REGULATORY TOXICOLOGY AND PHARMACOLOGY 7, 427–443 (1987)

Toxicology Then and Now

BERNARD L. OSER

Bernard L. Oser Associates, Inc., 2 Bay Club Drive, Bayside, New York 11360

Received May 27, 1987

A review of the history of the evolution of the science of toxicology from the original concepts of Paracelsus through the early development of analytical chemistry and its contributions to the detection of toxic substances in foods and drugs as these have led to modern regulatory rules for public protection is presented. The legal actions taken to protect against adulteration of food prior to the early steps by the U.S. Department of Agriculture that concluded with the passage of the 1906 Food and Drug Act are systematically documented. The history is reviewed of the Food and Drug Administration's role in the use of animal toxicity studies to develop reasonable criteria for safety of foods and drugs for man. Modern concepts of molecular distribution, metabolism, and excretion of substances in the animal body are discussed as these impinge on the so-called "protection index." The legal and often litigious controversies over the claimed carcinogenicity of chemical substances is documented with comments on the Delaney dilemma and the role of *in vitro* tests in toxicology. The review concludes with a discussion of the hazards of use of stochastic mathematical models to assess carcinogenicity and suggests that the criteria employed in the report of the Scientific Committee of the Food Safety Council are properly structured to give a contemporary evaluation of all the currently available data. References give documentation of events over the past 200 years that explain the present state of toxicology as a discipline. © 1987 Academic Press, Inc.

Major events in the development and application of chemistry and biology led to the emergence of toxicology as a major multidisciplinary science. It was not until the late 18th century that specific toxic agents could be identified and characterized with respect to exposure, dose, mode of action, and toxicity. Toxicology earned the recognition it has achieved only as recently as the present century and in an environment where socioindustrial conditions place greater emphasis on the safe uses of chemical substances than on their toxicity.

The most pertinent science that had to be developed before any of this could have been accomplished was chemistry. The 17th and 18th centuries marked the dawn of this science. The period from the mid-18th to the mid-19th century witnessed practical applications of chemistry in such diverse areas as the use of fossil fuels, the adaptation of coal gas for street lighting, the extraction of natural components from plant, animal, and mineral sources, and the production of chemical products for specific industrial and medicinal purposes. The development of analytical chemistry for the

0273-2300/87 $3.00
Copyright © 1987 by Academic Press, Inc.
All rights of reproduction in any form reserved.

detection of adulteration of foods began around 1800 before this "nefarious practice" was proscribed by law as a hazard to health and life.

WAS IST GIFT?

It seems appropriate to discuss the much-disputed concept of "poison." A logical way to begin is by reference to Philippus Theophrastus Aureolus Bombastus von Hohenheim, "called Paracelsus" (1493–1541). He was a Swiss physician and alchemist who was what today might be called an off-beat physician ("everything from charlatan to genius"). His ideas of therapeutic remedies were in conflict with those of most practitioners of his time; for example, he engaged in the "reckless" use of metal salts such as mercury compounds for the treatment of syphilis, which were commonly recognized to be poisons. In a defiant response to his critics (1,2), he contended that the only difference between a drug and "non-poison" was the dose. It was in this diatribe that he pronounced the familiar dictum which, in free translation, reads: "All things are poison and nothing is without poison. Only the dose makes the thing a non-poison." It is significant that the rarely quoted sentence immediately following reads: "For example, every food and every drink, if taken beyond its dose, is poison. . . ."

The doctrine that a substance can be ineffective at a low dose, therapeutically useful at a higher dose and, above that, actually toxic, served to motivate much of the early research. Dictionary definitions are rather simplistic in a modern context when chemicals are evaluated not only for harmful or lethal properties but for their safe ("non-poisonous") uses as additives to food. "Dose," under these conditions, must take into account not merely the amount of a substance ingested per unit weight of food, per serving, per day, or per kilogram body weight, but also the concentration in the food or beverage, the daily intake pattern, the chronicity (frequency and duration) of ingestion, supplementation or interaction with other added or natural components of the diet, the age and state of health of the host, human homogeneity, etc. "Dose" as used by Paracelsus, may accurately relate to an individual user under a given set of conditions, but not to a large segment of a population under widely varying nutritional, environmental, or occupational conditions.

THE DETECTION OF POISONS

A high point in the history of chemistry was achieved by Frederick Accum, born in Buckebourg, Germany, in 1776. At the age of 24, he emigrated to London where he found employment as a pharmacist's assistant. His zeal and industry attracted the attention of several well-known chemists, one of whom, William Nicholson, started the *Journal of Natural Philosophy, Chemistry and the Arts* in 1797 to which many eminent scientists, including Cavendish, Rumford, and Priestley, made contributions.

In 1801, Accum joined the Royal Institution as a "chemical operator" for Sir Humphrey Davy. He also established his own commercial laboratory which, for the next two decades, offered consulting services and sold chemical supplies and apparatus. Accum published profusely on a great variety of technical subjects among which was a series of papers titled "An Attempt to Discover the Genuineness and Purity of

TABLE 1

TOXIC ADULTERANTS FOUND IN FOODS, CA. 1775–1875[a]

Chemical	Food
Alum	Flour, bread, wine
Arsenic (trioxide)	Sugar
Boric acid	Butter, milk
Calcium sulfate (gypsum)	Bread, milk
Copper arsenite (Scheele's Green)	Confectionery
Copper salts (bluestone)	Water, pickles, peas, alcoholic beverages, tea
Formaldehyde	Milk
Iron oxide	Tea
Lead	Water (cisterns)
Lead acetate ("sugar of lead")	Milk, brandy
Lead chromate	Candy, tea
Lead oxide (red lead)	Cheese, mustard, annatto
Mercuric sulfide (vermilion)	Cayenne pepper, candy
Picrotoxin (cocculus indicus)	Beer
Salicyclic acid	Canned vegetables
Sulfur dioxide[b]	Dried fruits, canned vegetables
Sulfuric acid (oil of vitriol)	Vinegar, gin
Turpentine	Gin

[a] Excludes economic adulterants such as spent or foreign leaves in tea, roasted acorns in coffee, chalk in milk, ground bones in bread, etc.

[b] Generally recognized as safe (GRAS) in these foods at currently used levels.

Drugs and Medicinal Preparations" which focused on the prevalence of adulteration of these products.

That Accum was one of the most prolific food technologists of his time is indicated by his treatises published in 1820–1821 on Brewing Wine from Native Fruits, Good and Wholesome Bread, and Culinary Chemistry, and a Treatise on Adulterations of Food and Culinary Poisons (3). The latter dealt with the earliest application of the infant science of analytical chemistry to the detection of adulteration in various categories of products of the milling, baking, and fermentation industries. His treatises and popular lectures on the subject sought to condemn this widespread practice, more because of its fraudulent aspect than the health hazard to which it subjected consumers. Too little was known about toxicology and how to observe the chronic effects of low levels of exposure to chemicals. Legal proscriptions were virtually nonexistent except on a local scale and even then were applied only to specific foods like tea, coffee, milk, and bread (Table 1).

Accum's popularity was short-lived. The merchants whose names were cited as the suppliers of the foods which he found to be adulterated were, of course, not enchanted by his whistle-blowing revelations. These malefactors sought and found revenge which brought a sudden tragic end to Accum's career as a food scientist. Under a search warrant, they ransacked his premises to discover some incriminating evidence against him. His "scandalous crime" consisted in cutting pages from books he had borrowed from the library of the Royal Institution. It led to his arrest and conviction on the charge of mutilating books. This was before the days of reprints or copying

machines, when clipping extracts from books and journals was probably not a rare or reprehensible offense.

Accum felt disgraced even by his scientific friends who, with few exceptions, failed to come to his defense. Despite his reputation and eminence as a chemist, he soon found himself without friends or business. He was released on bail, but while his appeal for a retrial was pending, he fell into deep despair over the hatred and abuse to which he was subjected. So, after 30 prosperous years in England, rather than stand the humiliation of another trial, he returned to Germany where he abandoned his career as a food chemist.

There can be little doubt that Accum was the first to apply analytical chemistry to the detection of toxic adulterants in foods and drugs. Notwithstanding his brush with the law, I would vote for including him among the pioneers who contributed to laying the foundations of toxicology.

A LAW IS BORN

Adulteration of food continued with practically no legislative abatement for the next 30 years (4, 5). Although various authors picked up the cudgels in support of the campaign against food adulteration it did not begin to gain force until Dr. Arthur Hill Hassall's work as a microscopist came to the notice of the editor of *The Lancet,* Mr. Thomas Wakley. He established under Hassall's direction The Lancet Analytical Sanitary Commission. A 659-page book titled *Food and Its Adulterations* (6, 7) was published in 1855 covering "solids and fluids consumed by all classes of the public," viz., cereals, sugar, spices, coffee, tea, confectionery, bread, butter, preserves, water, and even tobacco and tobacco products. Hassall revealed the sources, labeling, and advertising of many of the samples he examined. An important feature was 159 drawings depicting the microscopic appearance of genuine and adulterated products. The Lancet Commission greatly influenced Parliament to pass the first English Food and Drug Adulteration Act whose centenary was celebrated in 1960. It proved rather ineffective, however, until an 1875 law prohibited the sale of adulterated food (8).

Modern innovations, such as the use of paper and plastic containers, vacuum and aseptic packaging, freeze drying, and radiation, have provided many more options to the food technologist, but canning, in its many aspects, still holds sway as the major process for the packaging of fruits, vegetables, beverages, soups, etc. The new technologies have generated toxicological problems of their own due to contact of foods with metallic surfaces, e.g., copper, iron, tin, lead (from solder), and migrants from can lacquers, resins, polymeric plastics, rolling oils (from metal foils), etc.

THE UNITED STATES FOLLOWS SUIT

The Civil War precipitated the establishment of gunpowder and munitions plants which became the predecessors of our giant producers of chemicals and drugs. During the latter half of the 1800s, perhaps as many as 200 papers were published indicating the prevalence of chemical adulteration of foods and drugs. Some states and municipalities passed prohibitory laws but there was no effective legislation at the Federal level aimed at curtailing this practice.

In 1895 Dr. Harvey Wiley was appointed head of the Bureau of Chemistry of the U.S. Department of Agriculture. One of his first acts was to initiate a campaign toward the enactment of comprehensive food and drug legislation. He recognized the value of demonstrating experimentally the injurious effects of chemical adulteration, although the principal target of his attack was illustrated by his statements that "there is more reason to be concerned with the economic adulteration of our food supply than with its safety," adding that "it is absurd in logic to attempt to prove anything harmless." As if to underscore the focus of his concern, he later wrote ". . . injury to public health, in my opinion, is the least important question in the subject of food adulteration, and it is the one which should be considered last of all. The real evil of food adulteration is deception of the consumer" (9).

Considering the rudimentary state of toxicological experimentation in animals (and perhaps in recognition of the aphorism that "the proper study of mankind is man"), Wiley undertook to investigate the effect on human subjects of diets containing high levels (up to several percent) of chemical preservatives in use at the time. He recruited 12 young male employees of the Department of Agriculture as "volunteers" for a feeding study. The group became known derisively as Wiley's "Poison Squad." Tests were started in 1902 and ran for a period of 5 years. Substances investigated were benzoic, boric, salicylic, and sulfurous acids, borax, and formaldehyde. The procedure was designed as follows: (1) a "fore period" to establish the amount of food needed to maintain constant body weight; (2) a 15- to 20-day "preservative period" to record subjective effects and disturbances in metabolism as reflected by food intake and analyses of excreta; and (3) an "after period" to allow any adverse effects to return to the normal state. No tests for specific organ changes were made and the conclusions drawn rested largely on the subjects' own judgement of their physical or physiological responses.

Detailed results of these studies were reported in Parts 1 and 2 of Bulletin 84 of the Department of Agriculture (10). Typical of the conclusions was the statement regarding boric acid and borax that "when continuously administered in small doses for a long period or when given in large quantities for a short period, (they) create disturbances of appetite, of digestion, and of health." Salicylic acid was found to be "a harmful substance of less virulence than has been generally supposed" but "even in small quantities (it) exerts a depressing and harmful effect upon the digestion and health and general metabolic activities of the body."

Wiley's war against economic adulteration continued with renewed force and remained paramount over issues of safety. For instance, he condemned the use of cheap substitutes such as saccharin for sugar and oleomargarine for butter. But despite the work of Dr. Bernard Hesse, an FDA consultant, on food colors and Wiley's own Poison Squad, the alleged injurious effects of chemicals played a secondary role in the administration of the new law. Experimental toxicology had not advanced sufficiently to serve as a scientifically and legally acceptable means of safety evaluation.

The first U.S. Food and Drug Act was passed in 1906 after almost a decade of disputatious Congressional debate. Enactment was finally accelerated only after public reaction to two events. One was the "embalmed beef" scandal of the Spanish American War and the other was the publication of "The Jungle," Upton Sinclair's savage indictment of the working conditions and filth in Chicago's stockyards.

During the years before and immediately after passage of "Wiley's Law" toxicity tests on animals and human volunteers were performed, mainly on food preserva-

tives and coal-tar colors, but they were quite primitive in both design and execution. However, they emphasized and gave impetus to the need for developing more reliable scientific methodology for safety evaluation.

There was considerable dissatisfaction within Congress and other government agencies over Wiley's extremely conservative administrative policies. His readiness to ban substances like sodium benzoate, sulfur dioxide, and saccharin, despite the advice of scientists on a referee board, culminated in so much controversy that after several years contemplation he resigned in March 1912.

TOXICOLOGY LEAPS FORWARD

A seminal factor in the birth of the modern science of nutrition, and consequently of toxicology, was the recognition that diets composed only of protein, carbohydrate, fat, and minerals could not sustain life. The work of such pioneers as Lunin in Russia, Hopkins and Drummond in England, and Osborne, Mendel, and McCollum in the United States led to the "vitamine" hypothesis, the term coined by Casimir Funk in 1912 (11) for the organic micronutrients essential for "complete" diets. The discovery over the next few decades of the entire complement of vitamins recognized today made it possible to design basal diets that would permit test species to reproduce normally and live out their full life cycle, a sine qua non of chronic toxicological feeding studies. It made it possible to distinguish between dietary inadequacy and toxicity per se.

The 1938 Amendment to the Food and Drug Act provided the impetus to consolidate and update toxicological procedures for the purpose of safety evaluation. Nothing approaching standardization of methodology had been achieved, but with considerable foresight, FDA had been developing procedures for internal use.

It would be remiss not to pay tribute to Dr. Arnold J. Lehman, Dr. Arthur A. Nelson, Dr. O. Garth Fitzhugh, and their colleagues in the Food and Drug Administration during the two decades preceding the 1958 Food Additives Amendment. This group of highly qualified pharmacologists, chemists, and pathologists deserves credit not only for developing the basic procedures for safety evaluation but also for conducting a great many subchronic feeding tests on typical samples of substances used in food production, processing, and packaging, as well as chronic studies of pesticides whose structure suggested a relationship to known carcinogens.

The contributions of the Division of Pharmacology were reported in the September 1949 and October 1955 issues of the *Food Drug Cosmetic Law Journal.* In response to pressure which had mounted from food and drug companies for advice and guidance as to the specifics of animal studies to satisfy regulatory requirements, FDA's "Appraisal of the Safety of Chemicals in Foods, Drugs and Cosmetics" was published in 1959 by the Association of Food and Drug Officials of the United States. While not declared as "official," it provided useful guidelines for the benefit of scientists in the service of regulated industries. It covered, albeit rather briefly, procedures recommended for acute, subacute, and chronic oral toxicity (Table 2) with particular reference to food additives and included chapters on dermal toxicity, carcinogenic screening, pathology, statistics, and the evaluation of drugs and endocrines (12). Many toxicologists felt that it was premature for FDA to "freeze" testing procedures employed in this still developing science by publication in official media such as the *Federal Register.*

TABLE 2

FDA 1959 PROTOCOL FOR ORAL TOXICITY TESTS

	Subacute		Chronic	
	Rats	Dogs	Rats	Dogs
Test groups	Control, plus 5 Dosages	Control, plus 3 Dosages	Control, plus 3	Control, plus 3
No. of animals (minimum per group)	10 M and 10 F	2 M and 2 F	25 M and 25 F	3 M and 3 F
Administration, route of	Oral	Oral	Oral	Oral
Time of observation	3 months	3 months	2 years	2 years
Clinical lab. tests	As indicated—see Biochemical Section		As indicated—see Biochemical Section	
Gross pathology	Autopsies on all animals		Autopsies on all animals	
Organ weights	Liver, kidney, spleen, heart, and testes or ovaries (thyroid and adrenals in special cases)		Liver, kidney, spleen, heart, and testes (thyroid and adrenals, after fixing, when suspected)	
Microscopic pathology	See Pathology Section		See Pathology Section	
Evaluation	The study shall establish (1) A maximum tolerated dosage; (2) The biological activity of the compound; (3) An estimate of the "no effect" dosage.		The study shall establish (1) A "no-effect" level; (2) A toxic level.	

The FDA group reported that emulsifying agents, preservatives, detergents, sanitizers, and chelating agents in use at that time were safe. Of the 19 certifiable FD&C colors, none showed evidence of carcinogenicity, but 3 (FD&C Orange Nos. 1 and 2 and Red No. 32) were dropped on the basis of toxicological rather than carcinogenic evidence. Of the remainder, 9 were subsequently banned principally because of neoplastic lesions in rats fed high lifetime doses. Of five classes of flavoring agents (ethers, esters, phenols, aldehydes, and alcohols) tested at levels of 0.1, 0.25, and 1.00% "none showed toxicity of sufficient degree to warrant against its use in foods." Among the pesticides tested for carcinogenicity—"probably more is read into this aspect than is warranted"—those that caused definite morphological alterations or were structurally akin to known carcinogens were especially scrutinized.

The premise employed for converting animal data to safe tolerances for man was basically the internationally recognized safety factor approach which is often misinterpreted and therefore justifies quotation from the current regulation (21 CFR 170.22): "Except where evidence is submitted which justifies use of a different safety factor, a safety factor in applying animal experimentation data to man of 100 to 1, will be used; that is, a food additive for use by man will not be granted a tolerance that will exceed 1/100th of the maximum amount demonstrated to be without harm to experimental animals." An obvious limitation of the 100-fold safety factor is that it cannot be applied in situations where the human dietary level is one or more percent, or in fact, anywhere approaching that level. For example, several vitamins would be toxic if ingested daily at 100 times their normal requirement. On the other

hand, as Dr. Lehman pointed out, "trivial" contamination could occur with substances of known or unknown toxicity in which case safety factors many fold greater than 100:1 could be justified. He wrote, "to assemble a list of possible trivial contaminants and to assign them a concentration value which can be regarded as 'de minimis' for all possible conditions would probably not serve a useful purpose," thus acknowledging the need for judging each case on its merits (13).

MUCH ADO ABOUT ZERO

It may be recalled that the adulteration section of the Food and Drug Act of 1906 prohibited the addition of a poisonous substance to food, with the exception that if it were not added (i.e., present naturally) it would be allowable provided the quantity as ordinarily consumed would not be injurious. There was little incentive to develop toxicological procedures since there were few instances where tolerances for added poisons were requested. With the advent of antibiotics, DDT, and many other synthetic organic pesticides during World War II, the problem of the safety of residues rose up in full force. Many weaknesses and loopholes in the law were recognized and corrected in the Food, Drug and Cosmetic Act of 1938. It relaxed the prohibition of poisonous or deleterious substances in food if they were required in production or unavoidable by good manufacturing practice. In such cases, the administrative agency was authorized to establish tolerances for residues in agricultural and animal products and in processed foods derived therefrom.

Analytical chemists met the challenge of developing procedures of continuously increasing sensitivity. However, it became obvious that the "zero tolerance" goal could be approached, but never completely attained. As a consequence of the realization that a "no residue" policy was literally untenable and impracticable, the subject eventually came up for judicial determination (*Monsanto vs. FDA*). FDA contended that the terms of the Food Additives Amendment, particularly with respect to the Delaney clause, allowed for no administrative flexibility. The court ruled otherwise (14).

It is of interest to recall that as long ago as 1902, Dr. Wiley, the father of our food and drug law, wrote in Bulletin 84 "It is only proper to give 'argument de minimis' full consideration and not to brush it aside as illogical and irrevelant. It is evident that any attempt to determine experimentally the effect of extremely minute quantities of any preservative, even when used continuously, would not be likely to lead to any definite result." However, he argued, by example, that any added burden on the kidneys, in circumstances where large quantities of nitrogenous foods are consumed, would be unjustified if indiscriminate use of multiple de minimis levels of preservatives were permitted. He suggested that "de minimis non curat lex" may be construed as "the law does not excuse the use of injurious substances because they may be present in small quantities"—an example of twisted logic since substances are not injurious when de minimis. I once proposed a corollary to this phrase, Dr. Wiley to the contrary notwithstanding: "de minimis non curat toxicologia" (15).

Eventually, FDA began to accept Dr. Lehman's concept of "trivial contamination" and devise policies for dealing with such situations. Three proposals, designated "constituents," "sensitivity of method," and "de minimis" have recently been defined by FDA and, in fact, used for regulatory purposes. My preference is for the latter concept

TABLE 3

CLASSIFICATION BY PRESUMABLE RISK SHOWING "PROTECTION INDEX" (PI) AND CATEGORIES OF SAFETY[a]

	PI[b] and category of safety for a daily per capita intake (in mg) of							
Class	$<10^{-5}$	10^{-5}–10^{-4}	10^{-4}–10^{-3}	10^{-3}–10^{-2}	10^{-2}–10^{-1}	0.1–1.0	1.0–10	>10
I	>250,000,000 A	>25,000,000 A	>2,500,000 A	>250,000 A	>25,000 B	>2500 C	>250 C	<250 D
II	>25,000,000 A	>2,500,000 A	>250,000 A	>25,000 B	>2,500 C	>250 C	>25 D	<25 D
III	(500,000) A	(50,000) B	(5,000) C	(500) C	(50) D	(?) D	(?) D	(?) D

[a] Class I and A least toxic; Class III and D most toxic. Cramer, *et al.* (16).

[b] $\text{PI} = \dfrac{\text{lowest no-effect level for class (mg/kg body wt)} \times 50 \text{ (kg body wt)}}{\text{maximum of intake range (mg) in column}}$.

since constituents, as commonly understood, include normal as well as added components, and the sensitivity of methods changes with progress in analytical methodology and instrumentation.

PROGNOSTICATION OF TOXIC HAZARD

Molecular disposition and more specifically detoxication mechanisms of many categories of organic chemicals have been the subject of investigations since the early decades of this century. Among the well-recognized biotransformation processes involved are hydrolysis, conjugation, oxidation, reduction, aromatization, ring opening, decarboxylation, dealkylation, and deamination. These reactions are mediated primarily by enzymes in the gastrointestinal tract or liver and to a lesser extent in various target organs and cells.

Reasonable assumptions of metabolic fate can be made taking into account limiting factors, for instance, the relative concentrations of substrate (toxicant) and the availability and rates of secretion of specific enzymes and reactants involved in conjugations (e.g., with glucuronic acid or glycine) or other biochemical processes.

A major contribution to the application of structure–activity relationships to the prediction of potential hazard is the "decision tree approach" of Cramer, *et al.* (16). This is accomplished by means of a branched "tree" whereby answers to a consecutive series of "yes or no" questions lead to the classification of compounds into four groups (A to D) of presumed toxicity, viz., negligible, low, intermediate, and high. Each class may then be divided into subgroups based on exposure levels. This system serves the dual purpose of classifying substances on the basis of their relative predicted toxicity and, when taken into account with estimated degrees of exposure (expressed on a scale ranging from 10^{-5} to >10 mg per capita per day), yields "Protection Indexes," which can be translated into the need for any further study (see Table 3) (16).

This system has been successfully applied to the classification of over 1500 flavoring substances, both naturally occurring and synthetic (17).

More recently a somewhat similar basis for predicting toxicity has been incorporated in the FDA Red Book (18, 19, 20), low, intermediate, and high "probable toxicity" being assigned to categories according to their functional groups, e.g., saturated aliphatic alcohols and complex heterocycles to Categories A and C, respectively. Together with daily exposure (expressed as parts per million dietary levels calculated from "adjusted poundage" or disappearance into the food supply) three classes of concern or minimum testing levels are derived.

It hardly needs to be emphasized that observed or expected dietary exposure levels are an essential feature of safety evaluation. They are generally calculated from food consumption data and the concentration of the additive used in specific foods or food classes but independently of the amounts of the same or closely related substances in the diet.

Exposure or intake levels of food additives have been calculated in different ways by various agencies and still remain one of the most difficult and controversial aspects of safety evaluation. Estimates are usually grossly exaggerated in terms of the frequency, duration, categories of food, and patterns of use.

A recent proposal which deserves serious consideration has been to evaluate the use of additives on an incremental basis, i.e., the extent to which they enhance the intake over the natural presence of the same or closely related substances in the diet (21) rather than on the basis of the daily intake of the additive per se (18).

TOXICOLOGY UPDATED

Without a detailed account of the development of animal testing procedures, the condensed outlines of the tests proposed in 1959 for subacute and chronic oral toxicity (Table 2) may be contrasted with current procedures described in the 1982 FDA publication of the Bureau of Foods dubbed "the Red Book" (18). It recommended increasing the number and range of dose levels in subchronic tests, increasing the number of rats or mice to 20 from 10 of each sex per group, and adding tests for blood clotting time, electrolytes, and enzymes to the series of routine clinical tests. The major change in the chronic studies was to augment the size of groups to 50 animals per sex and the organs examined grossly and histopathologically from liver, kidney, spleen, heart, and gonads plus any showing gross pathology at necropsy to about 40 organs and tissues as shown in Table 2.

Subchronic feeding studies are believed to reveal most types of toxic manifestations except those associated with fertility, reproduction, teratogenicity, and carcinogenicity. (The 90-day test could be extended to 9 or 10 months and yield significant data in all these respects, excluding carcinogenicity.)

Any evidence of such aberrations or structural configurations related to those of known carcinogens would naturally lead to the requirement for chronic (so-called "lifetime") studies in rats, mice, or both species, irrespective of whether results of short-term tests included a no-effect level (NEL) which, by the way, I later more accurately designated a "no *observed* adverse effect level" (NOAEL) (15, 22). Notwithstanding the number and variety of mutagenicity tests constituting a "battery" and the admitted absence of complete correlation with chronic bioassays of "known" carcinogens, the tendency for regulatory purposes has been to regard a single positive result to outweigh any number of negatives. Under the Delaney clause, a finding of

neoplastic lesions in a test group in excess of the controls, even if only hyperplastic or benign, has been regarded as predictive of malignancy, i.e., cancer.

This approach is tantamount to adoption of the "no-threshold" or "one molecule" hypothesis of carcinogenesis which is beyond experimental proof. It has led to an influx of mathematical extrapolations, based on putative theories for the multifactorial mechanism of carcinogenesis. Estimation of safe doses of single substances may vary over a millionfold range depending on the adequacy of the toxicological data base and the choice of statistical model.

In short, aside from more and larger test groups, the principal procedural modifications over the past 25 years have varied with respect to the number, type and frequency of clinical, hematological, and histopathological examinations, as well as the interpretive criteria.

Also contained in the Red Book are guidelines for long-term toxicity studies in dogs, for teratological studies in the rat, mouse, hamster, and rabbit, three-generation reproduction studies in rodents, and metabolism tests "using the same species and strain as those being used for most other toxicological studies on the same chemical." In the latter respect, it is emphasized that "differing pathways may invalidate a given species as a proper toxicological model for a specific compound" (18).

This sketchy account of the evolution of toxicology as employed for establishing safety with "reasonable certainty" is presented as background for a consideration of the adequacy, validity, and degree of assurance needed to implement the regulatory process.

JUDGES, PLEASE NOTE

That toxicology is not an exact science has become a cliche among its practitioners. Notwithstanding the detailed descriptions of experimental conditions and procedures promulgated under the heading of "guidelines," the options and leeways open to the investigator are so numerous that no two studies of a given substance in laboratories of "equal competence" ever turn out exactly the same. In fact, uncontrollable or uncontrolled differences are such that it is not unusual for repeated studies for safety to lead to opposite conclusions and in such cases the usual practice—"in the interest of prudence"—is to render negative judgements. As a result, many useful substances have been sacrificed on the altar of uncertainty (or have been so threatened), such as DDT, acrylonitrile, benzyl acetate, nitrilotriacetate, saccharin, cyclamates, and FD&C Red No. 3.

In our judicial system a defendant is declared either guilty or not guilty, rather than innocent or not innocent. Safety, like innocence, is theoretically a negative concept, beyond absolute proof. Hence, it is important to recognize the options in the design and execution of studies for safety which militate against absolute certainty. They include the choice of species, strain, and stock of experimental animals; their pre-, peri-, and postnatal nutrition; housing and environmental conditions; age and weight at the start of dosing; the frequency, duration, and route of administration of test doses; the choice, number, and reliability of physical, behavioral, biochemical, and hematological tests; metabolic and pharmacokinetic tests; and the extent and degree of postmortem pathological observations. With particular reference to the laboratory animal commonly assumed by toxicologists to be an appropriate surrogate for the human species, see my paper originally titled "Man is Not a Big Rat" (23).

Procedural variations account, at least in part, for the lack of interlaboratory reproducibility of bioassays and for the failure to correlate bioassays with predictions of carcinogenicity from mutagenicity tests. It should also be noted that many investigations for carcinogenicity deviate from proposed guidelines, for example, by the use of only one or two dosage levels below the assumed maximum tolerated dose. Nevertheless such bioassays have been used for comparing risk assessment via different statistical models. Decisions based on extrapolating no observed effect levels in animals to the actual conditions of use by human populations at assumed degrees of acceptability may convey the impression of precision which, in fact, is no greater than that of the toxicological data base.

Problems such as these often lead to the courts, which recalls the remarks of an eminent jurist that ". . . courts lack the technical competence to resolve scientific controversies; they lack the popular mandate and accountability to make the critical value choices that this kind of regulation requires. The court's role is rather to monitor the agency's decision-making process—to stand outside both the expert and the political debate and to assure that all the issues are thoroughly ventilated" (24).

THE DELANEY DILEMMA

Probably the most controversial aspect of our present food law from both the scientific and administrative standpoints has been the Delaney clause (which FDA had originally contended was superfluous). It makes no reference to the quantitative aspects of toxicological testing, i.e., the concentration or amount of the dose or the duration of the exposure period. It created the unprovable concepts of "no residue," "no threshold," and "absolute safety." The intent of Congress was not clear because of the vagueness of the phrase "found to induce cancer" whereby FDA assumes it can ban promoters as well as initiators even if they are not per se carcinogenic.

Many normal components of the diet, such as certain fats, spices, and products formed during cooking, can be shown to induce cancer in animals. Literal construction of the Delaney clause contained no provision to permit scientific discretion in the evaluation of the experimental evidence. An alternative provision (viz., "or if it is found, after tests which are appropriate for the evaluation of the safety of food additives, to induce cancer in man or animals") anticipated the possibility that parenteral dosage might be deemed appropriate but such routes are not considered suitable for toxicological testing of food additives.

The Delaney clause contributed to the popular perception that all chemical additives pose a carcinogenic risk in contrast with the ubiquity of naturally occurring substances (25) which can be demonstrated to be carcinogenic by rigorous animal tests but are assumed by law to be safe "if the quantity of such substance . . . does not ordinarily render it injurious to health" (26). In this connection the paper of Ames on dietary carcinogens is of particular interest (27).

Under the Delaney clause, nonmalignant (hyperplastic or benign) lesions are regarded administratively as precancerous and combined with malignant neoplasms for statistical analysis. This has led to an invasion of sophisticated mathematical proposals for the assessment of data based on the no-threshold and other putative theories for the multifactorial process of carcinogenicity. Estimates of safe levels of individual substances by these methods may vary over a millionfold range.

The cost in terms of time, dollars, and manpower in the futile attempt to "prove" noncarcinogenicity has been considered an obstacle to progress in research aimed toward enhancement or improvement of the food supply. It is safe to say that there has been no convincing epidemiological evidence that the ingestion of any food additive has been causally related to an increase in the incidence of human cancer.

MUTAGENICITY TO THE RESCUE?

The introduction by Ames of *in vitro* tests for mutagenicity (28, 29), using *Salmonella typhimurium* as the test organism, gave promise of a much-desired alternative to "lifetime" feeding studies in rodents. Subsequent modifications involving activation of the culture media with a liver homogenate (S9) resulted in better correlation of the Ames test with carcinogenicity as determined by animal studies, but a number of extensive studies indicated that there were approximately 10 to 15% false positives or false negatives. Nevertheless, this spurred microbiologists and genetic toxicologists on to further studies of *in vitro* and *in vivo* genetic tests based on diverse mechanisms for the attack on cellular DNA, the expectation being that a multiplicity of methods would compensate for the limitations of each.

When the idea of a "battery" of tests for mutagenicity was proposed, as many as nine different procedures were considered, based on the use of microorganisms, mammalian cells, insects (*Drosophila*), and rodents. Few laboratories had the facilities or experience to perform all these tests. However, in time, they neither individually nor cumulatively proved to be as reliable as predictors of carcinogenicity nor as inexpensive as was originally hoped. Recommendations were eventually made to limit the battery to three tests, preferably the Ames *Salmonella typhimurium* assays using four strains of organisms with S9 activation; the *in vitro* forward mutation assay using either *in vitro* cultured lymphoma cells or a host-mediated test in mice; and a dominant lethal assay in rats or mice. Notwithstanding this multiple approach to assessing potential mutagenicity, its consistency falls short of replacing chronic bioassays in rodents.

However, to the extent that an appropriate battery of tests yields uniformly positive or negative results, mutagenicity data, together with assumed metabolic fate based on molecular configurations, provide corroborative support for predicting potential carcinogenicity and particularly for setting priorities.

A PREGNANT PROBLEM

The thalidomide disaster in the early 1960s stimulated a revival of interest in reproductive and developmental toxicology. Despite extensive studies in rats, the teratological property of this tranquilizer was not discovered until after its permitted use in Europe. This emphasized the profound importance of recognizing species differences when extrapolating either positive or negative toxicological findings between species. Many investigations have established that differences exist not only in metabolism, i.e., the absorption, distribution, accumulation, and excretion of chemical substances, but in pharmacokinetics, i.e., the relative rates at which these processes take place.

Conventional studies of reproduction in experimental animals revealed the existence of congenital malformations at birth. Teratological procedures are specially designed to relate the time of exposure or administration of a drug to the period of organogenesis, that is, to distinguish between embryotoxicity and fetotoxicity.

Despite the voluminous literature that has accumulated on this subject in the last quarter century, a recent Interagency Regulatory Liaison Group Workshop on Reproductive Toxicity Risk Assessment acknowledged that "the paucity of information for both experimental animals and humans does not allow the correlation of reproductive effects between species" (31). Nevertheless, the report added "In the absence of specific data to the contrary, adverse effects in experimental animals should be presumed to indicate a potential risk to human reproduction."

Procedures for single and multigeneration studies of reproduction and teratology are described in the Bureau of Foods' Red Book. It should be emphasized, however, that at the levels of exposure to food additives in the human diet, in contrast with high-dose experiments in rodents, such adverse effects occur rarely, if at all.

COLORS REVISITED

The 1938 Act permitted FDA certification of coal-tar colors only if "harmless and suitable" for use. This entailed testing each production batch for purity (i.e., chemical identity) and freedom from toxic residues. Any coloring agent that was derived, or assumed to be capable of being derived, from a coal-tar source was considered to be subject to certification.

During the period 1953 to 1958, FDA experienced one of the most contentious problems in its entire history (32). It involved intramural and extramural controversy, public hearings, and subsequently litigation over the certification or delisting of food, drug, and cosmetic colors, with special reference to Orange No. 1 and Red No. 32 used to enhance the color of the skins of oranges. Experiments on rats and dogs had shown these certified colors to be toxic and suspiciously carcinogenic.

This protracted dispute wound up in the U.S. Supreme Court which ruled that "where a coal-tar color is not harmless [per se] it is not to be certified; if it is not certified, it is not to be used at all" regardless of the food product in or on which it is used.

In 1955, an episode occurred in which children developed diarrhea after consuming Halloween "trick or treat" candy containing excessive amounts of a certified color (FD&C Orange No. 1). In view of the fact that "harmless" was interpreted in the absolute sense, FDA delisted that color for food use.

In 1958, a pharmaceutical company (Hoffmann–LaRoche) reported the first industrial synthesis of β-carotene, the precursor of vitamin A which bears no structural relation to coal-tar colors and therefore was subject to certification. Nevertheless, it turned out, upon inquiry, that for regulatory purposes FDA regarded it as capable of being derived from coal-tar products. Public hearings were held but failed to reach a definitive conclusion on this issue. FDA then proposed amending the statute, to which Congress agreed by enacting the Color Additives Amendments of 1962. It permitted provisional listing of existent food, drug, and cosmetic colors, both natural and synthetic, subject to future determination of their permanent status.

QUANTITATIVE RISK ASSESSMENT

In the administration of the laws governing safety, some of the terminology has undergone implicit, if not explicit, redefinition broadening certain basic concepts within the legal framework as perceived by the regulatory agency. For example, the distinction between toxicity (the capability of a substance per se to cause harm) and hazard (the probability that harm might result from its use or misuse) was originally well recognized. Now, however, "risk" is confusingly equated with both toxicity and hazard. The words "safety" and "safe" were not specifically expected to be included in amended legislation to take into account reasonable, rather than absolute, certainty under the conditions of intended use.

"Risk is conceived to involve the evaluation of toxicity by means of animal tests, since retrospective human experience is generally not sufficiently defined or controlled and, at best, is relatively meager." Risk assessment is the precursor to the determination of risk acceptability which is now considered to be a sociopolitical responsibility rather than a decision-making process within the domain of qualified scientists.

Against this background, it is pertinent to consider the effort to quantify risk assessment so as to make safety decisions objective rather than judgmental. This has involved the introduction of statistical procedures based on hypothesis and conjecture, particularly in the area of carcinogenesis where the pathogenetic mechanisms are diverse, highly speculative, or unknown. Depending on the choice of mathematical model, estimates of a "safe" dose at a given (ineluctably arbitrary) acceptable risk level (e.g., a lifetime risk of cancer of one-in-one-million persons per year) may vary over a millionfold range. Thus it is illusory if not fallacious to assume that quantitative risk assessment permits more precise or objective evaluation of safety than is possible by applying scientific judgment based on experience to conventional procedures for the estimation of "acceptable daily intakes."

QUESTIONS REGARDING ASSESSMENTS

The traditional method of applying a safety factor to the observed no-effect level in animals has its limitation as described above. In its place, it is now proposed to make "quantitative risk assessments" by means of mathematical models based on putative mechanisms for carcinogenesis which take into account the slope of the dose–response curve. The stochastic models most often used in this connection are the probit, logit, one-hit, multistage, Weibull, and Armitage–Doll models.

Published data (33) however reveal that statisticians are often too generous in their assumption of the validity or adequacy of the biological data base and often plot curves from too few points in the zone of graded response. The net result has been that depending on the choice of models, estimates of the no-effect level from a single study in rats vary widely when extrapolated to acceptable risk levels for man of 10^{-4} to 10^{-8} (but usually and arbitrarily set at 10^{-6}). The final result of these mathematical machinations is actually no more precise than that which can be achieved through the safety factor route based on a multilevel assay with a factor determined by the slope of the dose–response curve and the severity of the effect.

A current trend among toxicologists is to deprecate the routine nature of bioassays designed to meet the needs for regulatory control. It is said that regulatory judgments of "safe" or "not safe" must rest on, or at least are facilitated by, numbers. However, biologically oriented scientists are very much aware of the problem of assigning numbers or scores to the diverse and often imponderable observations that collectively are expected to lead to a definite conclusion. This difficulty is illustrated by the descriptive terms applied by the National Toxicology Programs to characterize the result of a carcinogenicity study, viz., by classifying the evidence as "clear," "some," "equivocal," "none," or "inadequate." How this is to be applied for regulatory purposes, and by what agency, is not "clear." I think I can conclude this discussion of quantitative risk assessment in no better way than by quoting from the final chapter of the report of the Food Safety Council's Scientific Committee (34–40):

> Human risk assessment is a very inexact exercise, based largely upon theoretical assumptions concerning interspecies extrapolations. (34)
>
> . . . The qualitative nature of the biological data used to predict human risk must be taken into account. (35)
>
> . . . the equating of the "no observed effect level" and the "no effect level" is statistically naive. . . . (36)
>
> . . . inadequate or questionable animal test data should preclude the use of any mathematical models, since the results are likely to be totally misleading. (37)
>
> . . . mathematics cannot rise above its source. (38)
>
> Because the mechanisms of carcinogenesis are not understood, even those mathematical models drawing on biological theory cannot claim to be universally correct. (39)
>
> . . . blind use of low dose linear extrapolations or the conservative one-hit model . . . will often unnecessarily preclude the use of substances which are of significant societal or industrial value . . . and appears to us to be scientifically indefensible. . . . (40)

REFERENCES

1. HOLMSTEDT, B., AND LILJESTRAND, G. (1981). *Readings in Pharmacology,* p. 29. Raven Press, New York.
2. DEICHMANN, W. B., HENSCHLER, D., HOLMSTEDT, B., AND KEIL, G. (1986). What is there that is not poison? A study of the Third Defense by Paracelsus. *Arch. Toxicol.* **58,** 207–213.
3. ACCUM, F. (1820). A Treatise on Adulterations of Food, and Culinary Poisons. AB'M Small, Philadelphia.
4. ORFILA, M. J. B. *Traite de Poisons, ou Toxicologie Generale,* 1st ed., 1814, 5th ed., 1852. publisher unknown, Paris.
5. BLYTHE, A. W. (1985). *Poisons. Their Effects and Detection.* William Wood & Co., New York.
6. HASSALL, A. H. (1855). *Food and Its Adulterations; Comprising the Reports of the Analytical Sanitary Commission of "The Lancet."* Longman, Brown, Green, and Longmans, London.
7. HASSALL, A. H. (1857). *Adulterations Detected; or, Plain Instructions for the Discovery of Frauds in Food and Medicine.* Longman, Brown, Green, Longmans, and Roberts, London.
8. British Parliament (1860). An Act for Preventing the Adulteration of Food or Drink; British Parliament (1875). Sale of Food and Drug Act.
9. ANDERSON, O. E. (1958). *The Health of a Nation.* Univ. of Chicago Press, Chicago.

10. USDA Bureau of Chemistry, Bulletin 84 (1904 and 1906). *Influence of Food Preservatives and Artificial Colors on Digestion and Health.* Part I: *Boric Acid and Borax.* Part II: *Salicylic Acid and Salicylates.* U.S. Govt. Printing Office, Washington, DC.
11. FUNK, C. (1914). *Die Vitamine.* J. F. Bergmann, Wiesbaden; (1922). *The Vitamins.* Williams & Wilkins, Baltimore.
12. Staff of the Division of Pharmacology, Food and Drug Administration, Department of Health, Education and Welfare (1959). *Appraisal of the Safety of Chemicals in Foods, Drugs and Cosmetics.* Washington, DC.
13. LEHMAN, A. J. (1959). Some Interpretative Problems in Evaluating the Safety of Food Additives. Chapter in Ref. 12, p. 101.
14. U.S. Court of Appeals, Washington, DC. (1979). In the Matter of Acrylonitrile Copolymers Used to Fabricate Beverage Containers, November 12, 1979 (testimony unpublished).
15. OSER, B. L. (1969). Much ado about safety. *Food. Cosmet. Toxicol.* **7,** 415–424.
16. CRAMER, G. M., FORD, R. A., AND HALL, R. L. (1978). Estimation of toxic hazard—A decision tree approach. *Food Cosmet. Toxicol.* **16,** 255–276.
17. OSER, B. L., WEIL, C. A., WOODS, L. A., AND BERNARD, B. (1985). Recent progress in the consideration of flavor ingredients under the food additives amendment. 14. GRAS substances. *Food Technol.* **39,** 108–117.
18. U.S. Food and Drug Administration, Bureau of Foods (1982). *Toxicological Principles for the Safety Assessment of Direct Food Additives and Color Additives Used in Food.* U.S. Food and Drug Administration, Bureau of Foods, Washington, DC.
19. RULIS, A. M., HATTAN, D. G., AND MORGENROTH, V. H. (1984). FDA's priority-based assessment of food additives. I. Preliminary results. *Regul. Toxicol. Pharmacol.* **4,** 37–56.
20. RULIS, A. M., AND HATTAN, D. G. (1985). FDA's priority-based assessment of food additives. II. General toxicity parameters. *Regul. Toxicol. Pharmacol.* **5,** 152–174.
21. STOFBERG, J., AND KIRSCHMAN, J. C. (1985). The consumption ratio of flavouring materials: A mechanism for setting priorities for safety evaluation. *Food Chem. Toxicol.* **23,** 857–860.
22. OSER, B. L. (1971). Food additives: The no-effect level. *Arch. Environ. Health* **22,** 696–699.
23. OSER, B. L. (1981). The rat as a model for human toxicological evaluation. *J. Toxicol. Environ. Health* **8,** 521–542.
24. BAZELON, D. L. (1981). The judiciary: What role in health improvement? *Science* **211,** 792–793.
25. *Toxicants Occurring Naturally in Foods,* 2nd ed. (1973). Committee on Food Protection. National Academy of Sciences/National Research Council, Washington, DC.
26. Federal Food, Drug, and Cosmetic Act, as Amended, and Related Laws. Sec. 402(a) (1986).
27. AMES, B. N. (1983). Dietary Carcinogens and Anticarcinogens. *Science* **221,** 1256–1264.
28. AMES, B. N., AND WHITFIELD, H. J. (1966). *Cold Spring Harbor Symp. Quant. Biol.* **31,** 221–225.
29. AMES, B. N. (1971). The detection of chemical mutagens with enteric bacteria. In *Chemical Mutagens* (A. Hollaender, Ed.), Vol. 1, Chap. 9. Plenum Press, New York.
30. AMES, B. N., KAMMEN, H. O., AND YAMASAKI, E. (1975). Hair dyes are mutagenic: Identification of a variety of mutagenic ingredients. *Proc. Natl. Acad. Sci. USA* **72,** 2423–2427.
31. Interagency Regulatory Liaison Group Workshop on Reproductive Toxicity Risk Assessment (1986). *Environ. Health Perspect.* **66,** 193–221.
32. HOCHHEISER, S. H. *Synthetic Food Colors in the United States: A History Under Regulation.* Thesis. Univ. Wisconsin (1982); Univ. Micro-Films Intl., Ann Arbor, MI (1986).
33–40. Scientific Committee of the Food Safety Council (June 1980). *Proposed System For Food Safety Assessment,* Chap. 11. Food Safety Council, Washington, DC.