then may contact the registration area where the possible match has occurred to obtain the death certificate or the required information.

NDI will be of immediate use in ongoing long-term studies which include mortality. Beebe **(23)** described the NDI as the most important recent advance in making vital statistics accessible to researchers. NCI's Surveillance, Epidemiology, and End Results (SEER) program plans to use the NDI to determine deaths of all persons in the SEER registries. This should reduce the number of people lost to followup by SEER and provide better information about survival. Currently, deaths of people who moved or cease to participate are not always recorded by SEER.

National Cancer Institute

NCI, 1 of 11 research organizations of NIH, receives more than one-fifth of all Federal health research funds (283). NCI operates the SEER program, which provides cancer incidence data

on approximately 10 percent of the population. Additional information on the SEER program can be found in chapter 2.

Centers for **Disease Control**

The overall mission of the Centers for Disease Control (CDC) (56) is "to prevent unnecessary illness and death and to enhance the health of the American people." CDC serves as a focus for DHHS efforts in the areas of disease prevention and control, environmental health, health promotion, and health education.

NIOSH, located within CDC, assists OSHA in establishing workplace health standards. Between 1972 and 1974, NIOSH conducted the National Occupational Hazard Survey (NOHS) to provide estimates of the proportion of employees exposed to potential health hazards in various industries. NOHS estimates of exposure are often used in assessing risks from occupational carcinogens. NIOSH periodically conducts studies to identify the health effects of particular industrial processes and to determine the health experience of selected employee populations. The National Surveillance Network, which is operated by NIOSH, collects data from State safety and health inspection programs.

Social Security Administration

The Social Security Administration (SSA) collects information on economic and demographic data in administering the social security system. SSA makes available an annual l-percent continuous work-history sample (CWHS) to outside users which provides information about employment, migration, and earning status. Six different types of files, all of which contain sex, race, and age data, are available to outside users. For purposes of confidentiality, the employee and employer identification are included in scrambled form. The usefulness of the CWHS for epidemiologic studies is limited by privacy constraints of access and other characteristics, e.g., only wages UP to the taxable maximum are reported.

SSA has recently initiated efforts to amass a 10-percent sample of the work force because of the Office of Management and Budget's (OMB) need for better estimates of intercensal population. This file would constitute the most detailed information on the structure of the labor force so that employment distributions by sex, race, age, wages, and wage changes, work force participation, industry, and regional migration patterns could be analyzed systematically.

The Disability Insurance Fund, managed by SSA, contains information regarding benefit computation and actions related to employee entitlement. SSA routinely prepares reports regarding specific disease entities and has published characteristics of workers disabled by cancer (283).

Exposure Data

The principal data deficiencies for assessing cancer risks are inadequate information about exposures and lifestyle characteristics. Since cancer has a long latent period, relating cancer in today's population to particular exposures might require information from 20 or more years ago. Even when information was collected, records may have been destroyed before they became useful in cancer epidemiology studies.

As the lead agency for regulating chemicals, EPA administers numerous exposure-monitoring programs. Several studies have been critical of EPA's monitoring data collection efforts, and as a result, EPA established a Deputy Assistant Administrators Committee to review and make recommendations regarding agency monitoring and information management activities (113). Three of the major conclusions found by the committee are:

1. A considerable quantity of collected ambient environmental information has not been analyzed or presented to top management.
2. The most serious problem found was the lack of consistent, integrated information on toxic and hazardous pollutants.
3. There is little coordination between EPA offices focusing on the same area. In addition, there is a lack of comparability and sharing of these data.

Exposure data in the workplace are limited even though a number of cancer causing agents **have** been identified in the occupational environment. NIOSH's National Occupational Hazard Survey collects exposure data in a sample of industries, and OSHA requires monitoring for 6 of the **20** substances regulated because of carcinogenicity.

Death certificates generally include questions on the usual occupation and industry or business of the decedent, However, those questions are not always answered and there is uncertainty about the accuracy of the information that is provided. NCHS, along with NIOSH, is currently assessing the feasibility of using and improving occupational descriptors on death certificates as a surrogate for exposure information. Approximately a dozen States now code the usual occupation of the decedent, and about half of these also code the reported usual industry or business of the decedent. One State, Wisconsin, now publishes such tabulated information in their annual public health report. Other States, with support from NIOSH, have executed special studies on occupational mortality based on data reported on the standard death certificate **(45,236).**

In England, information reported on the death certificate has been used as a basis for occupational mortality analyses every 10 years since 1851 (with the exception of the war year, 1941). In the United States, a study of this type was conducted in 1950. It involved coding over 300,000 death certificates for the occupation and industry of males aged 20 to 64. The information was used in conjunction with the decennial census information for that year to produce measures of relative mortality risk associated with occupation and industry of decedents (148,149,150,151,196).

The National Human Adipose Tissue Monitoring Program and HIS and HANES, which are described above, are the principal mechanisms for monitoring levels of toxics in the body. This information is used not only to determine normal baseline levels but also to identify populations which may beat high risk.

This assessment did not concentrate on monitoring methods and programs, but the general impression is that data collection efforts are incomplete and that many generate data of limited usefulness. For instance, measurement techniques are not always specified for collected exposure data and ignorance of the sensitivity of the instrument used makes it difficult to compare measurements from different times and sites. Furthermore, a nondetectable measurement does not necessarily indicate that a substance is not present, and may mean only that the instrument was not sufficiently sensitive to measure it. Such negative readings are often not reported, and when they are, they may be misleading. The efforts of organizations such as the EPA Committee (113) mentioned above to improve collection and analysis of monitoring data might be encouraged.

Chemical Information Systems

The lack of toxicological information about many substances and concern over perceived toxic substance problems prompted Congress to enact more than two dozen statutes dealing with toxics. Many of the statutes delegate information-gathering functions to Federal regulatory and research agencies.

Toxic Substances Control Act —TSCA

In 1976, Congress passed TSCA to strengthen the ability of the Federal Government to accumulate information on potentially hazardous chemicals and better enable the Federal Government to protect the public from toxic substances. TSCA required the establishment of several new programs at EPA and a completely new infrastructure had to be organized. This necessitated recruiting an Assistant Administrator and placed a large burden on a small staff. Subsequently, the programs **have** been slow to get off the ground (141). Recruitment has continually lagged behind authorized staff ceilings which may be inadequate to meet expected program needs. EPA estimated that approximately **1,500** people were needed in fiscal year 1979: 382 permanent positions were authorized, 313 were actually filled.

New Chemicals.—One of TSCA's primary objectives—as stated in the opening paragraph of the Committee on Interstate and Foreign Commerce Report (165), is to provide "for the evaluation of the hazard-causing potential of new chemicals before commercial production begins." TSCA requires manufacturers and importers to notify EPA at least 90 days prior to the manufacture or import of a new chemical substance by submitting a premanufacture notice (PMN). Along with notification, manufacturers are to provide specific information about the new chemical, including any test data which relate to the effects of the substance on human health or the environment. "New" is synonymous with not being listed on the inventory of existing chemicals and subject to TSCA's authority. In May 1979, EPA (105) issued a statement of interim policy covering submission and review of PMNs. Final PMN rules have yet to be issued. In the 2-year period, April 1979 to March 12, 1981, EPA received 488 PMNs, and about 800 are anticipated in fiscal year 1982.

TSCA was not the first Federal law to require a review of new chemicals entering the market. FIFRA and the Food, Drug, and Cosmetics Act, have registration/certification provisions for pesticides and pharmaceuticals respectively. Unlike those laws, TSCA requires neither licensing nor registration, but only notification of intent to manufacture. As mentioned, the PMN must contain any test data available to the submitter, but EPA cannot require testing of a new chemical substance just because it is "new." The act places the burden of proof on EPA to demonstrate that the information available to EPA:

> . . . is insufficient to permit a reasoned evaluation of the health and environmental effects of a chemical substance . . . and . . . [that] in the absence of sufficient information . . . [the substance] may present an unreasonable risk or will be produced in substantial quantities, and there may be significant or substantial exposure.

If such a finding is made, EPA may issue an order under section 5(e) to prohibit or limit the manufacture, processing, distribution, use, or disposal of the chemical. EPA has proposed two such orders and each time the company withdrew its notice and decided not to manufacture. On two other occasions notices were withdrawn when the companies learned orders to require more information were in preparation.

Lack of regulatory action on a new chemical by EPA does not imply that the substance is "safe" or has been "approved." However, it does grant the manufacturer the right to produce and use the chemical as desired, subject to any other regulations that may be applicable. Under section 5(a)(2) EPA can issue a "significant new use rule" (SNUR) for a chemical when there is concern that a specific use of the substance, other than those proposed in the PMN, may pose an unreasonable risk. Issuance of an SNUR requires that persons must notify EPA 90 days prior to manufacture or processing of a substance for a use subject to a SNUR. Through March 1981, one chemical specific SNUR had been proposed and EPA was considering SNURs on more than 40 chemicals. Once a substance is in production, EPA can require testing under TSCA section 4 and/or submission of human health and environmental monitoring data under a TSCA section 8 rule (see Existing ***Chemicals).***

EPA recently published a policy statement describing a recommended list of premanufacture tests for new chemical substances (114). The tests are identical to those under consideration by the Organization for Economic Cooperation and Development (OECD). OECD considers that its base set of tests would generate the minimum amount of information normally sufficient to perform an initial hazard assessment of a chemical (see table 31). EPA is recommending that flexibility be used by manufacturers to tailor this "base set" of tests for the particular chemical substance and intended uses. Use of this base set is voluntary due to EPA's lack of statutory authority to require testing of new chemical substances. The estimated costs of this base set, should all tests be employed, range from $53,000 to $67,850.

EPA has reported that no toxicity data were submitted in 60 percent of the first 199 notices received (141). In fact, 25 percent of the notices contained no data on physical or chemical properties. No chronic test data have yet to be

Table 31.—EPA's Recommended Base Set of Data To Be Included in Premanufacturing Notices

Type of data	Estimated cost
Physical/chemical data	
Data about 11 characteristics	$ 3,800
Acute toxicity data	
Acute oral toxicity	2,000
Acute dermal toxicity	2,800
Acute inhalation toxicity	3,300
Skin irritation	700
Skin sensitization	3,200-6,700
Eye irritation (for chemicals showing no skin irritation)	450
Repeated dose toxicity data	
14-28 day-repeated dose test(s) using probable route(s) of human exposure.	10,200-12,800
Mutagenicity ***data***	
Gene (point) mutation data	1,350
Chromosomal aberration data	18,000
Ecotoxicity **data**	
Data about killing of three lower organisms	**4,100**
Degradation/accumulation data	***3,100-11,850***

SOURCE Office of Technology Assessment, adapted from EPA (114)

presented to EPA in a PMN. In cases where toxicity data are given they are usually limited. The then-Assistant Administrator for Toxic Substances has indicated (192) that the lack of test data:

> . . . has placed an extraordinary burden on EPA's limited resources . . . . Furthermore, we [EPA] believe that our objective will not be achieved until industry assumes more of the burden of generating adequate risk information and assessing the risk of its products.

Unless additional information is received with the notices, EPA's reviews will be "based upon a fundamental lack of information and data. This in turn means that our information will be highly uncertain" **(192).** In order to evaluate a chemical's carcinogenic potential, EPA staff have had to rely on structure activity relationships and mutagenicity data when available.

Existing **Chemicals,** –Sections 4 and 8 of TSCA relate directly to the issue of acquiring adequate information for assessing the carcinogenic risk of existing chemicals.

Section 4 grants EPA the authority to require industry testing of a potentially harmful chemical if available information is insufficient for a reasoned evaluation of risk and if the substance: 1) may present an unreasonable risk or 2) result in substantial or significant exposure. TSCA established ITC to recommend chemicals for testing under section 4; ITC and EPA responses to it are described above. To require industry testing, EPA must demonstrate that available information is insufficient to conclude that the substance "presents an unreasonable risk, " yet supports the finding that the chemical "may present an unreasonable risk. " If available information were sufficient to show that the chemical presents an unreasonable risk, EPA could regulate the substance under section 6 of TSCA. One difficulty EPA faces in using this testing authority is how to define "may present an unreasonable risk. "

Section 8 of TSCA required EPA to compile by November 1977 an inventory of chemical substances manufactured in the United States. An initial inventory was published about 2-1/2 years late in June 1979 and updated in July 1980. Information originally requested for the inventory was limited and EPA has proposed rules requiring additional information for certain substances. In February 1980, rules requiring exposure information on approximately 2,300 substances were proposed (108). Additional information-gathering rules are scheduled for proposal in 1981.

Section 8(c) of TSCA also requires manufacturers, processors, and distributors to notify EPA of information which reasonably supports the conclusion that a substance "presents a substantial risk" of injury to health or the environment. In addition, Section 8(c) requires the maintenance of records indicating "significant adverse reactions" alleged to have been caused by a chemical substance. Allegation by employees are to be retained by industry for 30 years and all other allegations for 5 years. These records are to be submitted to EPA upon request, and EPA is investigating means of establishing an automatic reporting system whenever a certain number of allegations are received in a 12-month period for the same substance, process, or discharge **(109).** Final rules to implement the significant adverse reaction reporting requirement are expected in 1981.

Chemical Substances Information Network (CSIN)

In February 1978, under mandate of TSCA, EPA and CEQ established the Toxic Substances Strategy Committee (TSSC) to facilitate interagency coordination of chemical information collection, dissemination, and classification. The Departments of Health and Human Services, Agriculture, Commerce, Defense, Energy, Interior, State, and Transportation, and OSHA, CPSC, and NSF participate in TSSC. IRLG and the DHHS Committee to Coordinate Environmental and Related Programs are also members. TSSC has formed a number of committees for special tasks, and the data committee recommended the development of a broad-based network of data systems—CSIN. CSIN was adopted by TSSC, and if sufficient resources and personnel are committed, it should go a long way toward the goal of providing convenient access to information about chemicals. Its master file will contain all information collected under TSCA; a subfile, stripped of confidential data about trade secrets, will be available for public use. CSIN will identify about 1 million chemicals and for each one provide selected research and test data, references in the toxicologic and biomedical literature, and information about regulations that pertain to the chemical (345). The system is still in the developmental stages, although some aspects were expected to be operational in 1980 and the entire system is to be completed within a decade.

CSIN will not be a single new system; rather it will incorporate several systems already in use. To facilitate locating information about a single substance in more than one system, EPA is developing an unambiguous identification number system, which was another recommendation of TSSC's data committee. In this way, information about structure, chemical and physical properties, production volume, uses, application, distribution, and toxicity, now stored in different systems, can be linked together. Such systems do not necessarily contain information about human health effects, but they can be used in combination with health information systems.

Collection and Coordination of Exposure and Health Data

Congress has responded to concern about the collection and availability of data for assessing environmental health risks. It has mandated commissions and studies directed at improving data collection, storage, and dissemination.

The Health Services Research, Health Statistics, and Medical Libraries Act of **1974,** Public Law 93-353, mandated the U.S. National Committee on Vital and Health Statistics to assist and advise PHS with statistical problems bearing on health and the delivery of health services which are of national interest. A committee recommendation was important to the establishment of the NDI.

The 95th Congress passed two acts which included sections pertaining to data collection for assessing and reducing cancer risks. The Clean Air Act Amendments of 1977 (Public Law 95-95) established the Task Force on Environmental Cancer and Heart and Lung Disease, to focus efforts by EPA and various branches of DHHS on issues relating to these diseases. The task force is composed of representatives from EPA, NCI, NHLBI, NIOSH, NIEHS, NCHS, CDC, and FDA (340). The task force was directed to recommend, among other things (340):

> . . . a comprehensive research program to determine and quantify the relationships between environmental pollution and human cancer . . . [and] . . . recommend research and such other measures as may be appropriate to prevent or reduce the incidence of environmentally related cancer . . . .

Initial efforts were focused on defining problems, categorizing relevant research programs, and exchanging information among member agencies and other appropriate groups. Activities related to prevention and reduction of environmental risks are planned to begin during 1981.

The task force established several project groups to address specific areas of interest. Of particular relevance to this assessment is the Project Group on Exposure and Metabolic

Mechanisms, established in May 1979. The group is examining the interrelationship of exposure to a toxicant and body uptake, metabolism, and affected target organs. It is hoped that this information will increase the ability of researchers to predict a chemical's potential toxicity and establish which symptoms maybe associated with trace chemical levels in the body.

The Project Group on Standardization of Measurements and Tests is primarily concerned with obtaining reliable, comparable data on environmental and disease measurements. The group has focused on "potential contributions to the state of the art of environmental and occupational monitoring and testing that would complement, rather than duplicate or overlap, the efforts of individuals or agencies" (340). They identified two problem areas common to task force agencies:

1. There is a need for better resource allocation to optimize the quality of data since agencies and laboratories charged with monitoring activities typically have limited resources.
2. Researchers currently have limited means of assessing the relationship and validity of published environmental monitoring data.

The Project Group on Standardization of Measurements and Tests expects to develop recommendations in these areas.

The Health Services Research, Health Statistics, and Health Care Technology Act of 1978 (Public Law 95-623) mandated the Secretary of DHHS to initiate several efforts relating to the impact of the environment on health. He is to develop a plan for the collection and coordination of statistical and epidemiological data on the effects of the environment on health and prepare guidelines for the collection, compilation, analysis, publication, and distribution of this information. The law also authorized the Secretary of DHHS to "consult with and take into consideration any recommendations of the Task Force" in developing the plan.

NCHS (252) recently published *Environmental Health: A Plan for Collecting and Coordinating Statistical and Epidemiologic Data,* which reviews information currently available from Federal data collection systems. A series of recommendations are made to Congress for correcting gaps and deficiencies in environmental health data systems. These recommendations include the need for priority setting in new data collection efforts, interagency coordination in the environmental data collection process, assurance of the quality of data, and linking data on the environment and health.

NCHS **(252)** identified 64 ongoing systems in 18 agencies that gather environmental health-related data. At least two-thirds of these collect either information on cancer incidence/mortality or data on cancer risk factors. Most of the data collection systems identified are designed to:

- collect health-related data,
- measure environmental pollutants and individual exposures,
- test specific interrelationships, or
- link data on the environment and health.

Linking systems that collect different kinds of data appear to have the greatest utility for assessing associations of environment and health. For example, the Upgrade system, which is concerned with water quality and health is a joint effort of CEQ, EPA, NIOSH, and NCHS, and integrates data from the Bureau of the Census, mortality data from NCHS, and water quality data from EPA and the U.S. Geological Survey.

OTA encountered an example of the often-voiced complaint that data are not in a form easily accessible and useful to researchers. National mortality data were necessary to carry out the analyses reported in chapter 2, but those data were available in computer readable form only for the years since 1968. OTA had to computerize the data back to 1933 to carry out the analyses. The OTA computer tapes of cancer mortality data will be made available to researchers. These data include deaths by age, race, and sex for selected cancers, for each of the years 1933 through 1978. All of the data on these tapes are consistent with those available on paper from NCHS.

Not all informational systems release data for public use and some that are released are of limited usefulness. Data are often aggregated by geographical regions (e.g., by county *or* State), which precludes detailed analysis of exposures on health. Individual identifiers are frequently deleted because of privacy and confidentiality constraints. This unfortunately hinders intensive specific epidemiologic investigations which are facilitated by matching an individual's exposures and lifestyle characteristics with health status. Agencies with individually identified records can sometimes conduct followup studies to obtain additional information for investigation, e.g., NCHS recently initiated a followup study, 10 years after the survey, of participants in its HANES I survey, The followup will investigate current health status of participants as it can be related to previously collected data.

Collection of useful epidemiologic information may be indirectly affected by the Paperwork Reduction Act of 1980, which passed during the last days of the 96th Congress. The act is intended to improve use of existing data collection systems and to reduce the Federal paperwork burden placed on individuals, businesses and governments by 25 percent within 3 years. The act empowers the Director of OMB to review and approve Federal agency information collection requests. Various medical and epidemiologic research groups, including NIH, expressed concern that Federal research into disease prevention could be impeded by the Paperwork Reduction Act and advocated the exemption of biomedical and epidemiologic research **(64),** The Senate Committee on Governmental Affairs decided against this exemption because it determined that the act would not interfere with disease prevention research.

In addition to reducing the paperwork burden, the act directs the sharing of information among agencies to the extent authorized by law and the establishment of a Federal Information Locator System. This system will contain a description of all information collected by the Federal Government, and directions for obtaining the information by all agencies and the public. Successful implementation of this system should enhance the quality of information available for epidemiologic research.

Government Records and Record Linkage

Government records contain a wealth of information about individuals that could be of great value to researchers looking for associations between exposures and disease states. The SSA, Internal Revenue Service (IRS), Veterans' Administration (VA), the Bureau of the Census, and NCHS are organizations with extensive data collections. For the most part, details about individuals are not readily available because of legal and institutional barriers. Most critically, it is difficult to obtain records of the same person from two or *more* sources. This can be a major obstacle because one record rarely contains all desired information. For example, to do a study relating mortality to occupational exposures, one could extract "occupation" from the IRS record, "employer" and "industry" from information collected by the SSA, and cause of death from the death certificate filed in the State in which the death occurred. The NDI, which is described above, makes it easier to locate death certificates.

Individuals in several Government agencies have been promoting a Linked Administrative Statistical Sample, a project designed to bring together records of IRS, SSA, and NCHS (200). The project planners aim to provide an improved data base for mortality research by compiling statistical information from the participating agencies on a sample of individuals. The starting point will be the l-percent CWHS (see above). This effort has great importance as a pilot for future projects to study cancer mortality, particularly contributions of occupational exposures.

Records from different agencies have been linked before on an ad hoc basis, e.g., VA has cooperated in several studies (23), but the broader scale now proposed has brought some basic issues to the fore. The most fundamental barrier to researchers acquiring data from more than one file is that which is erected to preserve

privacy and confidentiality. Restrictions have been tightened considerably in recent years by the Privacy Act of 1974 and the 1976 amendments to the Internal Revenue Code. Although epidemiologists have not been implicated, well-publicized breaches of confidentiality and privacy have engendered suspicion among legislators and the public, about possible misuse of easily accessible files. MacMahon (217) has summarized the epidemiologist's point of view:

> To determine cancer risks among persons exposed to particular environmental factors, we need to be able to link information relating to the same individual at different times in his life and to determine whether an individual exposed in the past is now dead or alive and in what state of health . . . . Maximum confidentiality means minimum epidemiologic information and minimal effectiveness in identifying new cancer hazards. In my opinion, we are well beyond the point at which concern for confidentiality seriously impairs the extraction of valuable knowledge, even from routinely collected information. Working as an epidemiologist, one comes to recognize the readiness with which most people, patients or nonpatients, will supply even sensitive information if they believe the cause is reasonable. Somehow, the issue of confidentiality becomes more difficult when it is institutionalized or politicized. We must attempt to convince the public's representatives that a reasonable balance can be achieved.

Other linkages, and the means to accomplish them, have been suggested. One current proposal involves drawing a sample of several million individuals from those people who filled out the 1980 long census form for future matching to the NDI. This would facilitate matching cause of death with personal characteristics reported in the census. Another proposal has been made to add cause of death information to the SSA's l-percent CWHS. However, since collection of cause of death data is not part of the mission of SSA, it would presumably require money from outside SSA for implementation.

The cited desirabilit, of making records more available for research purposes runs into conflicts with society's intention of protecting the individual's privacy. The linking of individual records between agencies would allow a person with access to the system to obtain most or all Government-held information about any individual. The potential for abuse is apparent. At the same time, linkage, in the hands of a biomedical researcher, might quickly provide information about behaviors, exposures, and health that could be obtained only with great difficulty in any other system.

The Workgroup on Records and Privacy of the Interagency Task Force on the Health Effects of Ionizing Radiation (182a) has addressed the problems arising from the conflict between access to records for research and the right to privacy. It suggested changes in the Privacy Act, the Tax Reform Act, and commented on pending bills (in 1979) that would have permitted **access** to records under tightly controlled conditions. The workgroup's report is an excellent starting point for discussion of this complicated subject and provides possible directions for, Federal efforts.

SUMMARY

Interest in cancer prevention and acceptance of the idea that some substances cause cancer have spurred developments in methods to determine which substances are associated with cancer.

Epidemiology has played an important role in identifying both carcinogenic substances and exposures which are associated with cancer although the causative agent may not be known. When available, epidemiologic data about cancer risks are the most convincing. At the same time, identifying a carcinogen on the basis of human disease and death means that other testing methods failed to identify the agent before human illness resulted from it.

The most important laboratory method for determining carcinogencit y is the long-term bioassay in laboratory animals, generally rats and

mice. The 1960's saw marked increased interest in the test, and NCI has played a major role in designing appropriate test methods. NCI developed a large-scale bioassay program for testing, evaluating, and documenting the carcinogenicity of environmental chemicals. It has also supported the IARC program to review information on environmental carcinogens and publish findings in its authoritative monograph series.

While the bioassay has been improved and is widely accepted as an appropriate method to identify carcinogens, it is expensive (each one costs $400,000 to $1 million) and time-consuming (each takes from a minimum of more than 2 years to a more realistic 5 years). Because of those costs and limited long-term bioassay capacity, other tests are being developed.

The short-term tests measure biological effects other than carcinogenicity (often mutagenicity) in bacteria, yeast, cultured mammalian cells, or in intact lower organisms. Major international efforts have been completed and others are ongoing to validate the ability of these tests to identify correctly carcinogens and noncarcinogens. Some tests perform well in the validation experiments, and it is expected that they will play an increasingly important role in identifying carcinogens.

Deciding to test a chemical in a short-term test or a long-term bioassay involves studying its structure. Some classes of chemicals are more likely to be carcinogenic than others, some are present in high concentrations in the environment, and some are viewed with suspicion for one reason or another. Based on such information a chemical may be selected for testing. NTP is developing a tier-testing scheme which first assays a chemical in a number of short-term tests. The chemicals that appear most risky as a result of short-term tests will be accorded priority for further testing in long-term bioassays. This NTP program promises to improve both the use of information from short-term tests and the usefulness of the bioassay program.

The establishment of NTP, which appears to be moving toward an ordered, careful development of methods for testing and interpretation of results, and IRLG's appearing as a single voice for Federal regulatory procedures and decisions promise further improvements. Additionally, IARC's distinguishing between noncarcinogens and carcinogens and ordering among carcinogens on the basis of test results is an important step in increasing the usefulness of the results.

Not included in testing systems, but discussed here are already useful and potentially more useful data collection systems. Several of these systems were mandated by Congress to collect information about exposures and health status. Currently, inadequate resources and coordination may be hampering the performance of the systems. Another major source of information, the administrative data systems, were not developed as health information resources. However, both SSA and IRS collect some information which might be of great value for epidemiologic studies. Perhaps the most pressing need in adapting the administrative data systems for these uses is more consultation between epidemiologists and the data systems experts.

80-481 O - 81 - 11