EXPERIMENTAL AND CLINICAL Neurotoxicology

edited by

PETER S. SPENCER, Ph.D.

Associate Professor of Neuroscience
Director, Institute of Neurotoxicology
Albert Einstein College of Medicine, Bronx, N.Y

and

HERBERT H. SCHAUMBURG, M.D.

Professor and Vice-chairman of Neurology
Institute of Neurotoxicology
Albert Einstein College of Medicine, Bronx, N.Y.

WILLIAMS & WILKINS
Baltimore/London

Made in the United States of America

Library of Congress Cataloging in Publication Data

Main entry under title:

Experimental and clinical neurotoxicology.

Includes index.
1. Nervous system—Diseases. 2. Neurotoxic agents. I. Spencer, Peter S. II. Schaumburg, Herbert H. [DNLM: 1. Environmental pollutants—Toxicity. 2. Nervous system—Drug effects. 3. Nervous system diseases—Chemically induced. WL100.3 E96]
RC347.E99 616.8'04'7 79-23302
ISBN 0-683-07854-2

Composed and printed at the
Waverly Press, Inc.
Mt. Royal and Guilford Aves.
Baltimore, Md. 21202, U.S.A.

Chapter 48

Experimental Design for Animal Toxicity Studies

GERARD F. **EGAN,** STEVEN C. LEWIS, and ROBERT **A.** SCALA

INTRODUCTION

This chapter will consider some of the conceptual and practical considerations in the design of whole-animal toxicity studies. The conceptual issues will be concerned with the definition and evaluation of the problem, selection of approaches, consideration of variables, and the like. The practical elements of whole-animal experimental design, nominally worthy of a separate text, will be summarized and highlighted with respect to types of studies and design elements such as animal model, mode of exposure, selection of controls, and the sequence of steps in data evaluation.

GENERAL CONSIDERATIONS

Although it seems to be unduly obvious to make such a statement, it is necessary to have a clear idea about why the specific study is to be done. The investigator needs more than an overall concept **of** the total research program or the kinds of information desired. He must have a clear, explicit, and brief statement of why a specific experiment is to be performed. When the data are in hand and the interpretation is uncertain or ambiguous, one explanation may be an uncertain purpose. Normally, concise statements of purpose are defeated by succumbing to the very common temptation of adding just one more component to the study. In these instances, what starts out as a clear, straightforward experiment, becomes encumbered with side issues, add-ons, supplemental or secondary purposes, and the original intent is lost, obscured, or thwarted. Biological studies seem more likely to fail as they become removed from clear purposes and stepwise evoiution of understanding.

At each point in the development of the experiment, the nature and function of a control must be recognized. As one asks "Why are you doing this experiment?", it is also necessary to be able to determine whether the intended effect or observation has been made. This is ordinarily measured by comparison with the appropriate concomitant control.

PURPOSES OF LABORATORY ANIMAL STUDIES

Whole-animal laboratory studies in toxicology and related disciplines may be divided into three major types. These are simulation, characterization of adverse effect, and evaluation of safety.

SIMULATION

The earliest experiments in toxicology are thought to have been simple simulations, either predictive or diagnostic. In its most rudimentary mode, a simulation experiment involves dosing an animal with the same amount of material as the human would be expected to receive (or did receive). The route is likewise the same, and there follows a period of watchful waiting. From the earliest use of royal "tasters," where the experimental subjects were, unfortunately, other humans, through contemporary forensic toxicology assays, the simulation experiment has answered the question of "What would happen if..." More sophis-

ticated simulation studies stem from unanticipated, but not necessarily lethal, effects in man from environmental or other agents, and seek to reproduce the effects first grossly and then on successively more elemental levels.

CHARACTERIZATION OF TOXICITY

Modem-era toxicology has strongly emphasized characterization of toxicity as the initial stage in the evaluation of a new material. Though frequently confused with safety evaluation, the characterization of toxicity has different objectives and, consequently, differences in design. The major emphasis is on the identification of target organs and systems through the use of multiple doses and routes of exposure. By these means, the investigator attempts to understand the quantitative relation between dose by any given route and the magnitude of response in the whole animal, and in any organ or subsystem which can be measured by invasive or noninvasive means, including postmortem examination. Not to be ignored are differences in effects which may result from the use of different routes of exposure and the possibility of reversing the effect by concluding the exposure. During these studies, the investigator is also concerned with identifying those functional or other measures which are early indicators or measures of adverse effect. From this array of findings, it is possible to design an adequate series of safety evaluation studies.

EVALUATION OF SAFETY

Once the investigator has identified the specific toxic effects of a material, additional studies to determine safe conditions of use or exposure are in order. The design principles for such experiments are the use of a route which is comparable to the anticipated means of exposure. Food additive chemicals should be fed and not injected, for example. The doses should reflect use levels and reasonable exaggerations of use levels. The duration of study should be related to the anticipated pattern for man. Food additives, to return to the earlier example, should be fed for the lifetime of the test animal. Products of incidental-or infrequent use may be tested for shorter times, although chemical structural or other considerations could dictate long-term postexposure observations. The safety evaluation tests should employ those predictive clinical or other tests identified in the earlier studies. These should be regarded as the most efficient criteria of effect.

Perhaps the subject of greatest controversy and the most progress in recent times with respect to experimental design has been the choice of animal species. The traditional choices of animals for testing were based on cost, availability, size, susceptibility to the effect under study, and extent of the historical data base. In the majority of safety evaluation studies in the United States, the result of this selection process was the use of beagle dogs and white rats. For the last several years, however, the driving issues have been susceptibility to the desired effect (*i.e.* cancer-resistant *uersus* cancer-susceptible strains) and the pharmacokinetic similarity to man.

Pharmacokinetic Considerations

Increasingly, studies on characterization of toxicity have included assessments of the effect of the experimental animal on the material under study, in addition to the more obvious study of the effects of the material on the animal. Measurements of uptake, distribution, storage, bioconversion, and excretion, provide clues to the understanding of the mechanism of adverse effects, the identity of the active chemical agent, and possible species differences in effect levels on target organs. Most particularly, any pharmacokinetic study which can rule in or out a particular animal species with regard to relevance to man becomes a singularly powerful tool in the selection of test species for whole-animal experiments designed to evaluate the safety to man of a particular agent or material.

UNDERSTANDING *VERSUS* MEASUREMENT

The ultimate purpose of a whole-animal study should be an understanding of the effects produced by the agent administered, and not merely the compilation of data from certain tests, or the satisfactory completion of a protocol fixed according to some governmental or other regulation. While the latter may give the comfort of having marked all the boxes on a checklist, the experiments will have failed if they did not provide first a characterization of the effects and second, an understanding of underlying mechanisms. The checklist approach assures that hitherto unidentified or untested effects will be missed. Striving for understanding while not assuring that such effects will be found, certainly provides for the prepared and sensitive experimental setting in which they are more likely to be recognized. These statements are as axiomatic as the need for a firm purpose for each experiment, yet the literature gives generous examples of inattention to these principles.

SCIENTIFIC JUSTIFICATION FOR WHOLE-ANIMAL STUDIES AS A USEFUL MODEL FOR HUMAN DISEASE PROCESSES

Disease states can be produced in experimental animals by a variety of means. There are genetic

diseases whose prevalence and degree can be controlled to some extent by the breeding process: the availability, for example, of a colony of animals with a genetic trait toward hypertension provides a model in which to study the physiologic or pathologic consequences of the disease, and may be useful in screening of potential therapeutic agents. Disease states may be produced through surgical intervention and the ligation or extirpation of critical organs or tissues. The literature on adrenal cortical functions, and the consequences of the loss of these functions, is based, in part, on such animal studies. Replacement therapy experiments have followed and, particularly with endocrine disorders, the use of the animal model has led to successful therapies in man. The third mode of inducing disease states in animals is from exposure to chemical or physical agents. One of the most widely known is the production of diabetes mellitus subsequent to the destruction of the β-cells of the pancreas by the chemical agent alloxan. This animal model yields most of the clinical features of the disease in man, although the degree of similarity between experimental alloxan neuropathy and human diabetic neuropathy is in question.

Given the fundamental similarity of mammalian functions across many species, and the essential commonality of many cellular mechanisms and reactions, it is ordinarily to be expected that there would be a commonality of lesions whether genetically, surgically, or chemically induced. This obvious simplification becomes the opening premise in any specific instance and must be carefully tested by comparative study of the disease of "spontaneous" or unknown etiology in man and the induced disease in animals. Unfortunately, such whole-animal modeling is more likely to be confirmatory of the human disease rather than predictive.

HAZARD *VERSUS* SAFETY EVALUATION (DOSE-RESPONSE ASSAYS)

ACUTE TOXICITY INVESTIGATIONS

The foregoing discussions have been concerned with the purpose of conducting toxicity experiments and the rationale supporting the use of animal models to predict the potential deleterious effects of chemicals on Man or his environment. In general, toxicity investigations are designed to uncover and to characterize a range of harmful effects produced in intact biological systems as a consequence of chemical exposure. These studies also may be structured to provide some evidence of the dose-response relationship that may exist between a chemical and an observed biological change. The assemblage of all such toxicity data then can be employed to assess the degree of risk associated with exposure to, or use of, a selected material. Thus, the goals of toxicity experiments may be divided into two main categories:

A. Those which attempt to determine the range of specific toxic effects produced by a chemical, *e.g.* skin or eye irritation, behavioral modification, detrimental changes in reproductive function, cancer, etc.
B. Those designed to quantify the relationship that may exist between exposure at different levels and the degree of biological response in the test organism. The former are useful in characterizing a number of potential effects while the latter are employed in safety evaluation procedures.

Among the most simple and perhaps most maligned of all toxicity tests is the so-called LD_{50} or LC_{50} experiment. The expression is a calculated value representing a mathematical estimation of the dose (D) or concentration (C) producing death in 50% of the animals on test. As such, it should include some estimate of the error of that value, *i.e.* a probability range or a confidence limit. These limits, arbitrarily selected by the investigator, describe the probability of obtaining similar results if the test is repeated.

The LD_{50} test is conducted by administering the test material to groups of animals over a range of doses. The most common routes include oral, dermal, intraperitoneal, subcutaneous, and intravenous. The dermal test provides information on the skin-penetrating properties of the test substance. If properly conducted, dermal tests may also yield valuable information regarding the irritating and corrosive potential of the chemical.

The LC_{50} test is a series of acute inhalation trials wherein the duration of the exposure is held constant, *e.g.* 15 min, 1 h, 4 h, 6 h, etc., while the concentration of test substance in the breathing atmosphere is varied for each exposure. A variation of this procedure involves adjusting the time while holding the concentration constant. The so-called LT_{50} test then measures the time required to produce mortalities in 50% of the animals exposed to a given concentration of material.

Most LC_{50}/LT_{50} tests rely on a nominal calculation of the chamber concentration rather than an actual analytical determination of the exposure level. In those instances where the test species is highly reactive, it is imperative to carry out routine analyses of chamber atmospheres, to control exposure levels, and to correlate mortalities carefully with exposure concentrations.

The LD_{50} or LC_{50} value, and the graphical representation of such parameters, allows the investi-

gator to learn something about the toxicity and the potency of a chemical, particularly with respect to other chemicals. Figure 48.1 illustrates the dose-response relationships of three chemicals (A, B, and C), each having different slopes or different x intercepts. Plots A and B indicate the two materials have the same LD_{50} values. The dose required to produce 50% mortalities with both chemicals is the same. However, the difference in their slopes suggests that for a given dose increment, Compound B will produce more mortalities than Compound A. Furthermore, a comparison of the LD_{5}s of these materials indicates that, compared to Compound B, smaller doses of Compound A will produce mortalities less than 50%. Hence, A is said to be more potent. Plots B and C illustrate the dose-response relationships for two chemicals whose relative toxicities remain comparable over the range of doses tested. In this instance, the LD_{50} of C is less than that of B and, for any given dose, its (C) toxicity would be regarded as being lower than that of B.

The LD or LC_{50} values are useful, then, in comparing the toxicities and potencies of chemicals. These lethality tests also can provide some information regarding functional signs of intoxication as well as the time of onset and the duration of such signs. By careful attention to the character and progress of these signs, coupled with a thorough gross pathological examination of surviving animals (14-day postexposure), as well as those dying shortly after dosing, it may be possible to determine the mechanism of toxic action and the principal target organ. These gross examinations would routinely involve the heart, lungs, liver, kidney, spleen, gastrointestinal tract and, when deemed appropriate, the brain, gonads, adrenals, etc.

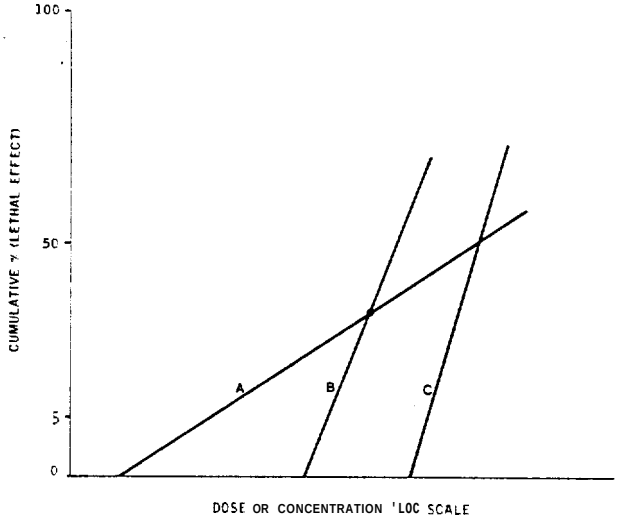

Figure 48.1 Hypothetical dose-response curve for three chemicals (*A*, *B*, and *C*) administered to a uniform population of biological specimens.

The misuse or maligning of the LD/LC_{50} value referred to earlier stems from a lack of appreciation of those variables which can have a significant influence on the outcome of the experiment. Although a more thorough discussion of experimental design variables will be presented in the succeeding section of this chapter, illustrative examples will be offered here of how such variables can impact on the outcome of acute, subacute, and chronic toxicity investigations. In the case of the LD/LC_{50} test, age, weight, caging conditions, dietary programs, species, route of administration, etc., can all exert a significant effect on the calculated LD/LC_{50} value. Unless it is possible to know the nature and extent of these variables as they influence the outcome of an LD/LC_{50} test, then, strictly speaking, any attempt at comparing such values would be highly suspect and without scientific base.

The estimation of the LD/LC_{50} value constitutes a first step in the process of assessing the toxicity of a material. The test has the advantage of being relatively simple, and it lends itself readily to statistical analysis. The calculated lethality index can serve as a yardstick by which to gauge the toxicity of that material. It also allows for the comparison of the relative toxicities of a number of chemicals. This comparative ranking has led to the use of such common descriptors as ***highly toxic, moderately toxic, nontoxic,*** etc.

The LD/LC_{50} test is, by definition, a parameter of acute toxicity. The latter term is used to describe those adverse effects occurring during an exposure and/or for a period of time up to 14 d postexposure. More often than not, the exposure or dosing regimen is a singular event whose duration may be measured in a matter of seconds, *e.g.* intravenous, intragastric, intramuscular, subcutaneous, topical administration, etc., or hours, *e.g.* inhalation exposure. Acute (short duration) exposures may result in toxic effects which, over a period of time, will be self-resolving. Such brief exposures, however, can also produce an effect which is still in evidence months to years later. For example, animals or humans exposed to elevated concentrations of hydrocarbon vapors for a period of minutes to hours will experience a range of central nervous system (CNS) effects. These may include headache, nausea, vomiting, dizziness, drowsiness, sleep, coma, and perhaps even death. If exposure is halted during the early stages of central nervous system involvement, the effects will gradually subside and the animal or individual will recover completely. Should the exposure continue to the point where brain oxygen levels have been seriously compromised, due to an interruption in the available supply of oxygen-rich blood to the brain or to an interference in the metabolism of oxygen by the

brain cells, then nerve cell damage could begin and the resulting toxic effect might be permanent.

The design of acute toxicity studies must take into consideration several significant factors, among which the following are probably the most critical:

A. Selection of species
B. Route of administration
C. Concentration of test material in vehicle
D. Duration of exposure and observation period.

The selection of species in acute toxicity investigations is more often than not based on practical considerations. A wide range of genetically homogeneous rats, mice, guinea pigs, hamsters, and even rabbits, is available to choose from. Depending on the nature and goals of the experiment, a given species is generally employed because of the ease in handling, space considerations, cost, the number of animals which are to be treated, etc. Rats and mice are typically selected for the various peroral and parenteral LD_{50} studies. The calculated lethality index can then be compared to values for other chemicals which are commonly derived from rat or mouse studies.

Because of the ease with which they can be manipulated, and the fact that they have sufficient body surface area to permit a variety of dermal toxicity investigations, rabbits are often selected to act as human surrogates in a number of dermal irritation, penetration, and sensitization studies.

Species variation in acute toxicity experiments has been well documented (4, 25), particularly with respect to the two most commonly used species for LD_{50} estimation, namely the rat and the mouse. Morrison and co-workers (20) reported that the oral LD_{50} in male albino mice for 5-(*N*-piperidine)10,11-dihydro-5*H*-(*a*,*d*)cycloheptene was 1160 mg/kg, while in male albino rats the value was found to be 6.6 mg/kg. The classic example of norbormide species sensitivity is also offered for consideration (rat oral LD_{50}—5–15 mg/kg *versus* 2250 mg/kg for mice) (20). Such differences tend to be the exception rather than the rule, as noted by Weil and Wright (36) and Weil and associates (34). The question of intraspecies variation is also one that must be considered. Thus, the LD_{50} for thiourea in the Hopkins rat strain was 4 mg/kg, while a value of 1340–1830 mg/kg was noted in the Norwegian strain (20). One could speculate that the differences in toxicity recorded among species or within a given species may be a function of relative liver microsomal enzyme activity. Such differences in enzyme activity may ultimately be genetically determined. Other biological factors which can influence the toxicity of a material, *e.g.* age, sex, etc., should also be weighed carefully when deciding on the elements of an acute toxicity protocol. The reader is referred to articles by Conney and co-workers (6), Yearly and Benish (38), and Lu and associates (18), for discussion of the impact of sex and age on the outcome of acute LD_{50} determinations. A brief review of their findings suggest the following:

A. Young animals do not have a full complement of mixed-function oxidase (MFO) enzymes in the liver. These enzymes are thought to be responsible for the metabolism and ultimate detoxification of a number of xenobiotes and a wide variety of endogenous metabolites, *e.g.* steroids, etc.
B. Older animals experience a gradual dimunition of MFO enzymes. As a consequence, their ability to detoxify exogenous materials is far less efficient than the young adult. They also tend to be more obese than their younger counterparts. Their fat tissues can allow for the deposition and retention of fat-soluble material for an extended period, prolonging, in some instances, the period of intoxication.
C. Young adult (*e.g.* sexually mature) animals of a given species tend toward greater metabolic and enzymatic homogeneity. This is probably a function of the stabilization of MFO activity, sex hormone influence, and relative CNS sensitivity (9, 15, 28).
D. In general, males and females tend to react similarly in acute toxicity tests. However, in those instances where the metabolism of the test materials affects, or is affected by, the metabolism of endogenous sex hormones, it is conceivable the resulting toxicity could be sex related.

The route of administration, *i.e.* oral, dermal, intravenous, inhalation, etc., is of paramount importance in determining the acute toxicity of a chemical. The manner in which a test material is administered in acute studies should reflect the most likely route(s) of exposure for humans, particularly as in those cases where the resulting data are to be employed in safety evaluation exercises. However, when the purpose of the test is merely to identify the acute toxic properties of a chemical, more often than not, the route of administration is peroral, dermal, or one of a number of parenteral routes. When selecting a route of exposure, it is necessary to consider the impact that certain organ systems can have on the metabolism of that test material. Thus, the toxicity of compounds administered via the oral route can be modified by a number of factors. Certainly uptake and transport via the enterohepatic system allows the liver to exert a major influence (detoxification or toxification/activation). The changing pH profile along the gastrointestinal tract can influence the relative absorption rates of ionizable test species. Stomach-

emptying time and gastrointestinal food content' also will play a role in the ultimate expression of a chemical's toxicity when administered via the oral route. It is evident that an intravenous administration of a material allows for the most efficient and rapid distribution of a test chemical throughout the body. Intramuscular and intraperitoneal administrations are less efficient, while the subcutaneous route permits a very slow release of material to the systemic circulation. It is also important to weigh the relative vascularity of the area to be treated.

The inhalation route represents a different set of technical problems. To conduct a proper, well controlled, acute inhalation trial, it is important and useful to know the test material's vapor pressure, boiling range, and the lipid/water partition coefficient, particle sues for aerosols and dusts, etc. With such information, it should be possible to determine the theoretical dose administered to the test species over a period of time. This determination presupposes a knowledge of the respiratory dynamics of that test animal, *e.g.* respiratory rate, tidal volume, minute volume, etc.

Respiratory exposures are often complicated by the tendency of test animals, rodents in particular, to ingest a considerable quantity of test material. This occurs as a consequence of grooming activities along with the ingestion of food and water which, while maintained in the inhalation chamber, become contaminated by the test material. In these instances, any observed toxicity would have to reflect the combined contribution of ingested and inhaled test material.

The size of the dose (volume),the concentration of the test material in its vehicle (diluent),and the nature of the vehicle may all have a significant influence on the outcome of the acute experiment. Thus, the administration of an oral dose in a volume which approaches or exceeds that of the test animal's stomach will only guarantee the artifactual character of the observed toxic response. In this instance, the biological response is but a manifestation of the unphysiological nature of the dosage volume employed. Regarding the impact of dilution, Griffith (11)has demonstrated that the concentration of test material in a diluent (vehicle) can have a major impact on the outcome of an acute oral experiment. Increasing the concentration of the test compound by a factor of *2.5* results in a concomitant decrease in toxicity by a factor of approximately **2.**

Although a well planned acute LD/LC_{50} toxicity test can provide useful information regarding a chemical's relative toxicity and the influence of species, age, sex, route, and mode of test material administration on the toxicity of that chemical. there are some major shortcomings associated with the acute test protocol. For example, the LD_{50} test does not provide an indication of toxicity *per* ***se.*** Rather, it is an experiment designed to determine the association between dose or extent of exposure and **a** predetermined end point, *e.g.* death. No information is developed regarding the persistence or the gradual development of any lesions, be they physiological or morphological in nature. The question of nonlethal doses bringing about the development of an irreversible lesion at a time subsequent to the routine postexposure observation period is also not addressed by the routine acute LD_{50} test. Thus, the carcinogenicity of 3-methylcholanthrene, 1,2:5,6-dibenzanthracene, and 3,4-benzpyrene, which can be evidenced after a single dose treatment and a 1–2-y observation period, is not at all reflected by the LD_{50} value or by observations made during the 7–14-d postexposure period (14). Other examples include the clinical effects of coumarin-based anticoagulants whose action is only manifest after the depletion of endogenous supplies of prothrombin. The same may be offered regarding the neurotoxic effects of tri-*o*-cresyl phosphate, lead, methyl mercury, and phenacemide (13, 14). That lesions induced in survivors of acute toxicity tests may become manifest after more than a year of observation is illustrated by the **work** of Schoental and Magee (27) and Wheeler and colleagues (37).

SUBACUTE AND SUBCHRONIC EXPERIMENTS

The subacute/subchronic experiment is a logical and necessary extension of the acute test, particularly regarding any effort to assess the safety of a chemical in the work environment, as a human food component or as a household product. In many instances, the acute toxicity study serves as a means of establishing exposure levels, *i.e.* concentrations in dietary regimens or in inhalation trials for the subacute/subchronic test.

Within the broad category of subacute/subchronic experiments, there are essentially two classes of studies:

A. Routine tests designed to determine the adverse effects of regularly repeated exposures over a period ranging from several days to 6–**9** months. In these studies, the intent of the experiment is to produce no observed effect(s) at one level and adverse effects in at least one higher exposure level. The parameters monitored may include biochemical changes, *i.e.* enzyme assays, hematology, clinical chemistries, urine analyses, organ and body weights, food intake, and physiological/functional, behavioral, and histopathological determinations.

B. Specialized studies designed to determine the metabolism and pharmacokinetics of a material, its mutagenic/teratogenic or neuropathogenic potential and, in some instances, the impact of repeated exposure on the immune mechanisms of the body.

The selection of species for these extended studies is generally pragmatic. Ideally, it would be preferable to choose a species whose metabolism of the test material is comparable to that in man. Rats, mice, and rabbits usually have been employed in such tests, because of their size and the extensive background on husbandry and nutritional needs that has been developed on them during the course of several years of experimentation. The dog is often used, particularly when the use of a nonrodent species, which is phylogenetically closer to man than rodents, may yield more meaningful and significant information.

The route of exposure ideally should reflect the anticipated route of contact in man. For some materials, where the inhalation route represents the nature of the real-world exposure, it may be possible to use a feeding study ***in lieu*** of the more costly inhalation trial. This is usually done only if it has been demonstrated that the metabolism and pharmacokinetic profiles of the test material are the same regardless of the route of administration, *i.e.* inhalation ***versus*** oral.

Testing involving rats should be initiated just after weaning so observations can be made during the period of most rapid growth. If no data are available regarding the metabolism of the test material in man, it is often deemed appropriate to employ a species which has been shown to be most sensitive in preliminary acute toxicity tests.

When the test material is to be administered as part of the dietary regimen, it is essential to keep the following in mind:

A. Some test materials are altered or modified by interaction with dietary components or as a consequence of exposure to air or moisture. In those instances where it is imperative to determine the toxicity of the chemical ***per se,*** it is then essential that chemical analyses be carried out to determine the stability of that chemical in the diet. Should these tests reveal that alterations in test material can and do occur with time, it may be necessary to replace the food supplies on a more frequent basis or to administer the chemical to the test animals by means of gavage or in an encapsulated form.

B. When test chemicals are added to the diet as a fraction of the total diet to achieve a predetermined dose level, *e.g.* mg/kg body weight per day, it is essential that the dietary levels be adjusted weekly or biweekly to reflect any variations in total body weight.

CHRONIC OR LIFE-TIME TESTS

Chronic or lifetime toxicity tests are performed to identify and characterize toxic effects whose pathogenesis requires several months before becoming manifest as a frank lesion, *i.e.* tumor, nerve fiber degeneration, etc. Such tests also may be carried out to quantify the nature and extent of those toxic responses requiring prolonged and/or repeated exposure before the disease state becomes evident.

The carcinogenesis bioassay is a special kind of long-term toxicity investigation designed to detect a specific kind of lesion, namely a tumor in a given organ system, *e.g.* hepatoma, lung adenoma, skin papilloma, etc. In contrast to this specialized test, the chronic toxicity investigation is geared toward detecting a wide range of pathological changes including tumors, blood dyscrasias, decrements in reproductive capacity, modifications in central and peripheral nervous system homeostasis, *i.e.* aging, behavioral changes, etc., and fibrotic changes in pulmonary tissue, modifications in basement membrane architecture, induction of autoimmune diseases, etc. With regard to the selection of test species for chronic toxicity investigations, including the carcinogenesis bioassay, it is usually impossible to determine with any degree of confidence the fidelity with which a given animal model mimics the metabolism and pharmacokinetics of a test chemical in man. Nor is it possible to know whether the dose of the chemical at the active receptor site would be similar in the chosen test animal to that occurring in man. Consequently, long-term toxicity investigations are often designed using an animal model which may be insensitive or otherwise inappropriate for predicting human health hazards. This certainly should be borne in mind when an investigation yields negative results for, in this instance, the test results should be considered as a "false negative." On the other hand, for the very same reason, a positive finding in the animal model may represent a "false positive" result. To minimize the prospects of such a dilemma, it is recommended that two unrelated species (rodent and nonrodent) be employed in chronic toxicity experiments, particularly when it is not possible or practical to determine the similarity in metabolic and pharmacokinetic profiles of the chemical in the test species and man.

The duration of treatment in either carcinogenesis bioassays or in chronic toxicity investigation may include a 1-year or half-life time exposure period followed by a 1-year to lifetime postexposure

observation period. Other long-term tests are designed to carry exposures until no survivors remain while in still others, sacrifices are initiated at some arbitrary point near the natural end of the test animal's lives, *i.e.* 24–30 months for rats, or when the survivors number 10% of the original test group, etc.

PROGRESSIVE AND RETROSPECTIVE STUDIES

Toxicology investigations are often carried out on materials for which there are no toxicity data. Such is typically the case for the vast majority of materials currently in commerce. Experimental investigations of this kind are termed progressive. In contrast to this class of study, an investigation may be triggered by the detection of an adverse effect during the manufacture or processing of a chemical, or as a consequence of its use in commerce. When a test is initiated under such circumstances, it is termed retrospective.

In some prospective animal studies, covert functional changes may go undetected. Often, such defects are only uncovered after human experience with the compound. Thus, Hawkins and Laurie (12) discovered that aminoglycoside antibiotics can cause hearing impairment, an effect which went undetected in progressive animal investigations. Furthermore, it is reasonable to assume the metabolism (detoxification) of an exogenous material could be compromised if the liver is diseased, as in the case of alcoholics. The change in metabolism in this instance, *i.e.* a prolongation of the biological half-life of the material, would probably enhance the toxicity of that substance and/or prolong the manifestations of its toxicity (2). Any such potential, yet highly probable, effect would not be predicted in a routine toxicity test.

Retrospective studies also have their own limitations. The association between prolonged exposure to elevated levels of benzene and the production of leukemia has never been confirmed in retrospective animal studies. The "beer drinkers myocarditis" observed in heavy beer drinkers in Canada and the United States in the mid-1960s) was difficult to confirm in animal tests until it was noted that protein-deficient diets, high levels of cobalt salts, and beer, were required in the test diets (10, 24).

In short, the progressive and retrospective animal toxicity studies must be pursued carefully to ensure an accurate interpretation of the test results. An awareness of the potential covariables, *e.g.* diet, preexisting disease, metabolic predisposition, etc., must all be given careful consideration before those interpretations can be deemed valid and useful.

EXPERIMENTAL DESIGN CONSIDERATIONS

At the outset of any experiment in toxicology, two questions need to be posed: (a) What is the purpose of the experiment or what information is being sought? (b) What is the nature and potential impact of those variables which, of necessity, become an integral part of any experimental design? The preceding section attempted to provide information needed to answer the first question. This section will concern itself with a detailed review of the major design variables which have a significant impact on the outcome of a toxicity study.

THE ANIMAL MODEL

As has already been noted, the choice of the animal model for a toxicity investigation is often based on practical considerations, *i.e.* the rat and mouse are frequently employed because of their size, ease of handling, short life spans and relatively low costs. These animal models have other characteristics which are important in the design of an experiment: (a) Their lack of a regurgitation reflex makes them suitable for those experiments requiring the administration of a test material intragastrically. (b) Certainly, it is advantageous to know the tumor incidence and the nature and extent of other serious lesions in a given animal model before electing to use it in a chronic investigation. Many rat and mouse strains have been inbred for over 50 years. In so doing, the gene pools in these lines have become quite stable and homogeneous. This, in turn, has provided for a variety of rodent models with highly predictable incidences of tumors and other serious disease states.

The notion of predisposition to selected pathologic processes deserves some consideration here. The pathogenesis of many disease entities requires several months before evidence of any detrimental change becomes manifest. Knowledge of the incidence, age of onset, and time-to-median-lesion development, is important in the selection of a species for a chronic toxicity test, particularly in the case of carcinogenesis bioassays. If an animal model has a high incidence of cancer in a specified organ, *e.g.* hepatocarcinoma, then it would be very difficult to obtain any meaningful data regarding the hepatocarcinogenic potential of a test material in that animal model. High background levels can be important, however, when the antitumor potential of a chemical is to be assessed. Also, a species with a predisposition to atherosclerosis would be quite useful for the study of drugs designed to modify cholesterol synthesis. There exist today numerous animal species which have been developed because of their predisposition toward selected disease

states. One has only to decide the purposes of an experiment before making the model selection. However, it has been argued that since Man is a random-bred animal, it would be appropriate to utilize an animal model with a heterogeneous gene pool. Although this argument is sound, it has its practical limitations. Thus, because studies generally involve the comparison of a limited (40–100) number of exposed animals with a control population of comparable size, it is important that there is not a large variation in the incidence of a particular disease in the control population. Otherwise, life expectancies may be compromised and the real significance of treatment-related disease processes might be masked by background pathological changes.

Interspecies differences in metabolism are also important variables to consider in the design of experiments. As indicated in the preceding section, differences in toxicity often can be associated with variations in metabolism. Most of these differences probably are linked to the genetic makeup of a species *(8).* For example, dalmations may be the only member of the canine species which metabolizes purines to uric acid and/or allantoin, a process similar to that occurring in man. Microsomal-mediated hydroxylations are often species-specific, *e.g.* 7-hydroxylation of coumarin in man ***uersus*** 3-hydroxylation in rats and rabbits (6, 8).

Along these same lines, the potential impact of exogenous materials on the microsomal enzyme system should be given careful consideration. The formation of a toxic metabolite or the detoxication of a chemical by this system can be affected significantly by pre-exposure to exogenous chemicals, *e.g.* phenobarbital, polychlorinated biphenyls, etc., which act to induce or otherwise enhance microsomal enyzme activity. In a similar fashion, it has been demonstrated that volatile hydrocarbons emitted from cedar wood chips used for rat and mouse bedding can induce the microsomal enzymes. Thus, in addition to the need to control species selection because of the importance of relative microsomal enzyme activity, it is equally important to control the environment of the animal cage as this may also affect that activity.

The selection of an appropriate animal model for a toxicity study also should reflect a consideration of relative species sensitivity. For example, the chicken is generally regarded as appropriate for the assessment of organophosphorus neurotoxicity in Man since both species respond with neuropathy after a similar dose and time delay *(5).* The chicken is also very sensitive to other chemicals such as methyl n-butyl ketone (19), but it is the rat which is generally regarded as a suitable mammalian species for neurotoxicity testing **(32).**

In addition to the problems inherent in the selection of an appropriate species and strain of animal, the importance of body weight, age, and sex should be assessed, particularly as these variables are among the easiest to control, and because they may often exert a detrimental impact on the outcome of the study. Thus, Balazs (1, 2) reported that the subcutaneous LD_{50} of isoproterenol hydrochloride is about 800 mg/kg in 200-g male Sprague-Dawley rats, but only **0.35** mg/kg in 600-g males of the same strain. Sex and strain, however, had no influence on the outcome of this LD_{50} experiment. Weight differences in animals of approximately the same age would suggest differences in total body fat mass. As such, the heavier animals have greater storage capacities for lipid-soluble materials which could prolong the period of intoxication associated with a chemical. Mention already has been made of the significant differences in metabolic (microsomal) activity between adult animals and those that are very young or very old. Age-related differences in toxicity also can be due to factors such as an incompletely developed blood-brain barrier. For example, the translocation of quaternized nitrogen-containing compounds across the blood-brain barrier is inhibited in the adult but not in the immature individual. In contrast to this age-related sensitivity is the finding that conjugation reactions are more fully developed in most young animals than they are in infants **(1,8, 20).** This variable could permit the regulatory approval of material as a drug for pediatric use when it might be highly toxic in the human infant. Finally, it should be appreciated that animals have age-associated degeneration of the Liver and kidneys as well as degeneration and regression of other organs and tissues. It would be reasonable to employ such older animals in tests designed to assess the bioactivity of drugs developed for geriatric use.

Wherever possible, chronic and subchronic toxicity investigations should be carried out using male and female animals, unless there are data to suggest sex-related sensitivities do not exist, or unless the goals of the program dictate the use of males ***uersus*** females.

EXPOSURE PARAMETERS

Among the important categories of variables having significant impact on the outcome of an experiment are those concerning the character of the exposure, *i.e.* route of administration, the length or duration of exposure and its frequency, the dose or concentration, etc.

Route of Administration

In acute toxicity studies, the route of administration is generally selected on the basis of conveni-

ence unless the stated objective of the study is to determine toxicity *vis-a-vis* a specified route of exposure, *e.g.* dermal, oral, etc. However, when it is necessary only to assess the general acute toxicity of a material, the latter may be administered intraperitoneally, subcutaneously, intramuscularly, intravenously, or intragastrically. The intravenous route bypasses absorption problems associated with the other listed routes. The use of the intraperitoneal route results in a rapid absorption of the test compound due to the high degree of vascularization in the peritoneal cavity and to the relatively large surface area of absorption. However, the use of this route allows for the passage of the test material through the liver via the hepatic portal route. In so doing, the test material may be metabolized (detoxified or activated) rapidly. This will also hold true for chemicals administered via the oral route. The oral administration of a material also may present problems beyond those associated with the impact of the enterohepatic circuit. It is conceivable that the rate of absorption from the gastrointestinal lumen to the surrounding vascular bed could be influenced by the quantity and quality of the food content in the tract. Furthermore, it is possible that the chemical composition of the test chemical could be affected by that food and/or by the intestinal flora. For these reasons, it may be prudent to consider employing more than one parenteral route of exposure in subchronic and chronic toxicity studies. In acute tests, this problem can be avoided by fasting the animals for the 12-h period immediately preceding treatment.

Chemicals administered via the subcutaneous or intramuscular routes will be absorbed more slowly than those introduced intraperitoneally. If the test material has any vasoactivity, *i.e.* vasconstriction or dilatation, then the rate of absorption via the subcutaneous or intramuscular route can be markedly affected.

The most complex route of administration is that presented in inhalation exposures. In addition to problems involved with the generation of contaminant (test chemical) concentrations at a predetermined level and at a constant rate, inhalation tests require a fairly extensive analytical control of the inhalation chamber test environment. Among the parameters which should be monitored, particularly in subchronic and chronic studies are: relative humidity, temperature, chamber air pressure, concentration of test material and, in those studies involving aerosols or other airborne particles, an analysis of particle size distribution and configuration (shape). Often, test chemicals are very hygroscopic and may, as a consequence of this interaction with water molecules, undergo spontaneous hydrolysis. Chamber relative humidity can affect the mucous linings of the respiratory system, particularly at low values, *i.e.* less than 30%, when mild irritation may result.

Chamber temperature is important in maintaining chemicals in a predetermined physical state, *i.e.* vapor *versus* liquid (aerosol). Furthermore, some polymeric materials which can exist in the vapor state at room temperature, *e.g.* dicyclopentadiene, can generate monomers (cyclopentadiene, in this instance) at slightly elevated temperatures. Chamber temperature is also of consequence to the health of the test animals.

Chamber air pressure provides an indication of test chamber stability relative to the exterior environment. As such, any potential leaks will be into rather than out from the chamber. Measurement of chamber pressure also provides an indication of airflow stability in the chamber.

Duration and Extent of Exposure

The duration of exposure may represent a variable requiring control, particularly in acute toxicity investigations. For example, the rate of infusion of a test material could affect the intravenous toxicity of a chemical. A relatively slow rate of administration via this route could be equal to the rate of detoxification of a compound. A rapid rate of infusion, however, might also prevent the effective interaction of a test chemical with the reactive site(s) in the body.

Acute dermal toxicity investigations often require the application of a test material to the skin of rabbits for a 24-h period under an occlusive bandage. Such a protocol is deliberately biased to maximize the likelihood that test materials will be absorbed through the skin. However, in the case of solvent evaluations, the defatting properties of the test material can produce a breakdown in the protective epidermal and dermal tissues, particularly under occlusive conditions. This could enhance dermal absorption resulting in an apparent increase in systemic toxicity. If the same volume of solvent was applied over a larger surface area under occlusive conditions, then the dermal toxicity probably would be decreased.

Finally, mention should be made of the utility of chronic (1-2-year) toxicity studies in determining "no effect" doses. With the exception of cancer, most toxic effects recorded in chronic trials can be detected in studies as short as 90 d. Weil and McCollister (35) have provided an extensive review on this subject wherein they concluded that a dose producing no toxic effects when thoroughly studied for 90 d will produce no effects, cancer excepted, when administered over the lifetime of the test animal. These authors also concluded that the most sensitive criteria of systemic toxicity are body

weight, ratios of liver and kidney weight to body weight, and kidney pathology.

The dose or concentration of a test substance must be well quantified. In the case of peroral or parenteral administration of test chemicals, it must be demonstrated that vehicles or diluents do not interact with, or in any way modify, the physical or chemical nature of the test material. The importance of dilution also must be weighed. Thus, the oral toxicity of chemicals can increase when administered in dilute solutions (1). This is thought to be due to enhanced intestinal absorption of the chemical. However, in some cases, increases in concentration bring about subsequent increases in toxicity primarily due to local irritation.

The total volume of test material and vehicle also should be controlled in acute studies. In general, the injected volume should be kept small in relation to the sue of the animal mode. Oral doses should not exceed 2–3% of the body weight while intravenous doses should be controlled to an upper limit of 0.5 ml for rodents and 2 ml for larger animal models. (8). Should it be necessary to inject volumes greater than those recommended, then multiple dosing at several injection sites should be considered.

Often, the relative toxicity of a test material is quite low. In such cases, it may be necessary to increase the concentration of the chemical in chronic toxicity test diets. In so doing, however, the relative percentage of nutrientain the diet may be compromised significantly. To control for this problem, paired feeding of control groups should be considered. Here, the paired controls are allowed only the amount of food consumed by the test group. Another approach is to include an inert material in the diet of the controls so that test and control groups experience the same dietary restrictions *vis-a-vis* daily intake of nutrients.

Some final comments are offered regarding the diluent or test vehicle. This material should have a very low order of toxicity and it should not interfere with, or in any way modify, the physical and/or chemical properties of the test material. The diluent should not enhance, inhibit, or otherwise interfere in the uptake, distribution, metabolism, or excretion of the test material. Its osmolarity, as well as the combined osmolarity of the test material and vehicle, should not affect local tissues. In short, the selection of a vehicle for use with those materials requiring dilution, solubilization, or suspension in a fluid media, should reflect a potential for significant modification in the actual toxicity of the test, regardless of the route and/or duration of exposure.

EXPERIMENTAL CONTROL'S

The design of any toxicity experiment should reflect a concern for standardization, that is, a need to include standards which permit the investigator to know with a high degree of precision the true meaning of any observed changes in the exposed or treated animals. Adequate controls in an experiment are particularly important with respect to animals, the test material, and the environment in which the animals are treated and/or maintained throughout the course of the experiment.

Animals

In general, there are three major types of animal control groups: negative, sham-treated, and positive. Negative controls are selected at random from the same population of animals as is used to form any test or treatment groups. The age (weight range) and sex should be similar to the test group(s). Such controls are maintained in animal holding rooms and are examined, and sampled for body weight, hematological and clinical chemistry determinations, etc., at the same time interval as are the treated groups. The negative controls are then sacrificed as per protocol requirements, usually at a time coincident with the sacrifice of the treated animals. A complete gross and histopathology work-up comparable to that employed on the exposed group(s) is then carried out on these controls.

Sham-treated controls are selected in the same manner as described above for negative controls. These animals are then treated in exactly the same manner as are the test or exposed animals, except the sham controls never encounter the material on test. For example, subacute or subchronic tests requiring intubation of the test material in a diluent would have a parallel control group which received the diluent via intubation. The volume of the vehicle so administered to these controls would be equal to the largest amount administered to the treatment group. Sham-treated controls for an inhalation test, regardless of its duration, would be maintained in an inhalation chamber and exposed to the same temperature, pressure, relative humidity and air flow characteristics as maintained in the treatment chamber. This group also would be exposed to a vehicle should the latter be incorporated in the test exposure regimens. An example was offered in an earlier section of this review regarding the need to use paired feeding in those feeding studies where the test material, non-nutritional in nature, represents a significant portion of the diet. Such paired feeding studies also would be considered to be sham-treatment controls. The purpose of the sham-treatment is to determine if the method of administering the test material contributes in any way to the toxic effects observed in the treatment group. Therefore, every effort should be made to incorporate this control group in studies, particularly when the nature of the dosing tech-

nique could be expected to influence the outcome of the study.

Positive controls are included in studies which focus on a specific biological event, *e.g.* sensitization, cancer, mutation, teratogenesis, neuropathy, etc. In such investigations, it is often considered appropriate, and indeed prudent, to maintain a control group of animals which is treated in the same time frame as are the test group(s) with a material known to produce the toxic effect under study. A positive control group allows the investigator to titrate the level of sensitivity of the treated groups and to compare this relative sensitivity to that recorded in other similar tests. It is then possible to know with greater confidence that a negative test result is more a reflection of the inherent low toxicity of that test material than a relative lack of animal responsiveness. The use of positive controls also permits the comparison of oncogenic potency between positive controls and treatment groups. Finally, the use of positive controls provides an indirect check on the reliability of a test laboratory. The choice of the positive control for an investigation should reflect an awareness of the chemical structure and possible similarities in metabolism, distribution, and excretion, between that control material and the test compound(s). This is particularly so with carcinogenicity bioassays and mutagenicity screens. However, teratology tests tend to rely on a handful of dissimilar compounds for the use as positive controls, *i.e.* aspirin, 6-aminonicotinamide, etc. For neurotoxicology investigations, it would be appropriate to select a material whose chemical composition or tissue response most closely approximates that of the test substance. The selection of positive controls should reflect variable sensitivities that exist among animal species ***vis-a-vis*** a particular positive control. For example, aspirin is not an effective teratogen in rabbits. Some nitrosamines cause a specific form of tumor in some animal models but not in others. Organophosphorus compounds are relatively weak neurotoxins in rats and guinea pigs but they are quite potent in chickens. It is well to point out that some investigators do not use positive controls because of the potential for contaminating the test group(s).

The Test Chemical

Among the variables that need careful consideration in toxicity investigations are those concerning the actual chemical to be tested. Sample purity and stability certainly should be ascertained before the onset of any investigation. Any such effort, however, presupposes the availability of an analytical technique which has sufficient detectability. selectivity, and reproducibility. As many studies will involve the dosing of animals with a compound in the parts-per-million or -billion range, the method of analysis will have to be capable of detecting extremely low quantities of the test compound while, at the same time, being able to discriminate that compound from the vast array of other chemicals potentially coexisting in the same environment, *i.e.* diet, chamber atmosphere, etc. The analytical technique should also be free from interference from any other chemicals existing in that environment. Finally, it should be possible to reproduce the quantitation of a material repeatedly and with time. Once such an analytical capability is in hand, then it is necessary to analyze the material to be tested to determine the nature and extent of any contaminants likely to be present with the test compound. Some studies are designed to assess the toxic properties of a pure chemical without interference from, or contribution by, contaminants present with the test compound. Other tests involve the evaluation of a chemical or product as used in commerce. In this instance, it is appropriate to examine the toxicity of the pure chemical and any associated by-products from its manufacture.

The importance of an adequate analytical technique is nowhere more important than when applied to the control of an inhalation toxicology experiment. In the course of preparing for, and actively carrying out such a study, the investigator must be able to monitor the chamber atmosphere for the contaminant or level of test material. It is necessary to be able to make rapid adjustments in the delivery of test material to the chamber to compensate for excursions that occasionally occur in chamber contaminant concentrations. Where feasible, a continuous analysis of contaminant levels should be provided. This can be accomplished with the use of infrared and ultraviolet spectrophotometric procedures. Gas chromatographs can be employed, but the analytical results lag by several minutes behind the actual sampling process. Regarding the latter, it is important to evaluate any procedure used to collect and/or transmit the contaminant from the chamber to the analytical unit. It is possible that the sampling train may introduce some interference in the actual level of contaminant. This can be accomplished by altering the contaminant's physical/chemical properties, thus interfering with the actual measurement of the test material. In other instances, the test material can interact with, or plate-out on, the surfaces of the sampling train. In short, the latter can represent a major factor in any effort to monitor chamber contaminant levels effectively.

Experiments involving the generation of aerosols or particulates present an added dimension to the problems of sampling and analyzing chamber atmospheres. These studies require the measurement

of particle (aerosol) sizes, and their distributions, in order to obtain some information regarding the respirable load presented to the test animal. Experiments of this type usually require some preliminary measurements of chamber contaminant levels to assure the adequacy of the contaminant generation apparatus and to demonstrate that the sampling and analytical systems do not interfere with the calculation of the true contaminant level. These preliminary tests are conducted in chambers without the complement of test animals, yet, it is conceivable that the presence of animals can have a significant effect on measured contaminant levels. Animal fur, cage surface areas, animal-generated humidity (respiratory and waste material) all can play a role in decreasing the expected level of test material in the chamber.

Environment

The last major category of experimental design variables is that involving environmental conditions. Included in this grouping are temperature, pressure, himidity, light cycles, caging conditions, and diet. Doull (8) and Morrison and associates (20) provide a thorough discussion of the impact of these environmental parameters on the outcome of toxicity investigation. Temperature changes have variable effects on the rate at which exogenous materials are metabolized. For example, the mouse intraperitoneal LD_{50} for amphetamines has been recorded to be about 6.0 mg/kg at 35°C, 155 mg/kg at 21°C and 84 mg/kg at 10°C. On the other hand, increases in relative humidity decrease the LD_{50} values for a number of chemicals. Pressure is more likely to present a problem in inhalation toxicity studies than it will in other areas of toxicology. The relative chamber airflow rates represent the source of this problem. Lighting conditions can play a major role in affecting the internal homeostatic patterns or biological rhythms of animals. This is particularly so with rodents. The latter are nocturnal and yet tend to be maintained in rooms which are lit during daylight or working hours. Although these animals can and do adjust rapidly to permanent changes in light cycles, they tend also to be sensitive to sporadic shifts in the light cycle. This sensitivity is reflected by changes in LD_{50} values, mixed function oxidase activity, circulating blood elements, etc.

The question of humidity already has been raised during the discussion of the control of inhalation chamber. contaminant levels. This environmental factor also may play a significant role in the establishment of infections in the respiratory apparatus though, to be sure, the humidity would have to be decreased well below 40%. In contrast to this, some investigators have reported that increases in relative humidity will result in a contaminant decrease in LD_{50} values.

Regarding caging conditions, Doull (8) notes that sensory input to the animal brain is capable of modifying hormonal and peripheral nervous system output. That input can be affected by the number of animals per cage, the nature of the cage, *i.e.* solid floor *uersus* wire mesh, the type of litter material, relative noise levels, activity patterns within the animal holding room, etc. As studies extend beyond the 2-week (subacute) type to the 2-year-plus (chronic) program, these factors take on greater significance regarding their potential impact on the outcome of the study.

Finally, diet control represents a key variable in the overall consideration of an experiment. A chemical analysis of the diet should be conducted to determine the level of metal contaminants, potential microsomal enzyme inducers, *i.e.* polycyclic aromatic hydrocarbons, carcinogen, *i.e.* aflatoxin, etc., protein and fat contents, etc. Depending on the goals of the study, these dietary variables could offer a significant contribution to any observed toxicity. Mention already has been made of the potential for interaction between dietary components and test materials. Such interaction could easily result in an inactivation of the test substance.

In short, by careful consideration and attention to the potential effects associated with these design variables, the investigator can expect to have a greater sense of confidence that those biological effects observed in conjunction with a toxicity study are indeed due to exposure to the test compound and not to an uncontrolled variable.

DATA EXTRACTION

GENERAL STUDIES

Any well-conducted toxicology study, whether directed primarily toward evaluation of neurotoxicity or not, should include observation and analysis of a number of non-neurologic parameters. The most commonly monitored of these indicators are whole body weight, organ weight, grossly observable functional signs, hematologic and clinical chemistry analytical data, and the observations recorded at the time of gross necropsy and subsequent histopathologic examination. The following is a superficial discussion of each of these general data areas.

For single-dose probe studies, body weights generally need be taken only at the beginning and end of the experiment. For subchronic studies of approximately 2 week duration, body weights should be taken on a daily basis; for subchronic studies of the order of 3-month duration, weekly body weights should be recorded. Beyond the 6th month of test for a chronic study (*i.e.*, 2 years/lifetime in rodents),

biweekly or monthly body weight determinations should be adequate. Although comparison of control and treatment groups by Student's *t* test **(29,** 30)... evaluation of mean group body weights at selected time intervals on test. . . is most commonly encountered in the literature, a more sensitive test ***can*** be applied. A linear approximation of the body weight plotted ***versus*** the logarithm of the time on test should be constructed for each animal. The slope of each of these lines may be determined graphically. A mean slope and the standard error of that mean for each treatment group then can be caiculated. These mean slopes may be treated **as** normally distributed independent data, and the mean slopes of the weight gain functions for each treatment group may be compared by Student's *t* test. A significant reduction in the slope of the body weight gain function for the "test group" should be considered indicative of nonspecific systemic toxicity.

Both absolute organ weight data (simple wet weight of the carefully dissected organs) and relative organ weight data (absolute organ weights normalized for total body weight) are commonly used by toxicologic investigators. Considerable controversy persists over the relative merits of each method of expression. With adequate attention paid to data variance, they are probably equivalently sensitive (when sex is taken into account and over-**all** body weights and ages are roughly equivalent). Nonparametric statistical tests are ideally applied with either absolute or relative organ weight data ***(vide infra).***

In the particular case of the use of brain weight data **as** a measure of neurotoxicity, several comments are in order. First, the CNS has a limited capacity for reactive hyperplasia ***(vide infra),*** in contrast to the kidney, heart, etc.; therefore, increases in CNS weight are quite unlikely except in the case of neoplasia or massive edema, both of which would be detectable by gross or histopathologic examination. Second, the weight of the brain is generally taken to be one of the most nearly constant of **all** the organ weights; second only to total body weight in its frequency of use, the brain weight is a basis for normalization of other organ weights. Third, the CNS apparently replaces many of the neuronal structures, lost through whatever process, with connective tissue (glial cells); the likelihood of detecting subtle changes by determination of gross weight is therefore very low. Fourth, organ weight is by its very nature a relatively widely distributed variable and so is not likely to be as sensitive an indicator of toxicity as are either functional or morphologic evaluation.

Measurement of hematologic parameters generally is taken to be a sensitive indicator of both nonspecific cytotoxic effects on rapidly dividing cells and of disturbances of mature cell morphology and physiology (*i.e.* typical erythrocyte biconcave conformation, osmotic fragility). It should be remembered that general anemias develop and progress only relatively slowly, this being the particular case with agents that are only weakly active. Reactive hyperplasia of the bone marrow (considerable reserve exists) may delay the appearance of, or completely ablate, the toxic response if only peripheral blood analyses are performed; examination of bone marrow is indicated in any toxicology study wherein hematotoxicity is suspected.

Total and differential leukocyte counts may indicate only the intercurrent state of the subject's overall health. Interpretation of these data should be the exclusive province of a hematologist trained and experienced with the species under study. These hematologic data very rarely can be considered pathognomonic, but rather they provide an opportunity to identify qualitatively the scope of a toxic response. In no cases should highly sophisticated statistical hypothesis **tests** be applied to these data; simple Student's *t* test or analysis of variance **(29,** 30) may be performed when $n \geq 30$ and the following are kept in mind:

- Changes in these parameters are highly nonspecific.
- There is considerable variation within what is considered the normal range.
- There is considerable quantitative uncertainty inherent in the methods of determination

Most grossly observable signs do not fulfill the criteria either of objectivity or of independence, the latter of which is required for statistical analysis. Compromise is struck in the case of grading (for severity) of histopathologic lesions and staging (for severity) of grossly observable signs. Because the intervals separating the discrete classifications are not uniform, only simple nonparametric statistical tests should be applied **(26).** The principal weakness of such scoring systems is classically demonstrated in the fact that there exists no uniformity of opinion among clinicians and/or pathologists as to the precise dividing lines between the classes/grades/stages. The best course is to restrict evaluation of these data to the judgment of the trained and experienced clinical professional and abandon statistical testing of data on grossly observable functional signs.

Analysis of the following clinical chemical parameters may be considered a standard diagnostic profile for descriptive toxicology:

- Blood glucose concentration
- Blood urea nitrogen concentration
- Serum glutamate-pyruvate transaminase (SGPT) activity

- Serum glutamate-oxaloacetate transaminase (SGOT) activity
- Creatine phosphokinase activity
- Lactate dehydrogenase activity
- Alkaline phosphatase activity
- Serum (or plasma) bilirubin concentration
- Serum cholesterol concentration
- Serum triglyceride concentration
- Serum (or plasma) inorganic phosphorus concentration
- Serum calcium concentration
- Serum potassium concentration
- Serum sodium concentration
- Serum chloride concentration
- Serum total protein concentration
- Serum uric acid concentration
- Urine specific gravity
- Urine protein concentration
- Microscopic examination of urine for cells/casts.

Each of the above tests is intended to survey the functional status of a particular organ system, but vanishingly few are so specific. For instance, both transaminase activity assays (SGOT and SGPT) are taken to reflect the hepatotoxicity of an agent. However, both may be elevated by result of traumatic injury, myocardial infarction, or by acute or chronic diseases of heart, muscle, and liver tissues. Other tests are exquisitely sensitive to the recent prior state of the subject; glucose and triglyceride assays are normally performed on 18-h-fasted subjects. Some variables exhibit fairly narrow "normal" ranges (*i.e.* ±5% for serum calcium concentration), whereas others vary widely (*i.e.* ±60% for serum triglyceride concentrations) **(3).** It is the evaluation of the complex, but coherent, sum of all of these data which forms the basis for interpretation; individual atypical values should not be overemphasized. **A** discussion of the significance of the results of each of these tests in both healthy and unhealthy individuals can be found in ***Documenta Geigy (7).***

As with any definitive toxicology experiment, a detailed gross dissection with all notable findings recorded at the time of necropsy is absolutely required. Here, **as** with other details of protocol, a balance must be struck in the judgment of the investigator between a comprehensive, but superficial, necropsy and an exhaustive, even semimicroscopic, dissection of specific organs and tissues.

SPECIALIZED STUDIES

Although adequate for most histopathologic evaluations in general toxicology studies, conventional histologic techniques for use of the light microscope are generally inadequate for determination of subtle neuropathologic changes. Spencer and Bischoff **(31)** have documented the preferred histologic preparation protocol ***(see also Chapter 50).***

Physioiogical abnormalities in the absence of morphological change have been documented. Such is the case when animals are exposed to acutely narcotizing concentrations of volatile low molecular weight hydrocarbon solvents. These **are** apparently entirely pharmacologic/biochemical events and are presumably completely reversible. These functional decrements in the absence of histopathology are customarily associated with acute overexposure, although the signs may recur frequently with repeated overexposure (and thus may mimic a chronic intoxication). Finally, very little direct evidence can be brought to bear on the comparative sensitivities and specificities of morphologic *versus* functional parameters.

Behavioral toxicology (a subunit of neurotoxicology) can be divided principally into two classes of studies. The first includes investigations of an ethological character in which the amount and qualitative type of spontaneous motor activity is monitored. The advantages of these studies are that they require no special equipment (given the disposition of the investigator to catalog behaviors manually), the variables are relatively easy to measure, the behaviors are spontaneous, and the tests have face validity **(23).** However, these tests are relatively nonspecific; the data are heavily influenced by the experimental environment (equipment, biorhythms of the animals, and the presence of distractions) and the age, health, nutritional status, and prior status of the subject. Furthermore, the face validity of the test is in question; a variety of agents (*i.e.* chlordecone, scopolamine) may produce increases in activity in one situation and decreases in activity in another **(22).** A particular behavioral change very rarely is diagnostic of a specific neurotoxic event. The prudent investigator monitors more than one behavioral indicator and under more than one experimental control situation. More than perhaps in any other of the data areas considered in this discussion, the prudent investigator employs only "two-tailed tests" of the nonparametric type for evaluation of data from such behavioral toxicology studies.

The second principal class of behavioral toxicology research is the study of the disintegration of an operant conditioned response. "A great strength of operant conditioners lies in their ability to manufacture behavior to specifications, manipulating the way an animal's responses pay-off" (16). In these studies, a subject is "taught" by conventional methods of behavioral shaping and operant conditioning to present a particular behavior in a particular environmental situation. Following exposure to the putative toxic agent, the subject is presented with

the specific environmental condition, and performance of the anticipated conditioned response is documented and analyzed. Thompson and Moerschbaecher **(33)** have shown that the slope of the learning curve for an operant condition may be a more sensitive indicator of behavioral toxicity than is disintegration of the pre-established condition after exposure to the putative neurotoxin. This observation is equivalent to the more sensitive evaluation of the time-dependent body-weight-gain function **as** discussed earlier.

Maximum nerve conduction velocity is useful for detecting the effects of certain neurotoxic agents. This test is most sensitive for identification of agents which preferentially damage the largest and most heavily myelinated motor fibers. The sensitivity of the test decays in inverse proportion to this specificity. That is, if large diameter nerve fibers are not affected by a toxic agent, regardless of size or degree of myelination, the maximum nerve conduction velocity as monitored by recording muscle contraction, would be expected to be affected very little, if at all. Early neurotoxic damage produced by most axonal neurotoxins (hexacarbons excepted) would go undetected by determination of maximum nerve conduction velocity, whereas such a test would be chosen to detect neurotoxic damage produced by agents which primarily damage the myelin sheath. In cases where neurotoxicity of the axonal type is suspected, determination of the amplitude of the sensory nerve action potential is indicated (17).

STATISTICAL TESTS

In studies of the effects of chemical agents on physiologic function, it is possible (and desirable) to use the "within subjects" design. With employment of the so-called Latin-square full-crossover design (20), each subject can serve as its own control in the study of partially or completely reversible neurologic decrements. In studies of irreversible functional decrement or in studies of morphologic damage (requiring the sacrifice of the test animal following treatment), the "between-subjects" design is required. In either case, a sufficient sample should be drawn to control cost effectively the probability of committing a type II statistical error (a failure to identify an experimental difference as real). The error variance (random noise) may be quite different for each of the above designs, even when the same agent is tested a pilot study (typically of about 10 animals in each of the control and treatment groups in the between-subjects design, or **20 animals** total in the within-subjects design) **will** aid greatly in estimating the population variances, which can, in turn, be used to estimate the optimum number of subjects to be studied in the final experiment **(30).** The diligent investigator should always consider a pilot study when dealing with a new variable (either independent or dependent). No study should ever be undertaken which does not include a rigorous concurrent control group (untreated or sham-operated) against which the subjects treated with the test agent are compared.

As a practical guidline in the selection of appropriate statistical tests, one can apply well L. E. Moses' "Principle of the Blunt ***Ax"*** (21). "If the ax chopped down the tree, it was sharp enough." Freely paraphrased, this can be taken to say that statistical overkill is almost invariably counterproductive.

Since many, if not most, neurotoxicologic end points are non-normally distributed, since aberrations of hypothesis testing are introduced by such non-normally distributed data, and since the nonparametric statistical tests do not assume normally distributed data, one should select from the cadre of nonparametric statistical tests for hypothesis testing in neurotoxicology. In general, the nonparametric tests are less powerful than are their parametric counterparts; said simply, one is less likely to discover a subtle difference by use of nonparametric hypothesis test than by use of an equivalent parametric method. The sacrifice, however, is mandated by the non-normally distributed data.

When *"n"* **530,** the simplest test is best. The Fischer Exact Probability Test (from a "so-called" 2 x **2** contingency table) is fully explained in ***Nonparametric Statistics (26).*** Since, in most cases, the problem to be examined is one of events which differ in kind rather than in magnitude (that is, "were the subjects normal or not normal?" rather than "were exposed subjects less able to perform a task than were the unexposed?"), the Fischer Exact Probability Test is the most appropriate to apply. When *"n"* exceeds **30,** the chi-square test should be used; the single excepting qualification is given in ***Nonparametric Statistics. (20):***

> Note: When the variable under scrutiny is the presence or absence of a phenomenon which by its very presence (neurohistopathology) is incontrovertible proof of effect, then the use of any statistical analysis is unnecessary (and undesirable).

For statistical evaluation of data for variables along a scale of relative intensities (*i.e.*, a > b > c, and a, **b,** c ≠ 0), the Mann-Whitney U-test is an excellent choice. The test takes full advantage of all the information from ordinal scaled data and is, therefore, exceptionally powerful among nonparametric tests. Among the data for which the U-test might be applied are the quantitative morphologic

parameters suggested by Norton (22), nerve conduction velocities, or sensory action potential amplitudes.

There is no magic formula for determination of an ideal sample size for experimental test. For a given level of power in any statistical test, there is an inverse relationship between the intrinsic variation in the measured variable and the sample sizes of control and treatment groups which are required to discriminate a difference of any given magnitude. In the absence of pilot trial data *(vide supra),* the purely apocryphal rule of thumb, $n = -30$, may be applied; the warning to be sounded is that such arbitrary experimental design may seriously compromise the power of any hypothesis test.

SIGNIFICANCE OF ANIMAL TESTING

The proper application of animal testing data to the human situation begins with distinctions among the various terms which classify the data base. There is almost universal disagreement on the exact meanings of the words *toxicity, hazard, risk,* and *safety;* but, for the present purpose, these definitions are offered.

Toxicity is the extrinsic property of a particular material to cause damage (morphologic or functional) to living organisms. Prefixes to the word more narrowly define the target—*i.e.* neurotoxicity, nephrotoxicity, hepatotoxicity. Tqxicity is a nominal scale; a particular toxic capacity either exists or it does not. However, the magnitude of toxic response is assumed to be straightforwardly dependent upon the magnitude of exposure.

Hazard is the seriousness of the particular toxic response. For instance, a material which is capable of causing only transient CNS depression might be stipulated to be less hazardous than a material which is capable of producing irreversible paralytic neurotoxicity.

Risk is the complex product of toxicity, exposure, sensitivity of the target, potency of the agent, and the coincidence of other environment circumstances which may alter the agent or target. **Risk** is the practical likelihood (finite probability) that a particular agent will cause a particular undesirable effect of a particular magnitude in a particular organism at a particular exposure level under a particular set of environmental circumstances. Each of these particulars can be generalized to a set (*i.e.* of agents or of environmental circumstances) or to **a** range (*i.e.* of exposure levels).

Safety, in the common usage of the word, is an absolute term describing a state in which either the agent is absolutely ineffective (nontoxic) or the target is absolutely insensitive (immune). If Paracelsus were presumed to be correct when he wrote *"All* substances are poisons; there is none which is not a poison," then safety is achievable only through "zero exposure." Restated, some nonzero risk, however immeasurably small, is associated with a nonzero exposure to a toxic agent.

The "toxicity test" is conducted—usually including high and low dose exposure levels, for the purpose of qualitatively identifying the entire set of deleterious effects which an agent is capable of causing in a given set of exposure circumstances. In its most complex form, the toxicity test may also attempt to demonstrate a "no discernable effect level" which is presumed to represent a combination of the effective biological threshold and the minimum detectable limit (*i.e. sensitivity)* of the test.

Of the several elements of "risk," two are the principal foci of animal testing, namely, toxicity and potency. The outcome of toxicity test determines whether or not the agent is capable of producing a toxic response at any dose under any circumstances. Such tests are commonly referred to as *hazard evaluation bioassays* by virtue of the fact that they are directed toward a particular kind of toxic response (hazard).

The so-called "safety test" (more appropriately named "relative safety test") is more complex and may employ several dose/exposure concentrations and schedules. The principal purpose of the test is to focus resources on the suppression of the minimum detectable limit (sensitivity) of the test and the consequently improved certainty of identification of the effective threshold. The safety test determines potency.

The absolute value of the empirically determined threshold and the confidence interval on that value may then be permuted with what is known or believed of the relative seriousness of the effect, of the sensitivities of the target organisms, and of the environmental circumstances of encounter to estimate relative risk.

REFERENCES

1. BALAZS T Measurement of Acute Toxicology. In: *Methods in Toxicology,* edited *by G.* E. Paget. Blackwell Scientific Publications, Oxford, 1970.
2. BALAZS T: Assessment of the Value of Systemic Toxicity Studies in Experimental Animals. In: *Advances in Modern Toxicology, Vol. 1, New Concepts in Safety Evaluation,* edited *by* M. A. Mehlman, R. E. Shapiro, and H. Blumenthal. Hemisphere Publish-

ing Corporation, Washington, D. C., 1976.
3. BIO-SCIENCE LABORATORIES: The Bio-Science Handbook, Specialized *Diagnostic* Laboratory Tests. Bio-Science Laboratories, Van Nuys, Calif, 1975.
4. BRODIE BB: Distribution and Fate of Drugs; Therapeutic Implications. In: Absorption and Distribution of Drugs, edited by T. B. Binns. Williams & Wilkins, Baltimore, 1964.
5. CAVANAGH JB: Peripheral neuropathy caused by chemical agents. CRC Critical Reviews in *Toxicology* 2,365, 1973.
6. CONNEY *AH,* SCHNEIDERMAN K, JACOBSON M, KUNTZMAN R Drug-induced changes in steroid metabolism. Annals of the New York Academy of Sciences 123, 98,1965.
7. *Documenta* Geigy, Scientific Tables, ed. 7, edited by K. Diem and C. Lentner. J. R. Geigy S. A., Basle, 1970.
8. DOULL J: Factors Influencing Toxicology. In: Toxicology: The Basic Science of Poisons, edited by L. J. Casarett and J. Doull. Macmillan Publishing Company, New York, 1975.
9. FOUTS JR, HART LG: Hepatic drug metabolism during the perinatal period. Annals of the New York Academy of *Sciences* 123,245, 1965.
10. GRICE H, MUNRO IC, WIBERG GS, HEGGTVEIT HA: The pathology of experimentally-induced cobalt cardiomyopathies. A comparison with beer drinkers cardiomyopathy. Clinical Toxicology 2,273, 1969.
11. GRIFFITH JF Inter-laboratory variations in the determination of acute oral LD_{50}. *Toxicology* and Applied Pharmacology 6, 726,1964.
12. HAWKINS JE, LAURIE MH: The ototoxicity of streptomycin. Annals of Otology, Rhinology, and *Larynogology* 61,789,1952.
13. HAYES WJ Jr.: The 90-dose LD_{50} and chronicity factor as measures of toxicity. Toxicology and Applied Pharmacology 11, 327, 1967.
14. HAYES WJ Jr.: Tests for Detecting and Measuring Long-Term Toxicity. In: Essays in Toxicology, Vol. 3, edited by W. J. Hayes, Jr. Academic Press, New York, 1972.
15. JONDORF WR, MAICKEL RP, BRODIE BB: Instability of newborn mice and guinea pigs to metabolize drugs. Biochemical Pharmacology 1, 352, 1959.
16. LATIES VG: How operant conditioning can contribute to behavioral toxicology. Environmental Health Perspectives 26, 29, 2978.
17. LeQUESNE PM: Clinical expression of neurotoxic injury and diagnostic use of electromyography. Environmental Health Perspectives 26, 89, 1978.
18. LU FJ, JESSUP DC, LAVALLEE A Toxicity of pesticides in young versus adult rats. Food and Cosmetic Toxicology 3, 591, 1965.
19. MENDELL JR, SAIDA K, GANANSIA MF, JACKSON DB, WEISS H, GARDIER RW, CHRISMAN C, ALLEN N, COURI D, O'NEILL J, MARKS B, HETLAND L Toxic polyneuropathy produced by methyl n-butyl ketone. Science 185, 787, 1974.
20. MORRISON JK, QUINTON RM, REINERT H: The Purpose and Value of LD_{50} Determinations. In: Modern Trends in Toxicology, Vol. I, edited by E. Boyland and R. Goulding. Appleton-Century-Crofts, London, 1968.
21. MOSES LE: Principle of the Blunt Ax. In: Biostatistics, p. 62, edited by A. Goldstein. Macmillan Publishing Company, New York, 1964.
22. NORTON S: Is behavior or morphology a more sensitive indicator of central nervous system toxicity? Environmental Health Perspectives 26, 21, 1978.
23. REITER L Use of activity measures in behavioral toxicology. Environmental Health Perspectives 26, 9, 1978.
24. RONA G, CHAPPEL CI: Pathogenesis and Pathology of Cobalt Cordiomyopathy. In: Recent Advances in Cardiac Structure *and* Metabolism, edited by E. Bajusz and G. Rona. University Park Press, Baltimore, 1973.
25. RUMKE CL Some Limitations in Animal Tests. In: Evaluation of Drug Activities Pharmacokinetics, Vol. I, edited by D. R. Laurence and A. L. Bacharach. Academic Press, London, 1964.
26. RUNYON RP: Non-Parametric Statistics, A *Contemporary* Approach. Addison-Wesley Publishing Company, Reading, Mass., 1977.
27. SCHOENTAL R, MAGEE PN: Chronic liver changes in rats after a single dose of lasiocarpine, a pyrrolizidine (senecio) alkaloid. Journal of Pathology and Bacteriology 74, 305, 1957.
28. SETNIKAR I, MAGISTRETTI MJ: The toxicity of central nervous system stimulants in rats of different ages. Proceedings of the European Society for Drug Toxicity 4, 132, 1964.
29. SNEDECOR GW, COCHRAN WG: Statistical Methods, Ed 6. Iowa State University Press, Ames, Iowa, 1967.
30. SOKAL RR, ROHLF FJ: Biometry. W. H. Freeman and Company, San Francisco, 1969.
31. SPENCER PS, BISCHOFF MC: Contemporary Morphological Techniques for Evaluating Peripheral Nerves. In: Management of *Peripheral* Nerve Problems, edited by O. K. Omer and M. Spinner. Saunders, Philadelphia, 1980.
32. SPENCER PS, SCHAUMBURG HH: Distal axonopathy: one common type of neurotoxic lesion. *Environmental* Health Perspectives 26, 97, 1978.
33. THOMPSON DM, MOERSCHBAECHER JM: Operant methodology in the study of learning. Environmental Health Perspectives 26, 77, 1978.
34. WEIL CS. CARPENTER CP, WEST JS, SMYTH HR J R Reproducibility of single oral dose toxicity testing. American Industrial Hygiene Association Journal 27,483, 1966.
35. WEIL CS, McCOLLISTER DD: Relationship between short and long term feeding studies in designing an effective toxicity test. Journal of Agricultural Food Chemistry 11,486,1963.
36. WEIL CS, WRIGHT GJ: Intra- and interlaboratory comparative evaluation of single oral test. Toxicology and Applied Pharmacology 11,378,1967.
37. WHEELER AG, DANSKY D, HAWKINS HC, PAYNE HG, WEIKEL JH: A toxicologic and hematologic evaluation of cyclophosphamide (CYTOXAN) in experimental animals. Toxicology and Applied Pharmacology 4, 324, 1962.
38. YEARLY RA, BENISH RA: A comparison of the acute toxicities of drugs in newborn and adult rats. Toxicology and Applied Pharmacology 7,504, 1965.