Predicting Neurotoxicity and Behavioral Dysfunction From Preclinical Toxicologic Data

May 1986

Prepared for

CENTER FOR FOOD SAFETY AND APPLIED NUTRITION
FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
WASHINGTON, D.C. 20204

under

Task Order #3
Contract No. FDA 223-83-2020
PREDICTING NEUROTOXICITY AND BEHAVIORAL 
DYSFUNCTION FROM PRECLINICAL TOXICOLOGIC DATA

May 1986

Prepared for
CENTER FOR FOOD SAFETY AND APPLIED NUTRITION
FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
WASHINGTON, D.C. 20204

under
Task Order #3
Contract Number FDA 223-83-2020

edited by
Richard W. Leukroth, Jr.

LIFE SCIENCES RESEARCH OFFICE
FEDERATION OF AMERICAN SOCIETIES
FOR EXPERIMENTAL BIOLOGY
9650 Rockville Pike
Bethesda, Maryland 20814
FOREWORD

The Life Sciences Research Office (LSRO), Federation of American Societies for Experimental Biology (FASEB), provides scientific assessments of topics in the biomedical sciences. Reports are based upon comprehensive literature reviews and the scientific opinions of knowledgeable investigators engaged in work in relevant areas of biology and medicine.

This report was developed for the Food and Drug Administration (FDA), Center for Food Safety and Applied Nutrition (CFSAN) in accordance with the provisions of Task Order #3 of Contract No. 223-83-2020. The report was prepared and edited by Richard W. Leukroth, Jr., M.S., Associate Staff Scientist, LSRO, FASEB, under the direction of an ad hoc Expert Panel on Predicting Neurotoxicity and Behavioral Dysfunction from Preclinical Toxicologic Data.

The report is presented in two parts: 1) the deliberations of the ad hoc Panel (Part A) and 2) the Proceedings of a symposium on this topic (Part B). The ad hoc Panel was comprised of seven invited speakers, who delivered presentations at the Symposium, and seven additional participants who participated as members of an invited symposium discussion panel. Panel members were chosen for their qualifications, experience, and judgment with due consideration for balance and breadth in the appropriate professional disciplines. Members of the ad hoc Panel and others who assisted in the preparation of the report are identified in Part A (see p.55-57). The ad hoc Panel reviewed and evaluated the scientific literature as well as the proceedings of the Symposium in executive session and developed its findings and conclusions on the basis of these discussions.

As announced by FASEB and FDA in the Federal Register, the ad hoc Panel held an open symposium on September 30, 1985 entitled Predicting Neurotoxicity and Behavioral Dysfunction from Preclinical Toxicologic Data. The symposium provided an opportunity for interested organizations and persons to appear before the ad hoc Panel to make oral presentations of data, information, and views on predicting neurotoxicity and behavioral dysfunction from preclinical toxicologic data. The symposium proceedings contain a series of invited review papers presented at the open symposium, and information and views provided by scientists who attended the symposium. Individuals and organizations who made oral presentations or submitted written materials are identified in Part B of this report (see Part B, p.XI).

The symposium presentations and the ad hoc Panel's evaluation of the presentations and other information, data, and views were made independently of FDA and any other group, governmental or nongovernmental. The speakers, Panel members,
and LSRO accept responsibility for the accuracy of their respective portions of this report; however, listing these individuals in Part A (see p.55-56) does not imply that they specifically endorse each study conclusion. This report was reviewed and approved by the LSRO Advisory Committee (which consists of representatives of each constituent Society of FASEB) under authority delegated by the Executive Committee of the Federation Board. Upon completion of these review procedures, the report was approved and transmitted to FDA by the Executive Director, FASEB.

While this is a report of the Federation of American Societies for Experimental Biology, it does not necessarily reflect the opinion of each individual member of the FASEB constituent Societies.

May 30, 1986
Date

Kenneth D. Fisher, Ph.D.
Director
Life Sciences Research Office
SUMMARY

This report provides a review of information on neurotoxicity and behavioral dysfunction and presents conclusions and recommendations reached by an ad hoc Expert Panel. The report describes available basic research, test methods for detecting neurobehavioral toxicity, and the need for expanding conventional applied toxicology testing to include indices of behavior and nervous system function. Included is a review of regulatory approaches to neurotoxicity testing, the availability of neurotoxicity information from conventionally applied toxicology protocols, and an evaluation of the adequacy of this information. The report is presented in two parts: 1) the evaluative findings of the Expert Panel (Part A) and 2) the proceedings of an open symposium on Predicting Neurotoxicity and Behavioral Dysfunction from Preclinical Toxicologic Data (Part B).

From the public viewpoint, the objective of applied toxicology is to minimize risks at acceptable costs. Accumulating evidence suggests that there is no such thing as a perfect testing screen. The toxicity of some test substances may evade detection only to be discovered from basic research, clinical observation, or epidemiological data. For the issue of neurobehavioral toxicity the Panel concluded that it is helpful to acknowledge information from multiple sources.

There is a need for the FDA to improve the manner in which it currently assesses the potential for neurobehavioral toxicity from conventional applied toxicology mechanisms. The information presently available is incomplete and misleading with deficiencies in study protocols, data collection methods, and identification of appropriate endpoints of neurotoxicity. Current guidelines should be modified to ensure that the limited information potentially available from toxicology protocols is not lost and to enhance uniformity in recording and reporting of functional, neuropathological, biochemical, and behavioral information. The Panel concluded that current FDA test guidelines require that the cumulative findings of such information be reported in a separate statement describing the structural and functional integrity of the nervous system.

In developing a strategy for detecting neurotoxicity, the Panel acknowledged that tests are available to aid in predicting neurobehavioral toxicity and that the choice of test procedures for initial neurotoxicity screening should sample a wide range of neurologic functions. The Panel recommended a two-tiered strategy coinciding with the level of inquiry or FDA concern.

The primary tier would function to identify neurotoxic potential by incorporating in early stages of toxicity testing (e.g., acute, subacute, and subchronic protocols) a functional observation battery, a measure of integrated function (e.g., motor
activity, learned behavior), and available information on structure-activity relationships. This initial stage of neurotoxicity testing identifies the range of central nervous system involvement and the categories of neurologic function to be tested in the second tier. The Panel concluded that positive findings in primary tier testing provide sufficient justification to require additional testing.

The secondary tier would employ a series of supplemental tests to: 1) further characterize the nature of neurotoxic effect; 2) determine the extent of central nervous system involvement; and 3) develop dose/time effect relationships in the advanced stages of conventional toxicology testing (i.e., long-term, chronic, and reproduction and development). Second tier testing would include measures of motor and sensory function as well as tests of cognitive abilities. The extent of testing is dependent upon early test findings and their interpretation by the petitioner and FDA, as well as the vested interests of the petitioner.

Any battery of screening tests should be expected to change with time based on new research data, clinical population studies, and a re-definition of cost effectiveness. There is a need to maintain flexibility when preparing test guidelines for neurobehavioral toxicity anticipate additional and/or improved test methods. Changes to excepted testing practices should not be made for trivial reasons, and when changes are instituted there should be consistency among regulatory agencies.
# TABLE OF CONTENTS

## PART A

### Report of the Expert ad hoc Panel

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreword</td>
<td>iii</td>
</tr>
<tr>
<td>Summary</td>
<td>v</td>
</tr>
<tr>
<td>I. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>A. Background</td>
<td>1</td>
</tr>
<tr>
<td>B. Scope of the Study</td>
<td>2</td>
</tr>
<tr>
<td>C. Technical Approach</td>
<td>3</td>
</tr>
<tr>
<td>D. Information Reviewed by the Panel</td>
<td>4</td>
</tr>
<tr>
<td>1. Symposium related information</td>
<td>4</td>
</tr>
<tr>
<td>2. Information from other study groups</td>
<td>4</td>
</tr>
<tr>
<td>3. Review of toxicology testing requirements</td>
<td>6</td>
</tr>
<tr>
<td>II. General Considerations</td>
<td>13</td>
</tr>
<tr>
<td>A. Behavioral Changes as Indices of Neurotoxicity</td>
<td>13</td>
</tr>
<tr>
<td>B. FDA Approach to Behavioral and Observational Data</td>
<td>15</td>
</tr>
<tr>
<td>C. Evaluating Nervous System Function in Conjunction with Good Laboratory Practice Guidelines</td>
<td>16</td>
</tr>
<tr>
<td>1. Study design</td>
<td>17</td>
</tr>
<tr>
<td>2. Collection of quantity data</td>
<td>20</td>
</tr>
<tr>
<td>3. Data evaluation</td>
<td>22</td>
</tr>
<tr>
<td>III. Availability of Behavioral and Neurotoxicity Information from Conventional Testing</td>
<td>25</td>
</tr>
<tr>
<td>A. Types of Information Available</td>
<td>25</td>
</tr>
<tr>
<td>1. Observational information</td>
<td>25</td>
</tr>
<tr>
<td>2. Reproduction and developmental information</td>
<td>26</td>
</tr>
</tbody>
</table>
3. Clinical and biochemical laboratory studies ........................................... 28
4. Pathological information ................................................................. 28

B. Reporting of Available Information ................................................. 28

IV. Adequacy of Neurotoxicity Information from Conventional Testing ........... 31
   A. Adequacy of Information as Determined by Studies of Well-Established Neurotoxicants ................................................................. 31
   B. Adequacy of Information as Determined by the Study Model .................. 32
   C. Adequacy of Information from Behavioral Indices ................................ 33

V. Strategy for Applied Neurobehavioral Toxicity Testing .......................... 35
   A. Levels of Inquiry in Applied Neurobehavioral Toxicology ...................... 35
      1. Selection of study protocol .......................................................... 36
      2. Selection of animals for study ...................................................... 36
   B. Tier Approach ................................................................................. 37
   C. Primary Tier Testing ................................................................. 37
      1. Screening for the neural component of concern .............................. 38
      2. Quantitative basis for minimizing observer bias .......................... 39
      3. Structure-activity relationships .................................................. 40
   D. Secondary Tier Testing .................................................................... 40

VI. Summary of Conclusions and Recommendations ................................... 43

VII. Literature Cited .............................................................................. 47

VIII. Study Participants ........................................................................... 55
PART B
Proceedings of the Symposium

Page

I. Symposium Overview ........................................... I-1
   Donald E. Hutchings, C. Wayne Callaway,
   Thomas J. Sobotka
   Literature Cited ............................................ I-8

II. The Use of Traditional Toxicologic Data in Assessing Neurobehavioral Dysfunction .......... II-1
   Marshall Steinberg
   Abstract ..................................................... II-1
   A. Introduction ............................................. II-3
   B. Regulatory Guidelines for Toxicity Testing ... II-3
   C. Laboratory Observations of Study Animals ..... II-11
   D. Animal Models ........................................... II-17
   E. Scientific Interrelations ............................... II-18
   F. Data Evaluation .......................................... II-20
   G. Summary ................................................. II-20
   Open Discussion ............................................ II-21
   Literature Cited ............................................ II-30

III. Complexity of Neurotoxicological Assessment ....... III-1
    William F. Sette
    Abstract .................................................... III-1
    Introduction ............................................... III-3
    A. Federal Mandates for Neurotoxicity Testing
       and Risk Assessment .................................. III-4
    B. Hazard Identification ................................. III-7
    C. Hazard Assessment ..................................... III-10
    D. Summary ............................................... III-14
Mechanisms of CNS Injury in Behavioral Dysfunctions

Richard B. Mailman

Abstract

Introduction

Overview of the Sites and Mechanisms of Neurotoxicity

A. Chemical Messengers

B. The Chemical Architecture of the Nervous System

Use of Neurochemical Methods

A. Neurochemical and Biochemical-Strategies Proposed for Evaluating Neurotoxicity

B. Changes in Concentration or Content of Neurotransmitter

C. Other Indices of Neurotransmitter Function

D. Caution in the Use of These Methods

E. The Application of Biochemical Methods in Neurotoxicology

The Use and Abuse of Neurochemistry in Predicting Neurotoxicology

A. The Case of FD&C Red No. 3

B. The Case of MPTP

Future Role of Mechanistic Studies in Predictive Neurotoxicology

A. Improved Methods of Prediction and Detection

B. Summary
Open Discussion ........................................ IV-23
Literature Cited ........................................ IV-29

V. Behavioral Indices of Neurotoxicity:
   What Can Be Measured? ............................... V-1
   Hugh A. Tilson

   Abstract ............................................... V-1

   Introduction .......................................... V-3

   A. Status of Neurobehavioral Toxicology for
      Regulatory Purposes ............................... V-3

   B. Rationale for Use of Behavioral Testing
      Procedures in Toxicology ......................... V-4

   Classification of Behavioral Approaches .......... V-5

   A. Definition of Behavior ............................ V-5

   B. Classification Methods of Behavioral
      Testing ............................................. V-7

   C. Testing for Signs of Exposure ................... V-8

   D. Behavioral Testing Schemas ....................... V-8

   Examples of Neurobehavioral Tests ................ V-11

   A. Tests of Motor Function .......................... V-11

   B. Behavioral Tests for Alterations in Sensory
      Processes ........................................... V-14

   C. Tests for Arousal or Reactivity ................ V-17

   D. Tests for Learning and Memory .................. V-18

   E. Schedule-Controlled Operant Behavior .......... V-25

   F. Naturally Occurring Behaviors ................... V-27

   G. Special Application of Behavioral
      Procedures ......................................... V-27

   Considerations in the Use of Behavioral
      Procedures ......................................... V-30
A. Mechanism of Action ......................... V-30
B. Defining an Adverse Effect .................. V-32
C. Functional Reserve .......................... V-32
D. Statistical Considerations .................. V-32
Summary ........................................... V-33
Open Discussion ............................... V-34
Literature Cited ............................... V-42

VI. Reliability, Sensitivity and Validity of Behavioral Indices of Neurotoxicity ........ VI-1
Charles V. Vorhees

Abstract ........................................ VI-1

Reliability of Behavioral Indices of Neurotoxicity ........................................ VI-3

A. Reliability of Behavioral Methods ........ VI-4
B. Sensitivity of Behavioral Methods ........ VI-29
C. Validity of Behavioral Methods .......... VI-38
D. Reliability of Other Indices of Behavioral Neurotoxicity .......................... VI-43

E. National and International Context of Neurotoxicity Testing .................... VI-50

F. Summary ..................................... VI-51

Open Discussion ............................... VI-52
Literature Cited ............................... VI-55

VII. Psychoneuroimmunology, New Approaches to Neurobehavioral Testing .......... VII-1
Nicholas P. Plotnikoff

Abstract ....................................... VII-1

A. Introduction ................................ VII-3
I. INTRODUCTION

This report addresses the need for collection of toxicity data on the nervous system, available methods, and the current status of applied research in the science of neurotoxicity and behavioral dysfunction. It presents a series of commissioned review papers presented at an open symposium and the evaluative findings of a Life Sciences Research Office (LSRO) ad hoc Panel on Predicting Neurotoxicity and Behavioral Disfunction from Preclinical Toxicologic Data.

A. BACKGROUND

There is increasing recognition within the scientific community that the endpoints of toxicology testing have expanded to include neurobehavioral toxicity (see Part B, Table I.1, p.1-2 and 3). New information from research on testing methods (Buelke-Sam et al., 1985; Geller et al., 1979; Tilson and Mitchell, 1984), advances in basic aspects of the neurosciences (Abelson et al., 1985), and experience derived from unforeseen human toxicities such as Ginger-jake (Morgan, 1982; Smith et al., 1930) and Spanish cooking oil (Smith and Spaulding, 1959) have sparked continued interest in and public concern for neurotoxicology. The number of toxicants known to influence nervous system function continues to grow (Anger and Johnson, 1985; Damstra, 1978; Norton, 1980; Spencer and Schaumburg, 1983a; Weiss, 1978) and concern for the effects of potentially neurotoxic chemicals is expanding (National Research Council, 1977a, 1986; Spencer and Schaumburg, 1983a; World Health Organization, 1986).

The decision by the Environmental Protection Agency (EPA) Office of Toxic Substances (OTS) to issue formal guidelines for neurotoxicity testing illustrates the impact of the concern for neurotoxicity on scientific policy and response from government agencies (U.S. Environmental Protection Agency, 1985a). The advent of concern for potential neurotoxic effects broadens the scope of activities that regulatory scientists must consider when making decisions regarding potential toxicity. Information and data on neurotoxic consequences, as with other endpoints of toxicity, must provide insight from the broadest and most complete database possible (National Research Council, 1984a; Sobotka, 1986).

The conventionally accepted study protocols in current use by regulatory agencies such as the Food and Drug Administration (FDA) and the EPA for the safety evaluation of various chemical substances have been modeled after the recommendations of the National Academy of Sciences (National Research Council, 1975). Risk assessments are dependent upon toxicity data derived
from protocols for acute, subchronic, chronic, teratology, and reproductive preclinical studies in laboratory animals. These data are used in the risk assessment process to provide government agencies with information to evaluate the potential effects of substances which may result from specific (e.g., prescription or registered drugs) and nonspecific (e.g., food additives) exposure within the general population. Although there are both similarities and disimilarities in the types of information needed to characterize toxicity within these exposure categories, the nature and extent of toxicity testing may vary based on the intended use of the substance and the degree of toxicity associated with the substance.

As an ongoing process, regulatory agencies evaluate recommended and required protocols for toxicology testing to assure that the test methods are providing data useful in the regulatory process. This review of neurobehavioral toxicology testing is particularly appropriate as this field has expanded rapidly in recent years. Advances in testing technology and results of research studies are providing a basis for developing criteria for safety evaluation that include greater emphasis on neurotoxicology testing and assessment of behavioral dysfunction. There is an opportunity to integrate research findings into regulatory applications within this discipline.

B. SCOPE OF THE STUDY

The FDA, Center for Food Safety and Applied Nutrition (CFSAN) is responsible for development of scientific criteria to establish the safety of new foods and food additives. The agency is also responsible for assuring the continued safety of the U.S. food supply in relation to trends in use of these food ingredients and the constantly evolving scientific criteria for safety evaluation (Food and Drug Administration, 1982). Under the terms of Task Order No. 3 of FDA Contract No. 223-83-2020, the Life Sciences Research Office (LSRO), Federation of American Societies for Experimental Biology (FASEB), was asked to review current scientific opinion and available information on the use of neurotoxicology and behavioral testing.

The goal of this study was to evaluate the use of conventional toxicology testing for predicting neurotoxicity and behavioral dysfunction and to identify the research needs for improving the reliability and comprehensiveness of the prediction of such effects in humans. The FDA sought expert opinion about the nature, extent, and feasibility of possible approaches to improving both the understanding and the use of these kinds of data for protecting public health. The FDA seeks to utilize such opinion as one source of information in its considerations of refining criteria for safety evaluation and providing future guidelines for neurotoxicology testing. The FDA outlined the scope of work for this LSRO study by posing the following questions:
1. To what extent do the various traditional toxicity tests (particularly reproduction, teratology, acute, subchronic, and chronic studies), carried out at exposure levels high enough to produce clear toxic effects, give information about the nature and scope of potential neurotoxicity or behavioral dysfunction? What particular endpoints in these traditional toxicity tests serve to indicate neuronal dysfunction?

2. To what extent do the traditional toxicity tests mentioned above fail to yield information about the nature and scope of neurotoxicity or behavioral dysfunction? What aspects of neurobehavioral toxicity would or might be missed by relying only on traditional toxicity tests? What type of neurobehavioral test battery would be necessary to complement traditional testing in order to supply this information?

3. If neuronal involvement is indicated by traditional toxicity tests, what type of test battery would be needed to characterize better the extent and nature of neuronal dysfunction?

C. TECHNICAL APPROACH

To examine these scientific issues, LSRO convened an ad hoc Expert Panel and organized an open symposium of the Panel entitled "Predicting Neurotoxicity and Behavioral Dysfunction from Preclinical Toxicologic Data". The symposium was followed by an executive session of the Panel. This technical approach provided a forum for the exchange of ideas among representatives of government, industry, and other interested parties, and offered opportunities for the presentation of views and discussion of concerns related to the goals of this study.

The ad hoc Expert Panel (symposium speakers and discussants) were selected on the basis of expertise in neurotoxicity, behavioral assessment, and related disciplines. The Panel provided expert investigative experience in areas such as experimental toxicology; molecular and cellular mechanisms of neurotoxicity; the role of morphological, genetic, and immunological aspects of the nervous system; the effects of dietary and/or environmental chemicals on neurological abnormalities and behavioral dysfunction; and, the application of epidemiological methods in detecting nervous system etiology. A listing of the Panel members is given on page 56.

In conjunction with FDA, LSRO identified six topics to be addressed in the commissioned papers presented at the open symposium on September 30, 1985. Each speaker prepared a manuscript that provided a literature review and invited open discussion on specific areas of applied toxicology testing for
neurotoxicity and behavioral dysfunction from both the LSRO Panel and other symposium attendees (see Part B, p.IX-2 and 3). The revised symposium papers and the symposium discussions are reported in Part B.

Following the symposium, the Panel met in executive session to address the specific questions in the scope of work. These and other related questions were addressed by the ad hoc Panel and are presented in Part A of this report.

D. INFORMATION REVIEWED BY THE PANEL

In formulating their evaluation of the FDA questions the Panel drew upon information provided by the symposium papers, the discussions which followed these papers, additional information provided to the Panel in response to a published request in the Federal Register (Food and Drug Administration, 1985), and other information contributed by individual Panel members during their discussions. The Panel also reviewed toxicology testing requirements for various regulatory agencies.

1. Symposium related information

The LSRO-commissioned literature reviews served as a primary resource for the Panel. These symposium presentations provided insight from the viewpoint of the toxicology screening laboratory (see Part B, Steinberg, p.II-1-32), the regulatory scientist (see Part B, Sette, p.III-1-22), the neurochemist (see Part B, Mailman, p.IV-1-36), the neurobehavioral toxicologist (see Part B, Tilson, p.V-1-54), the developmental neurotoxicologist (see Part B, Vorhees, p.VI-1-60), and the research scientist in immunology (see Part B, Plotnikoff, p.VII-1-23). Additional information and concerns were presented to the Panel by scientists who participated in open discussions at the symposium (see Part B, p.VIII-1-14 and IX-1-3).

Information submitted to the Panel in response to the Federal Register announcement was made available to the Panel prior to the Symposium. The Panel discussed these materials and gave consideration to them in the preparation of this report (Part B, Chapter IX-1).

2. Information from other study groups

The Panel drew upon reports of other groups which have addressed aspects of test strategies for detecting nervous system toxicity (Table 1). Of particular interest were the findings from the National Academy of Sciences (National Research Council, 1975, 1977a, 1984a), the Collaborative Behavioral Teratology Study (Buelke-Sam et al., 1985), and the forthcoming
<table>
<thead>
<tr>
<th>STUDY GROUP</th>
<th>YEAR</th>
<th>SUMMARY OF RECOMMENDATIONS FOR NEUROTOXICITY TESTING</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHO¹</td>
<td>1986</td>
<td>Recommended two levels of neurobehavioural testing: a) the primary level includes a functional observational battery and tests for motor activity; b) the secondary level includes specialized tests for intermittent schedules of reinforcement, sensory function, motor function, and cognition. Also included are improved neuropathological methods (World Health Organization, 1986).</td>
</tr>
<tr>
<td>NAS²</td>
<td>1986</td>
<td>Identified research needs to measure exposure using biological markers and better laboratory techniques (i.e., short term tests to identify neurotoxicants, and quantitative models for low dose extrapolation) and to study the predictive value of structure-function relationships (National Research Council, 1986).</td>
</tr>
<tr>
<td>NCTR³</td>
<td>1985</td>
<td>Recommended that behavioral teratology testing include functional tests of progeny for sensory systems, neuromotor development, locomotor activity, learning and memory, reactivity and/or habituation, and reproductive behavior (Kimmel et al., 1985).</td>
</tr>
<tr>
<td>NAS</td>
<td>1984</td>
<td>Neurobehavioural toxicity testing should include studies on function (both conditioned and unconditioned) and morphology (neuropathology) (National Research Council, 1984b).</td>
</tr>
<tr>
<td>NAS</td>
<td>1977</td>
<td>Recommended conditioned (i.e., schedule-controlled responding for food) and unconditioned behavior (i.e., circadian, spontaneous, motor activity) procedures (National Research Council, 1977a).</td>
</tr>
<tr>
<td>NAS</td>
<td>1975</td>
<td>Proposed a decision sequence for behavioral evaluation including motor activity; systematic observation; tests of sensory, motor, and complex learned functions; and behavioral teratology (National Research Council, 1975).</td>
</tr>
</tbody>
</table>

¹ World Health Organization  
² National Academy of Sciences  
³ National Center for Toxicological Research  
⁴ Organisation for Economic Co-operation and Development
report by the World Health Organization (1986). A listing of other symposia and conferences reviewed by the Panel in considering the questions posed by FDA are presented in Table I.1 (see Part B, p.I-2 and 3).

3. **Review of toxicology testing requirements**

Several regulatory agencies in the United States and in other countries have published guidelines for toxicology testing which either specify or imply testing of possible toxicant effects on the nervous system (see Part B, Sette, p.III-3). This information is summarized in Table 2. The EPA has developed testing protocols to address specific aspects of behavior and neurotoxicity of pesticides and other toxic substances (U.S. Environmental Protection Agency, 1985a). Japan and the United Kingdom require behavioral neurotoxicity testing, but have not identified specific test protocols for regulatory purposes.

Current guidelines for toxicity testing of food ingredients by the Food and Drug Administration have been published (Food and Drug Administration, 1982). The Panel reviewed required FDA study protocols (i.e., acute, short-term, subchronic, long-term as well as carcinogenicity, reproductive function, and teratology testing) for available data elements that relate to neurotoxicity and behavioral dysfunction. This information is summarized in Table 3.
Table 2. Available Regulatory Guidelines for Neurotoxicity Testing

<table>
<thead>
<tr>
<th>AGENCY</th>
<th>YEAR</th>
<th>DESCRIPTION OF THE TESTING GUIDELINES</th>
</tr>
</thead>
<tbody>
<tr>
<td>EPA$^1$</td>
<td>1986</td>
<td>Guidelines for developmental neurotoxicity include: a) behavioral observation of dams and offspring; b) evaluation of physical landmarks; c) tests for motor activity, auditory startle habituation, and Biel water maze paradigm; and d) neuropathology (U.S. Environmental Protection Agency, 1986).</td>
</tr>
<tr>
<td>EPA</td>
<td>1985</td>
<td>Guidelines for neurotoxicity include: a) functional observational battery; b) motor activity; c) neuropathology; d) neurotoxic esterase assays; e) schedule-controlled operant behavior; and f) peripheral nerve function (U.S. Environmental Protection Agency, 1985a).</td>
</tr>
<tr>
<td>MAFF$^2$</td>
<td>1985</td>
<td>Guidelines recommend reproduction and developmental protocols including the examination of behavioral observations (i.e., locomotion, learning, sensory function, and emotion) of postweaning progeny in Segment 2 and 3 studies (Japan, Ministry of Agriculture, Forestry and Fisheries, 1985).</td>
</tr>
<tr>
<td>FDA$^4$</td>
<td>1982</td>
<td>Available guidelines require the description of any behavioral abnormality, clinical signs of toxicity or pharmacological effects, and gross and histopathological evaluation of the nervous system (Food and Drug Administration, 1982).</td>
</tr>
</tbody>
</table>

---

1 U.S. Environmental Protection Agency  
2 Ministry of Agriculture, Forestry and Fisheries  
3 Department of Health and Social Security  
4 Food and Drug Administration  
5 Organisation for Economic Co-operation and Development
Table 3. Available Data Sets For Animal Observation, Behavior, and Pathology of Nerve Tissues in Some Current FDA Guidelines For Toxicology Testing (Food and Drug Administration, 1982)\(^1\)

<table>
<thead>
<tr>
<th>TEST</th>
<th>PROCEDURES RELEVANT TO THE NERVOUS SYSTEM(^2)</th>
<th>REPORTING REQUIREMENTS(^3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Acute LD(_{50})</td>
<td>14 day observation period. Record clinical observations frequently on day 1 and 2 times/day thereafter. Individual observations to be recorded for time of onset, intensity, and duration. Observations to include all toxicological and pharmacological signs. Weekly body weight. Necropsy and conditional microscopic examination of gross lesions.</td>
<td>Report all findings from clinical observations, necropsy, and histopathological examinations.</td>
</tr>
<tr>
<td>II. Short-Term</td>
<td>Record all toxicological and pharmacological signs daily including time of onset, intensity, and duration. Body weight, food, and/or water consumption weekly. Ophthalmological examination pre-administration and at termination for control and high dose animals. Necropsy and histopathological examination of the rodent brain; for nonrodents, nothing.</td>
<td>Report a description of all toxicological and pharmacological effects and abnormalities/animal/day of observation. Include behavioral and clinical abnormalities and pathological findings. An attempt should be made to correlate effects observed in-life with postmortem findings.</td>
</tr>
<tr>
<td>III. Subchronic</td>
<td>Record daily all toxicological and pharmacological signs including time of onset, duration, and intensity. Body weight, food, and/or water consumption weekly. Ophthalmological examination pre-administration and at termination for control and high dose animals. Examination of cranial cavity at necropsy. Histopathological examination of brain (3 levels), spinal cord (2 levels), eye, pituitary, and sciatic nerve with muscle.</td>
<td>Report a description of all toxicological or pharmacological effects and abnormalities/animal/day of observation. Include behavioral and clinical abnormalities and pathological findings. Correlate effects observed in-life with postmortem findings.</td>
</tr>
<tr>
<td>TEST</td>
<td>PROCEDURES RELEVANT TO THE NERVOUS SYSTEM</td>
<td>REPORTING REQUIREMENTS</td>
</tr>
<tr>
<td>------</td>
<td>------------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>IV.</td>
<td>Record animal observations daily throughout the test. Body weight and food consumption weekly for 13 weeks and monthly thereafter. Any behavioral abnormality or any clinical sign of toxicity or pharmacological effects, moribundity, and mortality, should be recorded. Such observations are usually taken at time of dosing. Ophthalmoscopic examination pre-administration and at termination for control and high dose animals. Examination of the cranial cavity at necropsy. Histopathological examination of brain (3 levels), spinal cord (2 levels), eye, pituitary, and sciatic nerve with muscle.</td>
<td>Report individual data for all descriptions of toxicological and pharmacological effects and abnormalities/animal/day of observation. Include behavioral and clinical abnormalities, and pathological findings. Correlate effects observed in-life with post-mortem findings.</td>
</tr>
<tr>
<td>V.</td>
<td>Record animal observations daily throughout the test. Body weight weekly for 13 weeks and monthly thereafter. Food consumption weekly for 13 weeks and at 3 month intervals thereafter. Record clinical signs and mortality for all animals. During the course of the study, clinical signs may suggest the need for other clinical determinations or post-mortem examinations. Examination of the cranial cavity at necropsy. Histopathological examination of brain (3 levels), spinal cord (2 levels), eye, pituitary, and peripheral nerve.</td>
<td>Report individual data for all descriptions of toxicological and pharmacological effects and abnormalities/animal/day of observation. Include behavioral and clinical abnormalities, and pathological findings. Correlate effects observed in-life with post-mortem findings.</td>
</tr>
</tbody>
</table>

* These protocols are frequently combined into one study.
<table>
<thead>
<tr>
<th>TEST</th>
<th>PROCEDURES RELEVANT TO THE NERVOUS SYSTEM&lt;sup&gt;2&lt;/sup&gt;</th>
<th>REPORTING REQUIREMENTS&lt;sup&gt;3&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>VI.  Reproduction with Teratology Phase</td>
<td>Record individual animal observations daily. Pertinent behavioral changes, food consumption, and all signs of toxicity, including mortality, must be recorded. Record information on mating behavior, conception, parturition, lactation, and weaning. Note gross anomalies at birth. Record behavioral abnormalities observed in dams and/or offspring. Histopathology of the brain only when considered a target organ.</td>
<td>Report incidence and severity of all abnormalities.</td>
</tr>
<tr>
<td>VII. Three Generation Reproduction</td>
<td>Record individual animal observations daily. Pertinent behavioral changes, food consumption, and all signs of toxicity, including mortality, must be recorded. Behavioral abnormalities observed in the dams and/or the offspring must be recorded. Examination of all tissues showing gross pathological changes.</td>
<td>Report incidence and severity of all abnormalities.</td>
</tr>
</tbody>
</table>
Table 3 (continued)

<table>
<thead>
<tr>
<th>TEST</th>
<th>PROCEDURES RELEVANT TO THE NERVOUS SYSTEM</th>
<th>REPORTING REQUIREMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIII. Teratogenicity</td>
<td>Record individual animal observations</td>
<td>Report incidence and severity of all abnormalities.</td>
</tr>
<tr>
<td>(Used to determine embryotoxicity and/or teratogenicity)</td>
<td>daily. Pertinent behavioral changes, food consumption, and all signs of toxicity, including mortality, must be recorded. Note gross anomalies at birth. Record behavioral abnormalities observed in dams and/or offspring. One-half of each litter to be evaluated for soft tissue abnormalities.</td>
<td></td>
</tr>
</tbody>
</table>
II. GENERAL CONSIDERATIONS

The ad hoc Panel identified several matters that related to discussions of the FDA questions. These included the use of behavior as an index of toxicity, and the FDA approach to behavioral and observational data. In addition, the Panel noted that increased attention to good laboratory practices could form a basis for improving neurobehavioral information from applied toxicology testing.

A. BEHAVIORAL CHANGES AS INDICES OF NEUROTOXICITY

Many chemicals are known to influence the neuronal control of behavioral processes such as learning, attention, sensorimotor performance, sleep, and mood (O'Donoghue, 1985a,b). Only in the last 10 years has there been an increased focus on this broad range of effects and acknowledgement that altered neuronal function may occur prior to, in conjunction with, or in the absence of traditionally sought signs of toxicity (Sobotka, 1986).

While there are elements of behavioral indices in toxicology that are common to pharmacology, both the terminology and the approaches are slightly different (Mello, 1976). The disciplines of psychology and animal behavior provide still another perspective. These differences are related to the technical study approach inherent within each discipline. Study findings in toxicology are traditionally based upon biochemical and pathological data, while pharmacological studies typically seek to quantitate and characterize effects at specific target endpoints and may rely heavily on physiological and biochemical data. The animal behaviorist and the psychologist are trained to identify changes in the frequency of an activity (e.g., grooming, circling). By observing changes in the overall activity of animals, they seek to identify trends of normal or abnormal behavior. On the other hand, the toxicologist seeks to use behavior as a quantifiable indicator of toxicity.

Numerous physiological, biochemical, morphological, and psychopharmacological tests are used in basic research to study the nervous system; however, a majority of these procedures are highly technical. These research techniques do not readily lend themselves to the applied toxicology screening protocols needed to identify those substances that produce behavioral alterations. With the awareness that information provided from conventional testing protocols may not have sufficient scope to identify neurotoxic potential, attention has focused on expanding available data on specific measures of brain function (Sobotka, 1986). Because behavior represents the final outcome of brain function, research efforts have focused on relating behavioral patterns or changes to specific neuronal events or functional sequences within the higher nervous system.
Alterations in behavior provide a relatively sensitive indicator of exposure because behavioral procedures can be used as a study endpoint to sample over a broad range of effects or to focus on specific subsystem effects. Behavioral endpoints are important because they provide information on the integration of several underlying processes and neurofunctions including motor, sensory, attention, motivation, and reactivity (see Part B, Tilson, p.V-4). Behavioral procedures are of additional value because they are noninvasive and can be used repeatedly in longitudinal studies of chronic exposure, or to study persistent effects following acute exposure.

Weiss and Laties (1975, 1979) have noted that behavioral toxicity studies are essential to a sound basis for regulation of environmental pollutants. Nevertheless, toxicologists have been reluctant to accept behavioral data to describe neurotoxicity without corroborative evidence from morphological or biochemical studies. This reluctance has historically been related to emphasis on signs of severe toxicity (e.g., axonopathies), without consideration for neurobehavioral changes (e.g., sleep disturbances, learning impairment).

Critics of behavioral endpoints have asserted that they do not distinguish between pharmacological responses and toxicological effects as evident from pathological change, and that some behavioral endpoints are observer dependent, only qualitative, or both. However, the Panel concluded the assertion that behavioral toxicology lacks sufficient development as a scientific discipline to permit standardized testing procedures is no longer tenable (Annau, 1986; Geller et al., 1979; Tilson and Mitchell, 1984; Weiss, 1983; see Part B, Tilson p.V-1-54 and Vorhees, p.VI-1-60).

The Panel recognized that acceptance of behavior as a toxicologic endpoint raises several issues. Of critical importance is the recognition that a change in behavior may occur in the absence of a known measurable neuromorphological and/or neurochemical correlate. Although the traditional morphological and biochemical indices provide valuable information to the toxicology database, the number of neural and extraneural biochemical endpoints is far too large to be of use in initial screening for behavioral alterations (Part B, Mailman, p.IV-1-36), and the detection of finite morphological changes within individual neurons or groups of neurons requires sophisticated research equipment and knowledge (Spencer and Schaumburg, 1983b). The integrative nature of the nervous system frequently presents additional complications for the toxicologist as the observed neurotoxic effects may be influenced by extraneural mechanisms (e.g., hormonal changes).

Spencer and Schaumburg (1983a) and Anger and Johnson (1985) have defined neurotoxicology as the study of alterations in the nervous system which are described in terms of biochemical, morphological, or functional changes. The EPA
guidelines define neurotoxicity as any adverse effect on the structure or function of the central and/or peripheral nervous system related to exposure to a chemical substance.

The FDA may choose to use these examples as a basis to develop a clear definition of behavior as a study endpoint in preclinical testing. Such a definition should integrate the similarities among the varying approaches (i.e., behavioral, pharmacological, and toxicological). If the fundamental objective of the toxicology screening process is to detect the consequences, then it should not matter whether the neuron or the axon dies, the glial cell fails to function, or that edema is present. What is important is that the scientist use all reasonable and cost-effective tools to provide sufficient information to permit an accurate assessment of toxic effect on the nervous system.

The Panel concluded that behavior is a valid toxicological endpoint which can provide useful information regarding the functional integrity of the nervous system. This observation is supported by a substantial literature base on the use of behavioral endpoints in pharmacology and toxicology studies of known neurotoxicants in animals and humans. These studies can be used as a basis to improve and expand conventional testing methods that are currently used to assess neurotoxicity.

B. FDA APPROACH TO BEHAVIORAL AND OBSERVATIONAL DATA

As noted in Table 3, the FDA recognizes the importance of reporting behavioral and clinical abnormalities and requires a description of all toxicological and pharmacological signs in most testing protocols. The FDA guidelines afford the petitioner some freedom in identifying and reporting study findings related to neurobehavioral toxicity. The Panel observed that in practice petitioners neither report, nor does FDA require, detailed data of this type. This may reflect a misunderstanding of a fundamental objective of applied toxicology testing, the availability of specific directives, an oversight in the reporting process, present understanding of the biological significance of behavioral changes, or insensitivity of available test systems to distinguish between normal variation and neurobehavioral changes.

In contrast to current EPA guidelines, the Panel noted that the term "behavior" has not been defined by FDA nor have any specific endpoints been listed for evaluation. The term as it is presently applied in FDA toxicology guidelines has little in common with the neurobehavioral toxicity addressed by the Panel. However, although these test guidelines do not describe the specific type(s) of behavioral information that should be collected, the lack of definition permits easy adjustment at any time to a broader definition which could include observations specifically intended as indices of neurobehavioral toxicity.
As evident in Table 3, these toxicology guidelines already provide the FDA with both the means and the flexibility to require collection and reporting of additional behavioral data. Currently, implementation of FDA guidelines leaves development of a protocol for the collection of such data to the individual investigator and reserves the right to request that such information be provided when FDA scientists conclude that it is not adequately available in the data presented by petitioners. Inasmuch as there are many neurological subsystems that may alter behavioral function as a consequence of exposure to certain classes of compounds, the Panel concluded that investigators should be given additional guidance as to what should be reported.

The Panel also concluded that, with some modifications to conventional toxicity tests, behavioral changes can be identified to the extent that a preliminary identification of possible neurotoxicity can be made. In this context, observations from conventional toxicology protocols or adjunct testing for behavioral changes can function as the first tier of a system for determining potential neurotoxicity. A positive finding from such first tier testing would provide the basis for further testing in a multitiered system.

C. EVALUATING NERVOUS SYSTEM FUNCTION IN CONJUNCTION WITH GOOD LABORATORY PRACTICE GUIDELINES

With the acceptance of Good Laboratory Practice Guidelines (GLP) (Aitio, 1981; Food and Drug Administration, 1978), interlaboratory comparisons of findings from experimental animal studies have been less difficult because of defined protocols, regular testing procedures, and more consistent reporting of study findings. Institution of the GLP regulations has raised both awareness of and concern for the organizational process and the conditions under which laboratory studies with experimental animals are planned, performed, monitored, recorded, and reported. The Panel observed that the intent of GLP guidelines could go beyond the extent of present FDA application. Increased attention to GLP could provide a basis for improving information about neurobehavioral toxicity from available applied toxicology methods. The scope of GLP covers:

- Scientific matters concerning the test system.
- Technical matters concerning the performance of a study, reporting of study data, and record-keeping.
- Organizational matters concerning the test facility, its management and personnel responsibilities.
A major component of GLP regulations address matters related to controlling the influence of environmental stress on the test system. Numerous investigators have observed that environmental stress can affect animal behavior (Selye, 1980). In preclinical laboratory investigations, factors leading to environmental stress must be controlled in order to distinguish between the effects of exposure to the test substance and early indicators of behavioral change, normal variation within the test system, and environmental influences on the test system. Findings from studies that do not maintain rigid control of environmental factors may be misleading or provide FDA with inconclusive data.

Butcher and colleagues (1979) have reported that interlaboratory differences in study findings could be attributed in part to the influence of environment on test outcome. This presents a significant problem when interpreting study findings for many indices of toxicity (e.g., biochemical, behavioral). The Panel is aware that a recent FDA Request for Proposal (RFP 223-86-2095, Dietary Amino Acids and Brain Function) for testing the effects of aspartame and other related compounds will include considerations for determining whether the characteristics of treatment-related effects and experimental conditions (e.g., circadian changes) can influence the outcome of the study. A comprehensive list of environmental and management conditions to be recorded as a part of biological data banks has been reported by Altman and Fisher (1981).

Uncontrolled environmental stress factors can confound and/or compromise study findings (see Part B, Steinberg, p.II-11). Lack of consideration for animal care or housing requirements are known to influence diurnal activity patterns, induce fluctuations in hormone levels, produce hypo/hyperthermia, alter respiration, or introduce infection. Multiple species housing (e.g., dogs and monkeys) within the same animal room can initiate predator-prey behaviors and elicit pheromone responses. Deprivation of food or water represents a strong stress factor that can produce well-documented multilevel changes in laboratory animal physiology and central nervous system (CNS) function.

In addition to these rather obvious considerations, the GLP regulations also address a number of general considerations common to all laboratory studies which are equally relevant when evaluating the integrity of the nervous system in applied toxicology screening procedures. These include consideration for study design, the collection of data, and the evaluation and reporting of study findings.

1. Study design

A major emphasis of contemporary applied toxicology is the detection of substances that produce irreversible effects (e.g., carcinogen or teratogen). While there is no question that
the detection of carcinogens or teratogens is paramount to the public interest, it is equally important to monitor and report functional changes including effects upon the nervous system. The Panel observed that existing screening protocols are not likely to detect anything other than severe neurotoxicity unless the study protocol is initiated with an intent to disclose alterations in nervous system function.

a. **Statement of purpose**

In accordance with the GLPs, petitioners are required to include test methods which ensure that the objectives of the study plan will be evaluated. This is clearly stated in section 58.120 of GLP "Each study shall have an approved protocol that clearly indicates the objectives and all methods for the conduct of the study" (Food and Drug Administration, 1978). Available FDA testing guidelines (Food and Drug Administration, 1982) indicate that the study should include an assessment of the functional integrity of the nervous system as reported by pharmacological effects, and behavioral and clinical abnormalities.

The Panel recommended that toxicity studies designed for regulatory submission should contain a clear definition of the intent and use of study data with regard to nervous system function. In addition, the study design should measure and characterize, to the fullest extent possible, relevant endpoints of neurotoxicity and behavioral dysfunction within the selected test system.

b. **Selection of test system**

The extrapolation of findings from animal models to humans continues to be a controversial subject. At present there is no single test system in animals which adequately presents a true homology for behavioral testing in humans. Furthermore, the intricacy of the nervous system, cross species differences in neuronal pathways, as well as intraspecies genetic variation may preclude the development of such a system. Researchers could be misled by false positive or false negative errors if they look too closely for an animal model to serve as an exact homology to humans. The Panel noted the advantage of observing a functional change in a commonly used test model to serve as a first tier identifier for additional testing. If a test indicates that the study animal is more sensitive to an environmental perturbation, then measurement of this perturbation itself is the change that should be observed and may provide data useful in determining dose threshold. If a neurobehavioral change is identified by a functional observation, then there are numerous behavioral tests that can be used to characterize and measure the nature of the deficit (see Part B, Tilson, p.V-11-30).
Neurobehavioral changes have sometimes been overlooked. As examples, the Panel cited studies of the loss of peripheral vision with methyl mercury (Evans et al., 1975; Mukuno et al., 1981; Rice and Gilbert, 1982), functional deficits following in utero alcohol exposure (Streissguth et al., 1984), and of hyperactivity and sleep disturbances with synthetic opiates (Hutchings, 1985a; Hutchings and Fifer, 1986). In these examples, the changes were not detected in the conventional applied toxicology studies and supplemental studies in laboratory animals had to be designed retrospectively on the basis of clinical observations in humans. To detect the complex, multifactorial effects which can be associated with subtle changes in nervous system function, animal findings should be compared to clinical observations whenever possible. This may lead to additional testing in laboratory animals which may not have been initiated had the same observations been made independently and in isolation (Hutchings, 1985b).

The Panel concluded that there is a need to maintain flexibility in developing study protocols for neurotoxicity and that there should be a cooperative dialogue among clinicians, regulatory scientists, and investigators in order to develop supplemental study protocols to cross-validate findings between humans and animals.

c. Dose selection

The selection of dose is critical to the assessment of toxicity in test animals. The investigator should select exposure levels that ensure observation of a broad range of toxicological effects (i.e., no observed effect to severe toxicity). Studies have shown that, for some chemicals, measurable changes in nervous system function can be identified at doses lower than those commonly associated with severe toxicity.

A Committee of the National Academy of Sciences has calculated levels of numerous pesticides in drinking water and set Acceptable Daily Intake levels (ADIs) for many of these substances (National Research Council, 1977b). These ADIs were based on no-effect levels of exposure in a functional observational battery. Using simple, quantitative behavioral measures (e.g., motor activity in a figure-eight maze) to evaluate neurotoxicity, Reiter observed that the ED50 values (i.e., the effective dose at which a 50% decrement is observed) are often below those associated with severe signs of toxicity (Reiter et al., 1981).

In addition, developmental studies have shown that there is a range of impaired function that can be observed between the no observable effects level (NOEL) and the dose threshold for gross fetal malformation (Butcher et al., 1975a,b; Hutchings, 1985a). These approaches have led to applications
in environmental toxicology. For example, behavioral and/or neurological effects are frequently used in determining Threshold Limit Values (TLVs) and exposure limits set by the American Conference of Industrial Hygienists (ACGIH) and the National Institute of Occupational Safety and Health (NIOSH) (Part B, Tilson, p.V-3).

Initial high-dose testing is useful for providing information to identify the specific CNS components to be evaluated. Such information is helpful in selecting the most sensitive test procedure to employ in advanced study protocols. Behavioral tests should be used at dose levels below the highest dose which produces severe toxicity, and would be most useful at a dose range lower than that at which gross toxicity has been detected. To avoid false positive findings, these behavioral tests should be used only after the dose response has been determined from acute dose range-finding studies. Quality assurance can be maintained for these procedures by using a positive control substance. The design considerations of the Collaborative Behavioral Teratology Study provide insight for the importance of dose selection and the detection sensitivity of several behavioral test procedures (Adams et al., 1985).

The Panel observed that behavioral testing procedures have long been used in behavioral pharmacology and that these procedures are useful in describing dose-effect relationships (Seiden and Balster, 1984; Weiss and Laties, 1976).

2. Collection of quality data

The Panel noted that three elements are essential to the generation of quality data derived from conventional applied toxicology. These include standardized operating procedures, trained personnel, and a formalized mechanism for recording data.

a. Standard operating procedures

The GLP provisions require that every testing facility have written standard operating procedures (SOP) to insure the quality and integrity of the data generated during a study. The scope of the SOP requirement includes details which cover every procedure from receipt of animals at the beginning of a study to the storage of tissue specimens at the conclusion of the study. Of particular interest to neurobehavioral investigators are the SOP requirements detailing test system observations and laboratory procedures (i.e., GLP Subpart E). The Panel noted that in regard to neurobehavioral toxicity, the SOP should identify the items to be observed, describe standard procedures for collection of data, and list the information to be recorded. The FDA may receive information that is incomplete or misleading.
if the test facility does not provide sufficient guidance to laboratory personnel for the collection of neurobehavioral information in the SOP.

b. **Laboratory personnel**

Laboratory personnel provide the primary interface between the test system and the collection of study findings. The collection of quality data requires that both technical and nontechnical laboratory personnel be aware of SOP and understand their responsibility within the framework of the study protocol.

The GLP requires that an adequate number of appropriately trained personnel be engaged in the conduct of or be responsible for the supervision of a nonclinical laboratory study. Noting that many neurobehavioral procedures require exacting protocols and rely on observational findings collected at the technician/test system interface, the Panel expressed concern for the training of study personnel.

There is no replacement for observations recorded by the experienced and responsible laboratory technician. Irwin (1968) estimated that his comprehensive observational assessment could be learned with 2-6 weeks of training. Gad (1982) observed that technicians can be trained to screen animals after approximately 6 hours training with reinforcement conducted over a 1-week period. As with any training program the breadth of technician experience at the beginning of training will determine the time needed to learn and develop proficiency for a newly acquired skill.

The Panel concluded that there is a need to develop training and certification programs for laboratory personnel responsible for the conduct of behavioral testing and the recording of animal observations. Testing facilities should be encouraged to provide training programs to enhance personnel awareness of integrated CNS function, the rationale for good laboratory practice, and the recording of quality data required for assessing nervous system function. The laboratory assistant and technician certification programs supported by the American Association for Laboratory Animal Sciences (AALAS) should be encouraged to expand available training programs to include neurobehavior.

c. **Recording raw data**

Large quantities of data are collected during the course of a toxicity study (see Part B, Steinberg, p.II-3). The original observations and descriptions of matters pertaining to daily activities as reported by study personnel are of little
use if they are not systematically recorded for future evaluation. Details of the requirements for recording data generated during the conduct of a nonclinical laboratory study are described in Subpart G of GLP.

The individual data elements to be collected during the study are determined by the study protocol and are detailed in the SOP. The use of a systematic method to record raw data from laboratory procedures enhances consistency of observation. In addition, systematic collection methods help to organize data elements to facilitate their evaluation.

The Panel observed that most test facilities have already made a significant investment in their data collection systems. Nevertheless, the Panel noted the importance of including specific checksheets to guide laboratory personnel in recording required observations. The use of computer-driven data collection systems was strongly recommended as they increase diligence via pre-programmed prompting. Computerized data facilitates the sequestering of large volumes of data for the evaluation of trends between study groups over numerous observation periods. Lewis et al. (1985) have reported on the application of a computer-support method for analyzing behavioral observations.

3. Data evaluation

Matters pertaining to the content of the final report are addressed in the GLP (Subpart J). Both the study director and the statistical approach used to evaluate the data are an integral part of report preparation. The study director has overall responsibility for the interpretation, analysis, documentation, and reporting of study results (GLP, Subpart B). This is an all-encompassing task for one individual who is usually specialized in one discipline (e.g., toxicology, pharmacology, pathology).

The multidisciplinary nature of data derived in support of neurobehavioral toxicity frequently require special insight to evaluate nervous system function. When considering the extensive number of data elements which can influence CNS function it is necessary for the study director to rely on the expertise of scientists specially trained in the neurobehavioral and neurotoxicological sciences. Such scientists provide experience in integrating the numerous data elements from the variety of test protocols needed to assess neurobehavioral toxicity and provide expertise for special applications of statistical methods unique to behavioral data. In addition, the development of study protocols and the data from neurobehavioral testing procedures should be reviewed by scientists trained in neurobehavioral toxicology.
The Panel concluded that significant improvements in understanding the relationship between environmental influences and nervous system function could be made in existing study protocols and supplemental neurobehavioral toxicity testing if the FDA broadened its definition of "behavior" and investigators applied the goals and objectives as outlined in the GLPs. Under these conditions, both intra- and interlaboratory comparisons of neurotoxicity evaluations might exhibit less variation.
III. AVAILABILITY OF BEHAVIORAL AND NEUROTOXICITY INFORMATION FROM CONVENTIONAL TESTING

Accumulating evidence suggests that the nervous system plays a central role both as a prime target for toxic effects of many compounds and in mediating the systemic effects of other compounds. The extent to which the various conventional toxicology tests give information about the nature and scope of potential neurotoxicity or behavioral dysfunction relates to the availability of information about neurobehavioral toxicity and the extent to which this information is reported.

A. TYPES OF INFORMATION AVAILABLE

Conventional toxicology testing protocols provide a number of opportunities to collect preliminary information on frank neurobehavioral changes, selective measures of neurotoxicity, and histological evidence of neuropathology (see Part B, Steinberg, p.II-3). Table 4 lists the types of information currently available from the conventional applied toxicology protocols used by FDA.

Much of the available neurobehavioral information from these test protocols is in the form of observational or indirect indices which can be used to infer impaired nervous system function. Such indices include food and water consumption, body weight changes, mating and sexual behaviors, clinical observations (e.g., appearance, motor activity), and pharmacological signs (e.g., cholinergic storm). Many of these are not specific nervous system signs and symptoms and are often multisystem determined.

1. Observational information

With the exception of the acute LD$_{50}$ study, all conventional toxicity protocols provide for some evaluation of food consumption and body weight change in study animals. This type of information is useful for maintaining consistent animal dosage during the course of a study and in providing information on compound-related changes from predicted norms. For example, it may provide an early indication of changes in consumption behavior produced by taste aversion or preference for the test substance. However, data of this type provide little information on CNS integrity, and is nonspecific to alterations in nervous system functioning.

Most conventional toxicology protocols include clinical observations of animal conduct. These observations are usually recorded from cage-side monitoring of study animals during scheduled dosing or collection of body weight measurements.
Although Irwin (1968) was probably the first investigator to outline a structured method for the collection of observational data, numerous other investigators have updated his original procedures (e.g., Alder and Zbinden, 1983; Balazs, 1970; Gad, 1982). The information that may be recorded includes subjective observations of changes in motor function (e.g., convulsion, tremor, gait disturbances, paralysis), clinical observations of appearance (e.g., piloerection, hunching), abnormal behaviors (e.g., lethargy, hyperactivity, self mutilation), and observations of pharmacological signs (e.g., ataxia, prostration, changes in salivation or lacrimation). In addition, simple procedures are available to provide observation of several reflexes (e.g., Preyer's reflex, corneal reflex, righting reflex, and negative geotaxis).

A rapid death observed in the acute LD$_{50}$ study can infer an effect on the CNS (e.g., depression of respiration). In addition, pharmacological events such as a cholinergic storm, as characterized by increased urination and lacrimation, can be detected. Multiple dose levels in acute studies permit observations on the graded nature of these findings. Similar gross observational information can also be available from the subacute and chronic testing protocols which provide the added benefit of a longitudinal assessment of these indices.

2. Reproduction and developmental information

Reproduction and developmental toxicity studies also provide information on the nervous system if the data are viewed with a perspective for neurotoxicity. A variety of testing protocols are presently available from several sources (Food and Drug Administration, 1982; International Programme on Chemical Safety, 1984; U.S. Department of Energy, 1982; U.S. Environmental Protection Agency, 1986). Developmental studies provide a priori information related to sexual behaviors (e.g., mating), mother/offspring interactions (e.g., rearing, lactation), and growth and development (e.g., neonatal survival, body weight). Although reproductive performance requires the successful integration of many neurological functions, such information is not equivalent to a comprehensive clinical neurological examination. The detection of terata can provide direct visible information of gross perturbations to nervous system development. Information on neonatal performance from available FDA protocols is limited to frank observations from multiple generation reproduction studies. In contrast, EPA guidelines for developmental neurotoxicity require: a) behavioral observation of dams and offspring; b) evaluation of physical landmarks; (e.g., eye opening, incisor eruption); c) tests for motor activity, auditory startle habituation, and the Biel water maze paradigm; and d) detailed neuropathology (U.S. Environmental Protection Agency, 1986).
Table 4. Some Available Information from Conventional Applied Toxicology Testing Protocols.

<table>
<thead>
<tr>
<th>TESTING PROTOCOL¹</th>
<th>TYPES OF AVAILABLE INFORMATION²</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACUTE</td>
<td>GROSS OBSERVATION</td>
</tr>
<tr>
<td></td>
<td>• Morbidity/Moribundity</td>
</tr>
<tr>
<td></td>
<td>• Mortality</td>
</tr>
<tr>
<td>SHORT-TERM</td>
<td>OBSERVATION OF CLINICAL SIGNS</td>
</tr>
<tr>
<td>SUBCHRONIC</td>
<td>• Severe toxicity</td>
</tr>
<tr>
<td>CHRONIC/</td>
<td>• Abnormal behavior</td>
</tr>
<tr>
<td>LONG-TERM</td>
<td>• Motor changes</td>
</tr>
<tr>
<td>CARCINOGENICITY/</td>
<td>OTHER MEASUREMENTS</td>
</tr>
<tr>
<td>MUTAGENICITY</td>
<td>• Food/water consumption</td>
</tr>
<tr>
<td>TERATOLOGY</td>
<td>• Body weight</td>
</tr>
<tr>
<td>REPRODUCTION</td>
<td>• Organ weight</td>
</tr>
<tr>
<td>MULTI-GENERATION</td>
<td>• Pathology (gross and microscopic)</td>
</tr>
<tr>
<td>REPRODUCTION</td>
<td>• Hematology</td>
</tr>
<tr>
<td></td>
<td>• Clinical chemistry</td>
</tr>
<tr>
<td></td>
<td>REPRODUCTIVE PERFORMANCE</td>
</tr>
<tr>
<td></td>
<td>• Mating</td>
</tr>
<tr>
<td></td>
<td>• Parturition</td>
</tr>
<tr>
<td></td>
<td>• Offspring survival</td>
</tr>
<tr>
<td></td>
<td>• Teratogenicity</td>
</tr>
<tr>
<td></td>
<td>• Growth of offspring</td>
</tr>
</tbody>
</table>

¹ Several protocols may be combined within a testing program to form a single study. Examples include combined chronic toxicity with carcinogenicity and reproduction with a teratology phase.

² Availability of information is dependent on the selection of the test protocol.
3. **Clinical and biochemical laboratory studies**

Laboratory procedures frequently included in long-term studies provide information on hematology and clinical chemistry. While much of this information is more closely associated with functional alterations in other organ systems, conditions such as anemia and fluctuations in blood pH and blood gases are indicative of anoxic conditions which might be expected to interfere with nervous system function (Weiss and Laties, 1975).

Although there are numerous biochemical assays for specific neurochemical endpoints, these are rarely incorporated in conventional applied toxicology protocols. The typical exceptions are assays for cholinesterase activity following pesticide exposure. As noted by Mailman (see Part B, Mailman, p.IV-1-36), such procedures would not be expected to be of value for screening compounds other than those with cholinergic effects. Biochemical measures can provide concurrent validity with behavioral effects if they are systematically related to exposure. Such biochemical assays do provide useful information when applied to advanced studies characterizing the mechanism of action of the test substance.

4. **Pathological information**

The pathological information presently used in conventional toxicology screening studies includes a description of gross findings, organ weight data, and the histopathological evaluation of hematoxylin and eosin sections for three levels of brain, two levels of spinal cord, and in most cases, at least one segment of peripheral nerve for adult study animals. Fixation procedures are conventionally limited to buffered formalin preparations with little or no consideration given to organ perfusion. In applied toxicology, special staining procedures (e.g., Nissel stain) are the exception rather than the rule.

B. **REPORTING OF AVAILABLE INFORMATION**

Available FDA guidelines for the collection and reporting of the various categories of information described above are specific for traditional toxicology data sets but are frequently nonspecific or incomplete for data sets attributed to the newer areas of toxicology (e.g., behavior, immunology). As a rule, all collected information is reported and the bulk of the preclinical report consists of individual data sets for observational findings of survival and severe toxicity, food consumption, body weight, hematology, clinical chemistry, organ weight, pathology, and reproduction and developmental findings. Conventional toxicity studies as they are presently reported to FDA, even when carried out at sufficiently high exposure levels, provide little useful
information about neurobehavioral toxicity. This is because present methods for cage-side observations are neither sufficiently systematic nor sufficiently quantitative to be of value.

The incomplete reporting of observational information available from conventional studies becomes a confounding factor when discussing the subject of adequacy. The Panel sought to identify a series of toxicity studies to use as a reference in their discussion of available information. The discussion focused on the National Toxicology Program (NTP) studies which employ a standardized computer data collection system for animal observations collected at weighing intervals. The Panel noted that while this information is being collected in the testing laboratory under the terms of existing NTP protocols in a controlled and systematic fashion, it is not readily available in the final report. Under the heading of "Body Weight and Clinical Signs" these reports include details on the body weight change for each dose and sex. Occasionally in mouse studies there may be a notation of fighting behavior among male animals. In general, many of these reports contain a standard statement "no other compound related clinical signs were observed", but do not provide information to support this statement.

The Panel speculated that, should these observational data and other relevant types of data be made available for review by a trained neuroscientist, they might provide additional information as a retrospective analysis for known neurotoxicants that have been studied in the NTP program. Such an analysis could also provide much needed insight into the use of conventional protocols to detect early signs of neurobehavioral toxicity from observational data.

When collected in a systematic manner, observational data can provide useful information on the nature and scope of potential neurobehavioral toxicity as a preliminary level of testing. The use of standardized, computer-generated checklists to collect, store, and tabulate this type of information has enhanced the utility of these toxicity endpoints. However, the reporting requirements for these data which frequently include descriptive statements, appear to have fallen behind current technological advances. There is a need to catalogue this observational information into a prioritized functional observational battery (e.g., Gad, 1982; Irwin, 1968) with specific reporting requirements for their presentation to FDA.

There is no specific requirement to summarize relevant neurofunctional and related neurotoxicity data in submissions to FDA. The FDA guidelines do not require a specific statement about neurotoxic potential. In contrast, other countries require petitioners to include an integrative description of all test substance effects on the test system (see Part B, Steinberg, p.II-22). The Panel observed that a separate, interpretive
statement of all available and inferred information regarding the functional integrity of the nervous system is needed. This statement should be prepared by a qualified neurotoxicologist.

A note about international reporting of study findings is in order. As described in Table 2, the approach to neurotoxicity testing frequently varies for regulatory agencies both within the same country and between countries. For example, behavioral testing of progeny is an accepted practice in Japan and the United Kingdom, however, in the United States it is not included in study protocols for FDA submissions. Petitioners frequently find themselves in the awkward position of multiple agency reporting. By necessity many petitioners have already invested considerable resources toward the development of a program for behavioral teratology based on basic research into testing methods and the effects of positive control compounds. Various laboratories in the United States have behavioral test batteries in place and already have the capability of conducting and reporting behavioral information (Nolen, 1985). The Panel noted that variations in interagency guidelines may lead to inconsistencies in the comprehensiveness of information available from studies of the same compound tested by the same basic study design for submission to different agencies.

Current neuropathology screening is limited because the number of sections sampled, preparation techniques, and levels of analysis are too incomplete to be useful. The limited number of histopathological sections of the CNS provide only an extraordinarily rough index of neuropathology. The brain is a unique organ; one or two poorly or inconsistently selected sections of CNS tissue may not always present a representative picture of neurologic injury related to a highly specific neurotoxicant. In addition, by placing the burden for the selection of fixation and staining procedures on petitioners, the FDA may receive incomplete or possibly misleading information. Organ weight (obtained from uniformly collected specimens) or other physical brain measurements (i.e., length, width) are well established pathological indices that depict gross changes to this organ. Current FDA guidelines do not require the collection of these data. The Panel identified current EPA neuropathology guidelines (U.S. Environmental Protection Agency, 1985b, 1986) as a useful model for further consideration, should the FDA choose to modify current pathology requirements.

The Panel concluded that even if improvements for the collection of observational and pathological data are made to FDA toxicology protocols, the neurotoxicity data provided would in themselves be limited without additional emphasis on provisions for obtaining information about the functional integrity of the nervous system.
IV. ADEQUACY OF NEUROTOXICITY INFORMATION FROM CONVENTIONAL TESTING

The extent to which conventional testing protocols fail to precisely assess nervous system toxicity is difficult to determine because these protocols provide virtually no quantifiable measurements for this type of toxicity. The FDA approach is currently so broad for this target organ system that it precludes the detection of most neurotoxic effects with the exception of those cases where CNS effects are clearly evident.

A. ADEQUACY OF INFORMATION AS DETERMINED BY STUDIES OF WELL-ESTABLISHED NEUROTOXICANTS

To address the question of adequacy, the Panel discussed a series of examples that included environmental agents, occupational chemicals, drugs, and dietary constituents which are either well-established neurotoxicants or represent substances thought to be neurotoxic (e.g., methyl mercury, ergot compounds, FD&C Red No. 3, antibiotics, synthetic opioids, vitamin B). Some degree of neurotoxicity and/or behavioral dysfunction could be associated with almost all of the compounds discussed. Several Panel members expressed concern that some dietary constituents might act on a CNS neurotransmitter system, thereby resulting in behavioral effects which could occur in the absence of toxicity as detected by traditional toxicology endpoints. In discussions of these chemicals, the Panel noted that indices of neurobehavioral toxicity provide relevant information which is presently omitted from conventional testing protocols.

Based on the Panel's review of well-established neurotoxicants, both the selection of dose and range of effects sampled by the study endpoint are intimately related to the sensitivity of the test system and the question of adequacy of information in the database. The Panel noted that neurobehavioral changes associated with exposure to many of the substances discussed could have been undetected by conventional toxicity studies if tested at an inappropriate dose. At this time, the conventional toxicity studies sample a range of toxicologic endpoints (i.e., biochemical and pathological) that provide limited information on functional change. If the investigator selects a dose below the level of detection for the selected study endpoints, the test substance could be misconstrued as having no effect when in fact the selection of an endpoint that affords sampling over a broad range of function (i.e., behavior) could prove the opposite.

Tanimura (1985), reporting on the findings of the Japan Pharmaceutical Manufacturers Association survey for the use of behavioral teratology in drug testing, indicated that the dose causing behavioral effects is frequently lower than that inducing
teratogenicity and that behavioral disorders in the offspring of rats can be more sensitive indicators of developmental toxicity in some cases. In contrast, Norton (1986) has reported that neither morphological nor behavioral tests were more sensitive than neonatal body weight changes for detecting damage from gestational irradiation. She concluded that the failure to detect dose-related effects of ionizing radiation on behavior was not obvious because a) these effects may represent a qualitative change rather than quantitative change in behavior with increasing degrees of CNS damage, and b) the developing CNS has a potential for functional compensation in the presence of small amounts of damage. As noted earlier, Hutchings (1985a) reported that the subtle neurobehavioral changes (e.g., sleep disturbance) associated with methadone were not identified by conventional applied toxicology screening methods. In this example, the neurobehavioral toxicity was initially identified from clinical findings in humans and later demonstrated in the rodent model using a neurobehavioral testing procedure. These examples suggest that all behavioral procedures are not equivalent in predicting CNS damage in dose-related studies. A reasonable approach is to use a series of tests which evaluate more than one aspect of CNS damage to toxicant exposure.

B. ADEQUACY OF INFORMATION AS DETERMINED BY THE STUDY MODEL

Using a different approach to the question of adequacy, the Panel explored the use of the rodent model for reproduction and development. Several toxicant effects could be missed by depending on this model. For example, erectile dysfunctions, a major cause of sexual dysfunction in humans, would be overlooked as the rodent has an os penis. The study of infertility provides another example. Sperm production in the rat greatly exceeds needs, whereas in humans, sperm production or survival may be factors in infertility. Similarly, there is little comparison between menstruous and estrus species, and the luteal phase in humans is automatic whereas in rodents it is not. The effects of hormones, however, can be studied peripherally and are known to show potent behavioral effects.

The Panel concluded that conventional toxicology screening procedures currently used by FDA are by themselves not adequate as an index of neurobehavioral toxicity because they omit quantifiable measures of integrative nervous system performance and endpoints of neurobehavioral toxicity. Some additional adverse effects that might slip through conventional testing by the selection of an inappropriate dose or animal model, could be detected by including neurobehavioral procedures in existing applied toxicology testing protocols.
C. ADEQUACY OF INFORMATION FROM BEHAVIORAL INDICES

Assessment of the nervous system should include an evaluation of behavioral processes such as sensory, motor, arousal, cognitive, and integrative functions. Conventional study protocols presently used for applied toxicology by FDA could be modified to include specific tests for these functions. Tilson has reviewed the use of many currently available neurobehavioral test methods (see Part B, Tilson, p.V-1-54). Neurobehavioral tests are considered reliable, sensitive, and valid indices of neurotoxicity (International Programme on Chemical Safety, 1984; U.S. Environmental Protection Agency, 1985a, 1986; and see Part B, Vorhees, p.VI-1-60). Behavioral methods can provide an assessment of overall CNS functional integrity, whereas other methods in the neurosciences (e.g., neurochemical, electrophysiological) are generally directed towards specific subsystems.

Motor functions for which there are neurobehavioral tests include spontaneous motor activity (e.g., figure-eight maze), motor coordination (e.g., rotarod test), muscle strength (e.g., Meyer's grip strength, swimming maze), deviations in movement or posture (e.g., hindlimb splay), and tremor. Sensory function can be assessed from neurobehavioral tests such as sensory screening, reflex modification, and instrumental conditioning. Nonassociative and associative tests are used to evaluate learning and memory, while schedules of reinforcement and maze tests provide information on instrumental performance, and consumption behaviors are an index of naturally occurring responses. For a detailed discussion of these and other neurobehavioral procedures see Tilson (see Part B, Tilson, p.V-1-54).

Of the list of neurobehavioral endpoints described by Tilson (see Part B, Tilson p.V-1-54), only consumption behaviors and gross indicators of motor function are presently included by FDA in conventional testing protocols for adult animals. In contrast, EPA guidelines for neurotoxicity testing provide details for tests including: a) functional observational battery; b) motor activity; c) neuropathology; d) neurotoxic esterase testing; e) schedule-controlled operant behavior; f) acute and subchronic delayed neurotoxicity of organophorus substances; and g) peripheral nerve function. In other countries, neurobehavioral testing has been more widely accepted for inclusion in developmental studies (Barlow, 1985; Tanimura, 1985). Nolen (1985) has described an industrial approach to meet the neurobehavioral teratology testing requirements of other countries. The NCTR Collaborative Study describes the application of behavioral procedures in teratology studies (Buelke-Sam et al., 1985), and the EPA has established developmental neurotoxicity screening guidelines (U.S. Environmental Protection Agency, 1986). The Panel identified these sources of information as useful models for further consideration should FDA choose to include neurobehavioral study endpoints in applied toxicology testing protocols.
V. STRATEGY FOR APPLIED NEUROBEHAVIORAL TOXICOLOGY TESTING

The notion that neurobehavioral testing should be included in applied toxicology protocols has been a subject of repeated discussions for many years (see Part B, Table I.1, p.I-2 and 3). Tilson (see Part B, Tilson, p.V-8) has reviewed the historical background on the development of various checklists and test batteries for neurotoxicology. The application of behavioral methods in reproduction and developmental testing was reviewed by Vorhees (see Part B, Vorhees, p.VI-1-60). Some recommendations by several study groups have been summarized in Table 1. In keeping with the recommendations of other study groups and the available literature on testing strategy, the Panel recognized a multiple tier approach as the most practical means of improving information on neurobehavioral toxicity within conventional testing protocols.

A. LEVELS OF INQUIRY IN APPLIED NEUROBEHAVIORAL TOXICOLOGY

In applied toxicology there are four major objectives in the evaluation of the effects of test substances [i.e., detection, characterization, identification of the target organ(s), and development of dose/time relationships] (Doull et al., 1980; see Part B, Sette, p.III-3). To meet these objectives, conventional toxicology protocols typically screen for a broad range of potential effects initially and then determine a dose at which no effect is observed.

Neurobehavioral procedures can be adapted into the tiered approach within the applied toxicology framework and can provide a range of information at different levels of inquiry. The first level includes those procedures intended to detect the occurrence of a change by screening a broad range of effects, while the second level includes specialized procedures used to assess the degree of toxicity or the lowest exposure level at which the effect can be observed.

Screening procedures permit the evaluation of a large number of animals, are frequently simple to perform, and do not require extensive staff training. The data provided by these procedures are often subjective, quantal, and less sensitive than specialized procedures (see Part B, Tilson, p.V-4).

Specialized neurobehavioral tests are more specific for the category of CNS function to be evaluated and are generally used for those studies that characterize the extent of CNS involvement, identify the mechanism of action, or estimate the NOEL. Although these procedures need not always require highly specific automated equipment, they frequently require a commitment of time and manpower.
1. **Selection of study protocol**

The Panel acknowledged that it is cost prohibitive to conduct extensive neurobehavioral testing for every test substance evaluated by applied toxicology testing. It follows that the choice of neurobehavioral tests will depend on the desired level of analysis. Those neurobehavioral screening procedures that sample a wide range of functions can be used to complement existing acute, subacute, and subchronic protocols at a first level of inquiry. Specialized neurobehavioral procedures are useful in advanced testing (e.g., long-term, chronic studies) as a secondary level to further characterize the nature and extent of neurotoxicity when positive findings are observed at the first level of testing. The unique vulnerability of the developing organism necessitates that once a decision has been made to conduct reproduction and developmental testing*, tier 2 neurobehavioral testing procedures are required.

2. **Selection of animals for study**

Several Panel members expressed concern for the effects of extensive manipulation of study animals during behavioral procedures. There is evidence to suggest that manipulation can produce changes in the toxicological profile of the test animal (e.g., immunology). Such alterations might modify the response to a carcinogen and reduce the power of the study (see Part B, Plotnikoff, p.VII-1-23). The Panel concluded that to maintain consistency, observations should be recorded uniformly for all study animals (e.g., during body weight measurements) and that any other procedure should be conducted in a designated satellite group. This is a well-accepted practice in conventional testing protocols for the collection of interim data on biochemical, hematological, and pathological parameters. In a rodent study a satellite group of animals could be used repeatedly throughout the study since these are noninvasive procedures. The size of the satellite group is determined on the basis of the least sensitive dependent variable. In addition, the possibility of test-test interaction must be considered when multiple procedures are required within satellite group testing.

---

* When applying neurobehavioral procedures to reproduction and developmental protocols, the Panel refers to the Segment I and perhaps a modified (i.e., by permitting some dams to deliver) Segment II, FDA study design (Food and Drug Administration, 1982). This decision is based on the logic that treatment for a full gestation will maximize CNS effects which could be observed by behavioral testing.
The surplus study animals available at the conclusion of the $F_1$ generation mating in multiple generation reproduction studies could be effectively channeled to supplemental neuro-behavioral procedures at a minimal cost. Regulatory agencies in both the United Kingdom and Japan have already acknowledged the value of conducting neurobehavioral toxicity studies using these animals (Barlow, 1985; Tanimura, 1985).

The Panel noted that, in addition to neurobehavioral tests, the choice of test could also be directed by chemical classes of concern on the basis of structure-activity relationships (SAR). This latter approach may be useful for those effects which might not be detected from screening studies.

The Panel concluded that neurobehavioral procedures are compatible with the objectives of applied toxicology and that these procedures can be incorporated in a systematic approach to enhance the scope of conventional testing.

B. TIER APPROACH

There was a diversity of opinion regarding the number of tiers to be included in any a schema. Some Panel members were of the opinion that four tiers were needed to fully characterize neurobehavioral toxicity. In this approach the schema would provide a detailed outline that differentiated between screening (tier 1), characterization (tier 2), identifying the nervous system as the primary target organ (tier 3), and the determination of dose/time effect relationships (tier 4). Other Panel members expressed concern that such a schema was more complex than is needed at this time. While a four-tier system would provide detailed information on the nature and scope of neurobehavioral toxicity for those substances tested, it would extend testing beyond the goals of applied toxicology in support of basic research. The Panel concluded that this would be cost prohibitive and an unfair burden on the petitioner.

The Panel concluded that a tier approach corresponding to the level of inquiry within the applied conventional toxicology format is feasible. Such a system would consist of two levels of testing (i.e., primary and secondary tiers). The objective of this approach is to detect and identify the nature of CNS involvement before placing emphasis on supplemental testing to characterize the extent of neurotoxicity related to the test substance.

C. PRIMARY TIER TESTING

Primary tier testing denotes that the test substance has an effect on the CNS. It serves as the initial indicator of the range of CNS change that can be observed, and whether these
effects occur at doses at which other signs of toxicity are not observed. These are important distinctions because at a high enough dose (i.e., fatal dose) many chemicals cause nonspecific CNS depression. The primary tier would help place this phenomenon in perspective by identifying a set of criteria for specialized neurotoxicological assessment and by indicating those categories of CNS involvement to be tested in the second tier.

Several Panel members raised the issue of negative findings related to questions of a) a reasonable level of testing or b) sensitivity of the tests used at this level of inquiry. The Panel recommended that the primary tier should include a test to survey a broad range of CNS effects (i.e., a functional observational battery) and one or more neurobehavioral tests to provide quantifiable data on integrated function (i.e., motor activity, learned behavior). Information on molecular structure-activity relationships when available can also be of value in the primary tier.

1. Screening for the neural component of concern

Alterations of a broad range of CNS functions including peripheral, sensory, muscular, and integrative neural components can be documented by clinical observation. Laboratory observations of frank functional deficits can be described for laboratory animals by a functional observational battery (FOB). The Panel identified current EPA guidelines for FOB testing as a useful model for further consideration (U.S. Environmental Protection Agency, 1985c). The FOB is a useful screening procedure because it provides a systematic method to enhance clinical observations of neurotoxic effects and can be used to complement procedures currently employed in applied toxicology protocols. A standardized FOB details the environmental testing conditions (i.e., test arena), the observations to be evaluated and the procedures for their evaluation; suggests a rating scale; identifies the level of technician training; and provides for a reporting format within a defined standard operating procedure.

The Irwin screening procedure (Irwin, 1968) is recognized as the forerunner of contemporary FOBs. Originally designed as a comprehensive observational procedure to determine the profile of various classes of pharmacological agents, it has been widely incorporated into acute LD₅₀ testing procedures in applied toxicology. Numerous investigators (e.g., Gad, 1982; Tilson et al., 1979, 1980) have modified the Irwin design to meet individual study needs (i.e., neurobehavioral toxicology research, industrial application). For an example of the listing of available observational endpoints and details pertaining to these methods refer to Gad (1982), Irwin (1968), Tilson (see Part B, p.V-8), Tilson and Mitchell (1984), and Tilson et al. (1980).
The Panel noted that available FOB designs have varying levels of complexity. As with any applied testing procedure the operational cost of the FOB increases as a multiplicative function of the processing time for the number of functions to be performed per animal by the number of animals to be processed and the number of observation periods during the study. The original Irwin screen estimated that 50 observational endpoints/animal could be assessed and scored within a 3-minute testing interval. In contrast, Gad (1982) suggests that 27 observational endpoints of sensory and motor function/animal can be evaluated in 3-8 minutes. When incorporated into the clinical observation procedures of applied toxicology protocols (i.e., acute, subacute, subchronic, and reproduction protocols), the impact of the processing time factor can be minimized. Although a certain amount of redundancy within the testing procedure can provide reassurance for observational findings, a refined selection of observational endpoints could maximize efficiency and avoid redundancy in testing for the same neuronal component. Such a refined selection of study endpoints could be made without compromising the scope of the test procedure.

2. Quantitative basis for minimizing observer bias

Although the FOB provides information on the presence or absence of a variety of reflexes, the data are semiquantitative (i.e., quantal, scalar, or interval) and subject to observer bias. To balance the primary tier in applied neurobehavioral screening, the Panel acknowledged that the addition of an observer, judgment-free index of neurobehavior is essential. Quantitative data from such a neurobehavioral procedure can be used to indicate the nature of a graded response across dose groups. Of the variety of neurobehavioral procedures that provide quantitative data, the Panel discussion focused on measurements of motor activity.

Spontaneous motor activity (SMA) in rodents is a high-frequency complex behavior. Individual components of this activity can be measured using automated devices that provide quantitative data free of observer judgment (Reiter and MacPhail, 1979). Changes in motor activity can reflect alterations in one or more sensorimotor functions, arousal, or motivational states. Reiter (1983) has shown that even though SMA is a nonspecific neurobehavioral endpoint, changes in SMA can be a sensitive measure of neurotoxicant exposure. As an example, he cited that the automated SMA procedure can detect signs of cholinesterase inhibited pesticide toxicity at ED50 levels. These levels are below doses that produce severe toxicity.

Some SMA procedures have been modified by shortening the processing time interval to enhance their acceptability in applied toxicology protocols. Because this procedure is time
consuming and requires housing the animals in a modified environment, the Panel concluded that use of satellite groups of animals should be considered. The procedure is best applied to acute, subacute, and subchronic study protocols where a variation in response can be evaluated across multiple dose levels.

The Panel concluded that a positive finding with the FOB and/or SMA procedures would provide sufficient justification to require additional testing to further characterize the nature of the neurotoxic effect (i.e., secondary tier testing).

3. Structure-activity relationships

The use of chemical classification and structure-activity relationships (SAR) is an accepted practice in pharmacology to assist in describing the profile of therapeutic agents (Goodman et al., 1985). Sette (see Part B, p.III-10) and Mailman (see Part B, p.IV-21) have reviewed additional background information for considering relationships based on physical or chemical properties and chemical classification in the hazard assessment process. An advantage to these procedures is that study findings of a representative chemical can frequently provide some measure of prediction for the effect of related chemicals. The SAR information can serve as an adjunct within the primary tier and may provide background experience relevant to the nature and scope of the neurotoxicant. The FDA currently utilizes a category system of chemical structures to assign new chemicals to levels of concern for required testing (Food and Drug Administration, 1982). This system is applied widely to carcinogenicity testing and could be adapted for application to neurotoxicity testing.

The Panel concluded that knowledge of SAR should be used wherever such information is adequate to identify classes of compounds with known neurotoxic potential for required neurobehavioral toxicity assessment. There is a need to further develop SAR-based approaches to neurotoxicology.

D. SECONDARY TIER TESTING

Secondary tier testing addresses those specialized tests which are employed to further characterize the nature and scope of the neurotoxicant's effect on the specific neural component identified by primary tier procedures. The objectives of the secondary tier are to further characterize the nature of the neurotoxic effect, determine the extent of nervous system involvement, and refine information for dose/time effect relationships.
Within the framework of applied toxicology testing, the Panel concluded that these specialized neurobehavioral procedures could be incorporated into chronic, reproduction, and developmental study protocols. Secondary tier procedures require a satellite group of animals because the test battery may be time consuming, requiring both specialized instrumentation and technician training, as well as additional handling of the animals in a modified environment. Although information derived from this detailed testing could be useful for determining the mechanism of action for the test substance, this was not considered to be a requirement of applied neurobehavioral toxicity testing. Identification of the mechanism of action requires a level of testing that goes beyond neurobehavioral screening methods and is dependent on detailed neurochemical and neuropathological research investigations.

A comprehensive review of the strengths and limitations of available neurobehavioral testing procedures has been prepared by Tilson and Mitchell (1984). Tilson (see Part B, p.V-1-54) has described the application of several of these procedures in applied neurobehavioral toxicity testing. Primary and secondary tier procedures for developmental testing have been reviewed by Vorhees (see Part B, p.VI-1-60).

The Panel concluded that the extent to which a petitioner would pursue secondary tier testing is dependent upon findings obtained from primary tier testing, their interpretation by the petitioner and FDA, as well as the vested interests of the petitioner. Should a decision be made to proceed, the selection of the appropriate testing procedures should be determined on the basis of open discussion between the FDA and the petitioner.
VI. SUMMARY OF CONCLUSIONS AND RECOMMENDATIONS

These conclusions and recommendations address issues of neurobehavioral toxicity in the framework of the responsibilities of Federal regulatory agencies to protect public safety. They represent the opinions of the Expert Panel on Predicting Neurotoxicity and Behavioral Dysfunction from Preclinical Toxicologic Data. The Panel convened to address three questions on neurotoxicity posed by the Center for Food Safety and Applied Nutrition, FDA (see p.3). The Panel based its conclusions and recommendations on two important considerations. First, neurobehavioral dysfunction is an identifiable scientific issue in which new scientific information has emerged rapidly during the past 20 years because of progress in all areas of the neurosciences; and, second, this new knowledge is having a substantial impact on applied toxicology.

Conclusion:

The present understanding of neurobehavioral toxicity derived from conventional toxicology studies is compromised by deficiencies in study protocols, data collection methods, and identification of appropriate endpoints of neurotoxicity.

Recommendation:

a. There are a number of behavioral changes and neurotoxicological events that could be observed during conventional acute, subacute, subchronic, reproduction, developmental, and chronic toxicology studies provided that guidelines on appropriate evaluation procedures can be adopted. There is a need to catalogue and prioritize such information and to implement a systematic set of functional observational procedures for use in safety evaluation studies. Current guidelines should be modified to ensure that the limited information potentially available from current protocols is not lost and that there is inter and intra-laboratory uniformity in recording and reporting this information.

b. Because behavior is a functional indicator of the net sensory, motor, arousal, motivational and other integrative processes that occur in the central and peripheral nervous systems, behavioral tests are useful in the assessment of neurotoxicity; however, observational rating procedures are not always sufficient by themselves as screening methods for neurobehavioral toxicity. Quantitative, observer-independent tests of behavior (e.g., motor activity)
should be conducted under controlled conditions. Automated tests of motor activity provide an opportunity to obtain quantitative data on behavioral changes free of observer judgment and have been shown to reflect toxicant-induced changes in nervous system function.

Tests other than motor activity should be considered (e.g., instrumental learning). The selection of these tests should be determined by knowledgeable and experienced neuroscientists.

**Conclusion:**

The limited neurotoxicity data that are available from conventional toxicity screening studies as currently conducted are not reported in a manner that permits a meaningful assessment of nervous system integrity.

**Recommendation:**

Guidelines for reporting data should be developed that detail the functional, neuropathological, biochemical, and behavioral data to be collected from existing protocols. The cumulative findings of these data should be reported in a separate statement describing the structural and functional integrity of the nervous system.

**Conclusion:**

The choice of tests to be included in an assessment of neurotoxicity will depend on the level of knowledge sought. It is cost prohibitive to include extensive behavioral testing for every substance in the initial stages of a neurobehavioral toxicity screening battery. The procedures needed for initial neurobehavioral toxicity screening should sample a wide range of functions.

**Recommendation:**

a. A tier approach to detecting neurobehavioral toxicity should be included in acute, subacute, subchronic, and developmental protocols. The initial tier would include a standardized functional observational battery and an automated test of motor activity.* When positive findings from behavioral screening

---

* Some members of the Panel recommended that a test of instrumental learning also be included in the first tier.
tests are obtained, they would serve to indicate changes in nervous system function and provide a basis for specialized testing (secondary level) to characterize further the nature and extent of neurotoxicity. Second level testing would include measures of motor and sensory function as well as tests of cognitive abilities.

b. Current knowledge on molecular structure-activity relationships (SAR) should be used wherever such information is adequate to identify classes of compounds with known neurotoxic potential for required neurobehavioral toxicity assessment. There is a need to develop further SAR-based approaches to neurotoxicology.

Conclusion:

Although normal behavior and toxicant-induced changes in behavior and the underlying neurotoxic mechanisms are becoming increasingly well understood, there is no uniformly accepted testing method for evaluating behavioral toxicity within applied toxicology study protocols. Modifications in the current test guidelines to include consideration of potential neurobehavioral toxicity will enhance our ability to detect neurotoxic compounds in accepted toxicity study protocols. In the absence of a definitive testing protocol for assessing neurobehavioral toxicity, scientists should anticipate some neurotoxic compounds may not be detected by pre-clinical testing in laboratory animals.

Recommendation:

a. There is a need to maintain some flexibility when preparing test guidelines for neurobehavioral toxicity. Such an approach should anticipate that additional and/or improved neurobehavioral test methods from this field will be refined as understanding of basic neurotoxic mechanisms improves.

b. The FDA should reserve the right to request further testing of specific compounds when unanticipated neurotoxic reactions are suspected or demonstrated in humans. The protocol for these supplemental studies, specifically designed to cross-validate animal findings with clinical observations in humans, should be developed from a cooperative dialogue between clinicians, regulatory scientists, and concerned investigators.
• **Conclusion:**

Neurotoxicology is a multidisciplinary field that seeks to integrate data derived from behavioral, neurological, neurophysiological, neuropathological, and neurochemical studies to evaluate toxicant-induced changes in nervous system function. Both the collection and evaluation of these data should be reviewed by scientists experienced in integrating the numerous data elements from the variety of test protocols that are used in assessing nervous system function.

**Recommendation:**

a. The development of study protocols and the data collected from each neurobehavioral toxicology study (e.g., neuropathology, neurochemistry, neurobehavior) should be reviewed by scientists trained in neurobehavioral toxicology and the appropriate basic science discipline.

b. There is a need to develop training and certification programs for laboratory personnel responsible for the conduct of behavioral testing and the recording of animal observations. For example, the technician certification program supported by the American Association for Laboratory Animal Sciences could expand available training programs to include neurobehavior.
VII. LITERATURE CITED


VIII. STUDY PARTICIPANTS

**CO-CHAIRMEN**

<table>
<thead>
<tr>
<th>C. Wayne Callaway, M.D.</th>
<th>Donald E. Hutchings, Ph.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Center for Clinical Nutrition</td>
<td>Department of Developmental Psychobiology</td>
</tr>
<tr>
<td>George Washington University</td>
<td>New York State Psychiatric Institute</td>
</tr>
<tr>
<td>Washington, D.C. 20037</td>
<td>New York, New York 10032</td>
</tr>
</tbody>
</table>

**SYMPOSIUM SPEAKERS**

<table>
<thead>
<tr>
<th>Richard B. Mailman, Ph.D.</th>
<th>Marshall Steinberg, Ph.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department of Psychiatry and Pharmacology</td>
<td>Hazleton Laboratories Corporation</td>
</tr>
<tr>
<td>University of North Carolina School of Medicine</td>
<td>Herndon, Virginia 22071</td>
</tr>
<tr>
<td>Chapel Hill, North Carolina 27514</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nicholas P. Plotnikoff, Ph.D.</th>
<th>Hugh A. Tilson, Ph.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Department of Pharmacology</td>
<td>Laboratory of Behavioral and Neurologic Toxicology</td>
</tr>
<tr>
<td>Oral Roberts University School of Medicine</td>
<td>National Institute for Environmental Health Sciences</td>
</tr>
<tr>
<td>Tulsa, Oklahoma 74138</td>
<td>Research Triangle Park, North Carolina 27709</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>William F. Sette, Ph.D.</th>
<th>Charles V. Vorhees, Ph.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Effects Division Office of Pesticide Programs U.S. Environmental Protection Agency</td>
<td>Childrens Hospital Medical Center</td>
</tr>
<tr>
<td>Washington, D.C. 20460</td>
<td>University of Cincinnati Cincinnati, Ohio 45229</td>
</tr>
</tbody>
</table>
SYMPOSIUM DISCUSSION PANEL

Louis W. Chang, Ph.D.
Department of Experimental Pathology
University of Arkansas
for Medical Sciences
Little Rock, Arkansas 72205

Joseph L. Jacobson, Ph.D.*
Department of Psychology
Wayne State University
Detroit, Michigan 48202

Raef K. Haddad, Ph.D.
New York State Office of Developmental Disabilities and Mental Retardation
Institute for Basic Research in Developmental Disabilities and Mental Retardation
Staten Island, New York 10314

Victor G. Laties, Ph.D.
Division of Toxicology
Department of Radiation Biology and Biophysics
University of Rochester
School of Medicine
Rochester, New York 14642

Richard M. Hoar, Ph.D.
Findley Research Inc.
Fall River, Massachusetts 02722

Lawrence W. Reiter, Ph.D.
Neurotoxicology Division
U.S. Environmental Protection Agency
Research Triangle Park,
North Carolina 27711

James R. Wilson, Ph.D.
Institute for Behavioral Genetics
University of Colorado
Boulder, Colorado 80309

* Unable to attend
FOOD AND DRUG ADMINISTRATION
200 C Street, S.W.
Washington, D.C. 20204

C. William Cooper, M.S.
Contract Office Technical Representative

Thomas J. Sobotka, Ph.D.
Toxicologist

LIFE SCIENCES RESEARCH OFFICE
9650 Rockville Pike
Bethesda, Maryland 20814

Kenneth D. Fisher, Ph.D.
Director

Richard W. Leukroth, Jr., M.S.
Project Coordinator

OTHER CONTRIBUTING LIFE SCIENCES STAFF

Sue Ann Anderson, Ph.D.
Senior Staff Scientist

Beverly R. Lea
Technical Services Manager

Harolyn Cohen
Secretary

Judith Miller
Administrative Assistant

Gloria J. Cole
Secretary

Susan M. Pilch, Ph.D.
Staff Scientist

Barbara L. Durant
Secretary

Stephen H. Simpson
Technical Literature Assistant

Louise S. Erlick
Librarian

Rose Soulen
Secretary

Sandra Gordon
Secretary

Marty Watt
Secretary
Proceedings of the Symposium on

PREDICTING NEUROTOXICITY AND BEHAVIORAL DYSFUNCTION
FROM PRECLINICAL TOXICOLOGIC DATA

Presented by:

Food and Drug Administration
Center for Food Safety and Applied Nutrition

and

Life Sciences Research Office
Federation of American Societies for Experimental Biology

September 30, 1985
Bethesda, Maryland
I. SYMPOSIUM OVERVIEW

Donald E. Hutchings, Ph.D., C. Wayne Callaway, M.D., and Thomas J. Sobotka, Ph.D.

This symposium addresses the problems of predicting neurobehavioral toxicity of foods, drugs, and cosmetic ingredients based upon animal test results. Since the 1970s, more than a dozen scientific meetings have been devoted to various facets of neurotoxicologic assessment (Table I.1). The focus of these meetings has changed as the science developed and in response to new questions and regulatory issues. The initial meetings that convened in the early 1970s focused on the availability of test methods and the selection of test-paradigms. From the mid 1970s to the early 1980s, the focus was on the data generated from different classes of chemicals, with special emphasis on the sensitivity of various neurobehavioral test methods. In recent years, attention has focused on test strategies, intra- as well as inter-laboratory reliability, multidisciplinary testing, and validity of neurobehavioral data in predicting neurotoxicity.

Under present regulatory guidelines (Food and Drug Administration, 1982), the FDA has the authority to require special testing of any chemical substance when there is any evidence of a potential neurotoxicological problem. At this time however, the agency does not require specific neurobehavioral testing as a routine screening requirement under the existing FDA toxicology test guidelines. This symposium seeks to address the following major issues: a) what is the nature and extent of neurotoxicological information which is presently reported to the FDA from toxicological data collected according to existing test guidelines; b) what additional information can be gleaned from the data as submitted; c) to what extent do we need to expand these guidelines to include state-of-the-art neurotoxicologic assessment techniques?

Of particular concern for contemporary researchers, regulatory scientists, and industry is the controversy surrounding the use of neurobehavioral data in the detection of chemically-induced neurotoxicity and the extent to which such data derived from animal studies can be extrapolated to humans. While we can not presume to address all of the aspects of this controversy at this symposium, it is our goal to present a balanced review of the science involved and to consider a series of specific questions formulated by the Center for Food Safety and Applied Nutrition at the Food and Drug Administration about part of this controversy.
Table I.1. Representative symposia reports and conferences on neurotoxicity and behavioral testing in recent years.

<table>
<thead>
<tr>
<th>Title/Sponsor</th>
<th>Primary Focus</th>
<th>Source and Year of Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurotoxicology in the Fetus and Child (University of Arkansas)</td>
<td>Chemicals</td>
<td>Spyker-Cranmer (1986)</td>
</tr>
<tr>
<td>The Collaborative Behavioral Teratology Study (National Center for Toxicology Research of the Food and Drug Administration)</td>
<td>Methods, Reliability, and Sensitivity</td>
<td>Kimmel and Buelke-Sam (1985)</td>
</tr>
<tr>
<td>Nutrient Intake, Brain Biochemistry, and Behavior (American Society for Clinical Nutrition)</td>
<td>Methods and Nutrients</td>
<td>Pollitt and Read (1985)</td>
</tr>
<tr>
<td>Second International Conference on Neurotoxicology of Selected Chemicals (University of Arkansas)</td>
<td>Chemicals (1985)</td>
<td>Spyker-Cranmer</td>
</tr>
<tr>
<td>Application of Neurophysiological Techniques to Toxicological Problems (Society of Toxicology)</td>
<td>Methods</td>
<td>Woolley (1985)</td>
</tr>
<tr>
<td>Workshop on Neurotoxicity Testing in Human Populations (U.S. Environmental Protection Agency, University of North Carolina at Chapel Hill, and Albert Einstein College of Medicine)</td>
<td>Methods</td>
<td>Otto and Eckerman (1985)</td>
</tr>
<tr>
<td>Disease and Chemically Induced Neurological Dysfunction (U.S. Environmental Protection Agency)</td>
<td>Methods and Chemicals</td>
<td>Annau and Dyer (1983)</td>
</tr>
<tr>
<td>Title/Sponsor</td>
<td>Primary Focus</td>
<td>Source and Year of Report</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>The Effects of Foods and Drugs on the Development and Function of the Nervous System: Methods for Predicting Toxicity (Food and Drug Administration)</td>
<td>Methods and Chemicals</td>
<td>Gryder and Frankos (1980)</td>
</tr>
<tr>
<td>Test Methods for Definition of Effects of Toxic Substances on Behavior and Neuromotor Function (U.S. Environmental Protection Agency)</td>
<td>Methods</td>
<td>Geller et al. (1979)</td>
</tr>
<tr>
<td>Target Organ Toxicity - Nervous System (National Institute of Environmental Health Science)</td>
<td>Methods</td>
<td>Mitchell (1978)</td>
</tr>
<tr>
<td>Behavioral Toxicology: An Emerging Discipline (U.S. Environmental Protection Agency)</td>
<td>Methods</td>
<td>Zenick and Reiter (1977)</td>
</tr>
<tr>
<td>Target Organ Toxicity - Behavioral Teratology (National Institute of Environmental Health Science)</td>
<td>Methods</td>
<td>Kimmel (1976)</td>
</tr>
<tr>
<td>Behavioral Toxicology (The University of Rochester)</td>
<td>Methods</td>
<td>Weiss and Laties (1975)</td>
</tr>
<tr>
<td>Behavioral Toxicology Early Detection of Occupational Hazards (National Institute of Occupational Safety and Health)</td>
<td>Methods</td>
<td>Xintaras et al. (1974)</td>
</tr>
</tbody>
</table>
The stated mission of CFSAN/FDA is to "promote and protect the public health and welfare by assuring that the food supply is safe, nutritious, wholesome and honestly labeled." In the realm of government regulatory policy, the uninitiated scientist who is oriented in a specific research discipline, may become confused when the scientific issues are circumscribed in the unfamiliar language of legal regulatory terminology. The orderly and rational view of the scientist who traditionally attempts to define and quantify functional relationships between dependent and independent variables finds that such terms as "safe", "wholesome", and "nutritious" are too imprecise. Regardless of the ambiguous nature of these quasi-scientific terms, they espouse the public concern for safety and quality of life. With increasing frequency, scientists are called upon to assist regulatory agencies to clarify the scientific principles implied by such legal terminology.

The concerns of the average consumer are primarily related to product claims rather than to the scientific issues of safety. These later concerns are conveniently delegated to the FDA or other governmental agencies whose mission is to protect the public interest. This trust sounds deceptively simple. As a result of conflicting scientific findings and/or changing social priorities, decisions about safety and what it means are not always straightforward and certainly not easy.

Consider for a moment the conflicting information to which the public is frequently exposed. For example, several years ago the Mayor of San Francisco was publicly murdered. Under the extenuating circumstances in which the convicted killer claimed to be affected by an overdose of "junk-food", the court apparently reduced the defendant's sentence (Zimring, 1984). In the lay press this became known as the "Twinkie® defense" or the "toxic junk-food syndrome" but it is more appropriately referred to as the "diminished capacity defense" (Blodgett, 1985). In another example, a group of scientists recently reported on a study that megadoses of vitamin B caused severe, debilitating damage to sensory nerves (Schaumbürg et al., 1983; Windebank et al., 1984). In these examples, we can see the public dilemma. Twinkies® are as "American as apple pie", and approximately one-half of the adult American population takes some form of vitamin supplement on the supposition that they will improve their health. What is the public to make of these news items now that both appear to pose a serious public health hazard? The American Dietetic Association in conjunction with the National Council Against Health Fraud has recently endorsed a position paper on diet and criminal behavior prepared by the California Council Against Health Fraud (American Dietetic Association, 1985). This position paper recognized the need for scientifically sound interdisciplinary research with appropriate controls and the careful interpretation of data.
In cases such as these, the public has turned to the FDA for either answers or action. As described in a paper presented at a recent symposium on diet and behavior (Sobota, 1986), there are many misconceptions about the role of the FDA in these and other regulatory problems. The agency relies upon a variety of broadly defined safety standards depending upon the regulatory category of the food substance in question. For those substances which fall into categories of general foods or nutrients, safety is based upon the lack of evidence of actual or potential harm to consumers. In this instance, the burden of proof is on the FDA to produce such evidence before the agency is allowed to take any action against the substance in question. On the other hand, the regulations state that a reasonable certainty of no harm must be established for food and color additives. Industry is required to submit to the FDA results from pre-market safety tests prior to the approval of any new food or color additive. In both cases, the availability of credible scientific evidence is key to the agency's decision-making process.

The agency has traditionally approached the evaluation of safety along the lines of conventional toxicology, which include the generally accepted indices of toxicity, such as mortality, pathology, carcinogenesis, growth disorders, clinical abnormalities, organ system impairment or failure, reproductive disorders, and teratogenic potential. There has been a growing opinion among regulatory scientists specializing in the neurosciences, that such conventional measures of toxicity may not be sufficiently comprehensive to detect all potential adverse effects. There is recognition that drugs and environmental and dietary chemicals can influence the neuronal control of behavioral processes such as learning, attention, sensorimotor performance, sleep, and mood. This has generated considerable interest among the scientific and regulatory communities, as well as concern among the general public. Brain function is influenced by such chemicals. This has compelled toxicologists to reevaluate their conceptual definition of toxicity to include the fact that changes in neuronal function may occur in conjunction with, prior to, or perhaps even in the absence of other more conventional signs of toxicity. In terms of safety assessment, one of the many questions facing the FDA is whether to expand the database to include not only the traditional measures of pathological toxicity but also specific measures of brain dysfunction. A fundamental purpose of this symposium is to review the available information on this subject, to provide this symposium panel with the information needed to evaluate these questions, and to report the study findings to the FDA.

The first paper presents an overview of the traditional animal toxicologic screening protocols used by industry to obtain data in support of petitions for chemical use as submitted to government regulatory agencies. In this presentation, Dr. Marshall Steinberg describes the myriad of data elements collected from
these protocols and their use or potential application for neurotoxicologic assessment. The paper focuses on the importance of current methods for collecting and tabulating cage-side observations, model selection, environmental control, and technician training. In addition, the paper reviews the neurotoxicity test requirement differences between domestic and foreign regulatory agencies. It contains discussions of the value of including an interpretation of observational findings with those of other test results into the final report.

Dr. William F. Sette describes neurotoxicologic assessment from the viewpoint of the regulatory scientist who is responsible for hazard assessment. This paper is a review of the integral role of the central nervous system in controlling body function. It focuses on the complexity of differentiating between neurologic dysfunction and other forms of target organ toxicity as determined by data from conventional toxicity and behavioral studies. The author discusses options for improving existing study protocols, the use of explicit testing, and the development of generic testing requirements for specific classes of chemicals.

Advances in the neurosciences have led to improved understanding of the anatomical, biochemical, and molecular loci involved in the injury to, and adaptation of, the central nervous system (CNS) to toxic insult. In the paper on mechanisms of CNS injury and behavioral dysfunction, Dr. Richard B. Mailman describes the differences of toxicant dynamics and the plethora of possible molecular endpoints and biochemical targets in the central nervous system. By describing the examples of the effects of FD&C Red No. 3 and MPTP (1-methyl-4-phenyl-2,3,4,6-tetrahydropyridine) on the dopaminergic system, he notes that these problems will probably preclude the likelihood that neurochemical or biochemical tests will be useful for screening or detecting unknown neurotoxicants in the near future.

There are numerous laboratory techniques available to the toxicologist for assessing behavioral dysfunction in animal models. These tests which are characteristically noninvasive provide sensitive measurements of the effects of neurotoxic agents. The review of available behavioral testing methods by Dr. Hugh A. Tilson elucidates the criteria for classifying these tests. He offers examples of neurobehavioral tests currently used to evaluate the behavioral toxicity of chemicals that affect motor function, sensory function, learning/memory performance, instrumental performance, and naturally occurring responses. He describes the value of a systematized functional observational battery in the provisional assessment of potentially neurotoxic chemicals. An indication of neurotoxicity as detected at this level could be further characterized by any number of secondary tests.
Common to the complexity of hazard assessment in any field are the questions of sensitivity, reliability, and validity of the tests included in a study protocol. Dr. Charles V. Vorhees provides the panel with a review of the concerns and considerations related to these issues in the science of neuro-behavioral toxicology. He describes the findings of collaborative behavioral studies and reports on the behavioral trends observed in a study of historical controls. This paper demonstrates that behavioral techniques have shown a high degree of sensitivity, reliability, and validity. However, further research is needed to broaden the database and to provide additional examples for the correlation between neurochemical/neuropathological damage and behavioral dysfunction. At present, behavioral tests offer the most comprehensive means of evaluating CNS injury because other neurological assays are highly specific and difficult to conduct in a large scale screening program.

The closing paper at this symposium provides a review of advances in the new and rapidly developing field of psychoneuro-immunology. Dr. Nicholas P. Plotnikoff describes the research approaches used to demonstrate the interrelationship between central sites within the brain that have modulatory influences on the immune system. The intricate feedback loops by which the autonomic and neuroendocrine systems interact with the reticulo-endothelial system are discussed in relation to environmental stress and the role of pro-hormones. Measurements of the interaction among these systems may provide an opportunity to predict early changes in CNS function and behavior.
LITERATURE CITED


II. THE USE OF TRADITIONAL TOXICOLOGIC DATA IN ASSESSING NEUROBEHAVIORAL DYSFUNCTION

Marshall Steinberg, Ph.D.
Vice President and Scientific Director
Hazleton Laboratories Corporation
Renaissance Center
13873 Park Center Road
Herndon, Virginia 22071

ABSTRACT

In the United States, the regulation and testing of chemicals that may impact on human health is mandated by a number of different laws and regulated by several government agencies. Toxicologic screening in laboratory animals has created a major database for predicting adverse health effects of drugs, food and cosmetic ingredients, and manufactured chemicals. Test protocols and reporting requirements frequently differ between government agencies within the United States and between nations. Until recently, the emphasis for this required testing has been to identify those chemicals that may result in cancer or birth defects with little or no emphasis placed on neurobehavioral testing. The so-called "traditional toxicologic screen" presently includes acute, subchronic, chronic, teratology, and reproductive multispecies testing.

The report for each segment of the test screen contains a myriad of data elements that have been observed by numerous technicians in the testing laboratory. These laboratory tests vary from general animal observations of animal body weight and food consumption to tests for hematology, blood chemistry, urinalysis, and histopathology. In addition, specialized tests may include other physiological measurements, neurologic tests, eye exams as well as functional testing to evaluate performance (e.g., open field, swimming maze, and T-maze). Reliability, quality control, and the difficulty in converting technician observation into quantifiable data elements that are free from bias is a primary concern for all testing laboratories. This paper focuses on the neurotoxicologic and behavioral information that can
be gleaned from traditional toxicologic screens, the collection of study data, the effect of model selection, and the environmental considerations that may affect study outcome. The paper includes examples of compounds previously reported in the literature.
A. INTRODUCTION

If one describes behavioral toxicology as the study of the effects of toxicants in relation to an animal's conduct, then one might describe classical toxicology as peripheral organ function toxicology. The art of classical toxicology seeks to determine the effect of a dose level at which either morphology or organ function is changed, and then seeks to determine the level at which that change does not occur, the so-called no-effect level (NOEL). Classical toxicology requires a quantitation and an evaluation of the study findings.

Often data are collected that can be used in a manner not originally intended. Examples may be found in toxicology involving the immune system (e.g., spleen to body weight ratios, thymus weight, white cell counts) and neurobehavioral toxicology (e.g., animal activity, cage behavior, motor activity, food consumption). The purpose of this paper is to demonstrate that traditional test techniques provide the opportunity to collect data that can be of value when attempting to determine if a material produces neurobehavioral changes.

B. REGULATORY GUIDELINES FOR TOXICITY TESTING

In the conduct of toxicological testing, specific descriptive animal toxicity tests have been developed within the context of regulatory guidelines. The bases for this testing are the toxicological principles that: 1) the effects observed in laboratory animals are applicable to humans and 2) that the exposure of laboratory animals to toxic substances in high doses is necessary to reveal possible hazards in humans. Numerous federal and international government agencies use findings from toxicology studies of laboratory animals to assess potential risk to humans.

1. Classical toxicity testing protocols

The following listing of available test methods are presently used by toxicologists who are involved in risk assessment (Doull et al., 1980; National Research Council, 1975).

a. LD50 study

The LD50 study or some variation thereof is designed to identify acute toxicity dose levels and associated clinical signs. The use of this test is being reduced and the replacements measure either the minimum lethal dose or the approximate lethal dose or some variant. This change is being driven by the desire to decrease the use of animals for test procedures of questionable value.
b. **Skin or eye irritation**

These tests are designed to measure safety rather than toxicity. They are also of questionable scientific value and in vitro alternatives are being sought. Most animal tests of this type tend to produce a high level of false positives.

c. **Delayed hypersensitivity**

This test measures the ability of a material to complex with a protein and produce a T-cell mediated delayed hypersensitivity reaction. Variations of this test include activation of the chemical or protein complex by ultraviolet irradiation.

d. **Subchronic toxicity tests**

These tests are usually 28-day to 120-day exposures. Variations of this test involve the use of a variety of dose forms and dosing procedures. They may include dosed feed, dosed water, parenteral administration, dermal application, gavage, intubation, and inhalation. The route of administration is generally intended to mimic potential or intended human exposure. The route is extremely important and can effect the toxic response. With the exception of oncogenicity, and if the study is properly designed, this route of exposure will detect the major toxicity associated with a toxicant. Associated metabolic studies are also performed to determine the absorption, distribution, and excretion of candidate materials. With some modification to existing protocols, behavioral effects could be readily detected by subchronic toxicity testing.

e. **Chronic toxicity**

Chronic studies are generally designed to evaluate the effects of long-term exposure. The doses are generally lower than those used in subchronic studies. These studies are particularly useful in evaluating industrial and environmental toxicants. Pharmaceuticals intended to be given for long periods are also tested in this manner. The selected route of administration for these studies approximates the potential route of human exposure. Oncogenicity may be an endpoint but general toxicity is the intended target. These studies are also particularly useful in evaluating chemicals that tend to sequester in the body such as chlorinated hydrocarbons (e.g., DDT), heavy metals (e.g., mercury) and other materials.
f. **Reproduction tests**

These tests are designed to measure overall reproductive performance and generally are performed to produce two generations of offspring. They measure the effects of direct action on germ cells, effects on accessory organs and influence upon hormonal controlling mechanisms, among other things. There are numerous compounds acting on the central nervous system (CNS) which appear to have an effect upon reproduction (e.g., reserpine, amphetamine, and phenothiazines). The effect of the toxicant may be on the male, the female, or both.

g. **Teratology**

The measurement of congenital defects is accomplished by these studies. The defects may include morphologic, biochemical, or functional abnormalities produced either before or at birth. This type of testing has become more prevalent since the thalidomide tragedy in Europe. Time of exposure during pregnancy is extremely important to the development of the lesions. For example, diphenylhydantoin given to pregnant mice on days 9-10 of gestation will produce CNS effects which will not be seen if the pharmaceutical is given at other times during gestation (Harbison and Becker, 1969).

h. **Carcinogenicity**

The purpose of carcinogenicity studies is to determine if a compound produces an oncogenic response. This study may be combined with a chronic toxicity study. Frequently, a carcinogenicity study will be designed to permit a risk analysis to the human population. This will depend upon the use, or intended use, of the material. The studies may be permitted to extend over the entire lifespan of the test animal, but this may result in a response to a compound being obscured by spontaneously occurring tumors. Variants of such studies permit the identification of complete carcinogens, initiators, and promoters.

A large number of animals must be used to insure statistical validity. Weak carcinogens may not be detected by standard test design. Variants of this test permit the identification of dose levels which produce an acceptable risk level to man.
1. **Mutagenicity**

These studies are conducted in vitro, in vivo, or combined. Surrogate cell cultures, bacteria, yeast, flies, whole animals, and a variety of other life forms have been used. These tests are extremely useful in screening compounds and elucidating mechanisms of action. Mutagenesis tests measure the ability of a chemical to cause changes in the genetic material in a cell in ways that can be transmitted during cell division (Doull, 1980).

2. **Elements of toxicological data sets**

The performance of any of the traditional tests requires meticulous collection of the data followed by a careful evaluation of the different types of data and their possible interrelationships.

There are incredible amounts of data collected during the course of a chronic study; we estimate as many as 700,000 data points. For example, a subchronic 90-day toxicity study requires the collection, verification, comparison, and analysis of data points that include clinical observations, clinical pathology, analytical chemistry, anatomic pathology, environmental monitoring and when required, an evaluation by biological science specialists such as veterinarians, pathologists, and biochemists. The rodent protocol for a subchronic study routinely provides for twice daily observations of morbidity and morbidity and may or may not indicate requirements for recording these types of data. In addition, such a protocol generally provides for routine data collection, including recording body weights and clinical observations for rodents. This generally occurs once a week for the first 13 weeks of test and at intervals varying from once every 2 weeks to once per month thereafter. Some protocols call for more frequent data collection toward the end of the study. Clinical pathology samples are routinely collected at predetermined intervals during the life of the animals on study, and are usually associated with serial sacrifices. It is important to recognize that a properly designed and conducted toxicology study is a team effort by a group of scientists and technicians who contribute a variety of complementary skills.

3. **The availability of behavioral data from classical toxicity testing protocols**

When conducting a study under the classical guidelines, there is an opportunity to develop some behavioral information. However, to glean this information, one must look beyond the elements that comprise the individual data sets (e.g., food consumption, body weight, organ to body weight ratios) and
look at the animal as a functioning unit. One might consider this "holistic toxicology", with the individual data elements providing sign posts that are used in the search for the site of the biochemical lesion(s). In pursuit of the behavioral effects that might relate to this biochemical lesion, it is not only necessary to observe the animal's cage activity or curiosity when being weighed or placed in a new cage, but it is also necessary to evaluate how the animal responds when handled. These observations may provide useful information that animals in the high dose group were more lethargic or aggressive than the animals in the low doses or control groups. Another example of the importance of general observations is that of fighting behavior in the case of group housed male mice, which may be increased or decreased as a consequence of study treatment. Sometimes these observer impressions are not recorded either owing to no provision having been made for their capture, or were lost owing to the inexperience of the observer with a particular species or strain of test animal. In some instances the effects are too subtle to be noted under the testing conditions. In other cases, it may be necessary to conduct specialized studies (e.g., conditional avoidance, T-maze, strength tests, rotorod) to discern or quantify subtle biological effects, and to evaluate specific pathologic or biochemical changes (National Research Council, 1982).

There are interrelationships between the chemical lesion, the observable pathology, the functional change, and behavioral change (Figure II.1). The basis for all toxicity is a biochemical lesion. This biochemical lesion may or may not result in an observable pathology and the observable pathology may or may not result in a functional change outside the CNS. When evaluating data from toxicity testing, consideration must also be provided for the compensatory mechanisms present in almost all organ systems that may not permit the biochemical lesion to be expressed as an observable change.

4. Federal and international neurotoxicity testing guidelines

With the exception of the U.S. Environmental Protection Agency (EPA), under the terms of the Toxic Substances Control Act (TSCA) (U.S. Environmental Protection Agency, 1985), there are presently no formal regulatory guidelines for the conduct of behavioral testing in countries developing new and significant pharmaceutical or agrichemicals. Several countries require that behavioral testing be included in the testing program without specifying detailed guidelines for the type of testing to be performed (Table II.1) (Japan, Ministry of Health and Welfare, 1982, 1984). In the United States, the Food and Drug Administration (FDA) may require additional behavioral
Figure II.1. Interrelationships between toxicant and effect.
Table II.1. Regulatory guidelines for neurotoxicity and delayed neurotoxicity testing.

<table>
<thead>
<tr>
<th>TEST</th>
<th>FIFRA\textsuperscript{a}</th>
<th>TSCA\textsuperscript{b}</th>
<th>FDA\textsuperscript{c}</th>
<th>OECD\textsuperscript{d}</th>
<th>EEC\textsuperscript{e}</th>
<th>MAFF\textsuperscript{f}</th>
<th>MOHW\textsuperscript{g}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurotoxicity</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Delayed Neurotoxicity</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

\textsuperscript{+} Current guidelines that describe testing.
\textsuperscript{-} Current guidelines that do not describe testing.

\textsuperscript{a} Federal Insecticide, Fungicide, and Rodenticide Act (1972).
\textsuperscript{b} Toxic Substances Control Act (1976).
\textsuperscript{c} Food Substances Control Act (Food and Drug Administration, 1982).
\textsuperscript{d} Organisation for Economic Co-operation and Development (1981).
\textsuperscript{e} European Economic Community (Council of the European Communities, 1979).
\textsuperscript{f} Ministry of Agriculture, Forestry and Fisheries (Japan, Ministry of Agriculture, Forestry and Fisheries, 1985).
\textsuperscript{g} Ministry of Health and Welfare (Japan, Ministry of Health and Welfare, 1984).
testing, but only if supplemental behavioral tests can be justified on the basis of findings from other nonbehaviorally-oriented testing protocols. This is unlike other FDA toxicological testing guidelines (e.g., carcinogenicity testing food additives) which clearly define the testing requirements. The regulations set forth by the Ministry of Health and Welfare in Japan serve as an example of a requirement for behavioral testing of pharmaceuticals without defining what tests are to be included in the study protocol.

The Ministry of Health and Welfare in Japan frequently requires three types of behavioral testing for evaluations of new products. In their reproduction study testing requirements, two males and two females per litter might be examined using the tail-flick test and pupillary reflex test at 21 days postpartum. At 6 to 7 weeks postpartum, 15 males and 15 females per group might be selected to conduct an open field evaluation. At 8 to 9 weeks, animals from the same group of test subjects might be used to perform a water "T" maze test. This behavioral evaluation and variations thereof are acceptable to the Ministry of Health and Welfare as part of the data package required for registration of pharmaceuticals in Japan (Japan, Ministry of Health and Welfare, 1984). The registration of pesticides in Japan requires that a pharmacological activity screen be conducted (Japan, Ministry of Agriculture, Forestry and Fisheries, 1985). Elements of the pharmacologic screen indicate that some behavioral testing is desired; however, the details are unspecified.

The European Economic Community, Sixth Amendment Guidelines (Council of the European Communities, 1979) do not address behavioral testing. The Organisation for Economic Co-operation and Development (OECD) does address behavioral testing in a manner similar to the FDA (Organisation for Economic Co-operation and Development, 1981). The OECD guidelines, however, do address specific tests for determining the neurotoxicity of organophosphorus compounds. In 1982, the OECD recognized the limitations of its existing neurotoxicity testing protocols and recommended the development of guidelines for more general neurotoxicity testing to include functional behavioral assessments and neuropathologic examination (National Research Council, 1984).

A variety of behavior screens have been used in conjunction with standard classical toxicology test protocols (Geller et al., 1979). These have ranged from the use of clickers for response to auditory stimuli, to a battery such as the Irwin screen (Irwin, 1968). Some testing laboratories use the rotorod, treadmill, gait measurement, and other tests routinely in toxicologic evaluations as determined by either the class of compound to be tested, or by management decision.
C. LABORATORY OBSERVATIONS OF STUDY ANIMALS

Animal observations and clinical signs can be readily incorporated into traditional testing protocols by using a multi-level categorization coding system to direct the laboratory technician and develop a database for study comparisons. Several different systems exist and may be purchased from commercial sources. Common signs (Table II.2) can be examined, including motor activity, respiratory effects, convulsions, ocular signs, cardiovascular signs, salivation, piloerection, analgesia, muscle tone, gastrointestinal signs (to include the feces, emesis, and diuresis), reflex activity, and the condition of the skin. Once defined, these signs can be combined with expressions of gradation and coded for computer entry (Table II.3). This system may include data elements related to housing, food and water consumption, appearance, behavior, study status (e.g., reason for removing the animal from study), excreta, respiration, skin and pelage, the condition of the eyes, tissue masses, and other abnormalities both descriptive and positional. Each major category may be expanded to provide additional information. For example, the appearance category could be further defined by eight additional subcategories (Table II.4) to include data regarding whether the animal is hunched, thin or obese, whether the teeth have appeared on time and if there is a malocclusion problem. With proper handling and observation of the animal, the investigator can easily note head tilt, paralysis, or tremors. A categorization system such as this can also provide an opportunity for free text entry under the heading of "other" observations. The use of a computer can greatly facilitate the collection of these data.

1. Animal conduct

An expansion of the behavior category (Table II.5) offers the investigator an opportunity to characterize animal observations using terms such as languid, prostrate, hyperactive, circling, convulsion, or ataxia. Additional comments might include anorexia or sensitivity to touch. Opportunities for describing the behavior of test animals are provided, but as with the appearance category, findings not specifically subcategorized require the addition of descriptive text.

2. Food and water consumption

Food and water consumption data (Table II.6) are often useful to the toxicologist and may provide the first indication that the test material has exerted an effect upon the test animal. An increase or decrease in food consumption may either indicate a change in appetite, a change in food utilization efficiency, or a variety of other compound effects. In relation to food consumption, the investigator should also
Table II.2. Categorization approach to observing common clinical signs in study animals.

<table>
<thead>
<tr>
<th>COMMON CLINICAL SIGNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. MOTOR ACTIVITY</td>
</tr>
<tr>
<td>II. RESPIRATORY ACTIVITY</td>
</tr>
<tr>
<td>III. CONVULSIONS</td>
</tr>
<tr>
<td>IV. OCULAR SIGNS</td>
</tr>
<tr>
<td>V. REFLEXES</td>
</tr>
<tr>
<td>VI. CARDIOVASCULAR SIGNS</td>
</tr>
<tr>
<td>VII. SALIVATION</td>
</tr>
<tr>
<td>VIII. Piloerection</td>
</tr>
<tr>
<td>IX. Analgesia</td>
</tr>
<tr>
<td>X. Muscle tone</td>
</tr>
<tr>
<td>XI. Gastrointestinal signs:</td>
</tr>
<tr>
<td>a) Feces</td>
</tr>
<tr>
<td>b) Emesis</td>
</tr>
<tr>
<td>c) Diuresis</td>
</tr>
<tr>
<td>XII. Skin</td>
</tr>
</tbody>
</table>
Table II.3. Categories included in a typical clinical observation coding system.

<table>
<thead>
<tr>
<th>Clinical Observation Coding Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Status (e.g., normal, sacrificed)</td>
</tr>
<tr>
<td>Housing, Food, and Water</td>
</tr>
<tr>
<td>Appearance</td>
</tr>
<tr>
<td>Behavior</td>
</tr>
<tr>
<td>Excretion</td>
</tr>
<tr>
<td>Respiration</td>
</tr>
<tr>
<td>Skin and Pelage</td>
</tr>
<tr>
<td>Eyes</td>
</tr>
<tr>
<td>Masses</td>
</tr>
<tr>
<td>Other abnormalities (Descriptive text)</td>
</tr>
</tbody>
</table>

Table II.4. Available subcategories that can be listed under the appearance heading in a typical observation coding system.

<table>
<thead>
<tr>
<th>APPEARANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hunched</td>
</tr>
<tr>
<td>Thin</td>
</tr>
<tr>
<td>Obese</td>
</tr>
<tr>
<td>Teeth</td>
</tr>
<tr>
<td>Malocclusion</td>
</tr>
<tr>
<td>Head tilt</td>
</tr>
<tr>
<td>Paralysis</td>
</tr>
<tr>
<td>Tremors</td>
</tr>
<tr>
<td>Other observations (Descriptive text)</td>
</tr>
</tbody>
</table>
Table II.5. Available subcategories that can be listed under the behavior heading in a typical observational coding system.

<table>
<thead>
<tr>
<th>BEHAVIOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Languid</td>
</tr>
<tr>
<td>Prostrate</td>
</tr>
<tr>
<td>Anorexia</td>
</tr>
<tr>
<td>Hyperactive</td>
</tr>
<tr>
<td>Circling</td>
</tr>
<tr>
<td>Ataxia</td>
</tr>
<tr>
<td>Sensitive to touch</td>
</tr>
<tr>
<td>Convulsions</td>
</tr>
<tr>
<td>Other observations (Descriptive text)</td>
</tr>
</tbody>
</table>

Table II.6. Consummatory data that may provide the initial indication that a test material has exerted an effect on the test animal.

<table>
<thead>
<tr>
<th>Test Material Effects on Consummatory Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in appetite</td>
</tr>
<tr>
<td>Change in food utilization efficiency</td>
</tr>
<tr>
<td>Diarrhea</td>
</tr>
<tr>
<td>Dehydration</td>
</tr>
<tr>
<td>Change in water consumption</td>
</tr>
<tr>
<td>Change in food consumption</td>
</tr>
</tbody>
</table>
include observations for diarrhea and dehydration. Changes in water and food consumption are important not only as an indicator of a possible relationship to central effects, but should be interrelated with other observations such as animal activity. The investigator should not be misled by a change in the food consumption by an older animal as this could be related to a decrease in motor activity caused by a large tumor. A decrease in food consumption may also be indicative of a compound-related effect on motor function. Food consumption data should be used as a trigger to look for other events which might include changes in behavior.

3. Environmental factors

When evaluating behavioral changes, environmental factors must also be examined and controlled. It is possible in the conduct of a reproduction study that an unexplained decrease in the numbers of pregnancies may be owing to the fact that the lights were too bright in the vivarium. In the rat, both blindness and estrus alteration have been demonstrated following exposure to high intensity lighting (DeWitt, 1976; Weihe, 1976). For this reason, it is important that the total animal system (both model selection and vivarium design) be reviewed. This is particularly important when evaluating changes in animal behavior patterns. Environmental stress may also be reflected as behavioral change if the test system is subject to excessive changes in noise, temperature, humidity, and population density (Davis, 1978; Pakes et al., 1984; Weihe, 1976).

4. Motor activity

The loss of the righting reflex is an important element in monitoring motor activity as a part of regulated toxicology testing (Table II.7). The righting reflex is traditionally used to indicate lesions that involve either the CNS or neuromuscular injury. An observation of muscle tremors in study animals is also an important finding. Tremors may be caused by toxicant effects on the CNS, the peripheral nervous system, or the musculature. Tremors may be induced by some of the chlorinated hydrocarbon pesticides (Doull et al., 1980), while the loss of the righting reflex may be associated with exposure to pharmaceuticals (e.g., morphine, amphetamine, alcohol, and chlorpromazine). Anesthesia provides another indication of nerve muscle dysfunction indicating an effect at the level of the sensory organs as well as the CNS. Chemicals commonly involved in producing anesthesia are aliphatic short-chain hydrocarbons such as the halothanes. An increase or decrease in motor activity, breeding behavior, or curiosity can indicate an involvement with the somatic motor or central nervous system (Chan et al., 1982).
Table II.7. Suggestive CNS interpretation of clinical observations for motor activity.

<table>
<thead>
<tr>
<th>Clinical Observation Code</th>
<th>Subcategory Observed Signs</th>
<th>Site of CNS Neurologic Involvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor Activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Loss of righting reflex</td>
<td>Neuromuscular</td>
<td></td>
</tr>
<tr>
<td>Tremors</td>
<td>Neuromuscular</td>
<td></td>
</tr>
<tr>
<td>Anesthesia</td>
<td>Sensory</td>
<td></td>
</tr>
<tr>
<td>Decrease or Increase:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preening</td>
<td>Somatic motor</td>
<td></td>
</tr>
<tr>
<td>Curiosity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
D. ANIMAL MODELS

In risk assessment, the most appropriate model is the species that most closely resembles the human in terms of taxonomy and physiology. The most commonly used animal models in toxicological testing include the rodent (rat, mouse, hamster, guinea pig), dog, rabbit, and monkey. These animals came into favor because their size made them suitable for use in the laboratory. Regulatory agencies require testing in both a rodent and nonrodent mammalian model. In the selection of an animal test model, it is important to assess both the degree of homology and the extent to which it relates to the analogy.

1. **Metabolic considerations in species and strain selection of animals to be tested**

The importance of animal model selection has been well demonstrated in the experience of testing neurotoxins. The exposure to some cholinesterase inhibitors at doses lower than those associated with cholinesterase inhibition have shown that behavioral changes can be demonstrated by behavioral testing techniques such as conditioned avoidance. Laboratory studies with pesticides such as DFP (diisopropyl fluorophosphate), Leptophos, and Mipafox have proven these chemicals to be effective cholinesterase inhibitors (Powell and Lampert, 1977). Neurotoxicologic evaluation of these pesticides would offer a false negative finding in the rodent model; however, in the appropriate animal model (the chicken), these chemicals induce delayed neurotoxicity. It is important to note that for this class of compounds, the accepted animal model for delayed neurotoxicity is not the same as that normally associated with classical toxicologic testing. Therefore the delayed neurotoxic effects of organophosphates on the peripheral nervous system can not be observed in the classical rodent testing protocol. It is particularly important to select the proper model with the appropriate metabolic pathway when designing studies to discern behavioral changes in the classical toxicology models.

2. **The effect of age at the time of testing**

In the conduct of a toxicity evaluation, care should be taken to note the age of the test animal, as age has a significant impact upon the toxic response. Lead and tellurium are examples of compounds producing peripheral as well as central neuropathies more readily in younger animals (Doull, 1980). Conversely, many older animals are more susceptible to "dying back" neuropathies as with bromophenylacetylurea and delayed neurotoxicity associated with some organophosphates.

Consideration of age is a most important element in teratologic testing. As with other organ systems, the developing CNS passes through the process of embryogenesis to organogenesis
which is followed by histogenesis and further progression to functional development during the fetal period. The occurrence of a chemical insult at any time during this continuum may result in a change or alteration which is specific and directly related to time of development. Both Phase I and II developmental testing protocols are designed to study these changes during development (Wilson, 1973).

E. SCIENTIFIC INTERRELATIONS

The interpretation of study findings and the report preparation process is a complex art wherein the myriad of data elements are brought together and scientifically interrelated to evaluate toxicity. Both the duration of neurotoxicant exposure and the mechanism of action are important considerations in a neurotoxicity evaluation.

1. Duration of exposure

Injury to the nervous system may occur as an acute insult or a gradual degenerative process over a lengthy period of exposure to a neurotoxicant. The observed effect may differ depending on the duration of treatment. The neurological findings of studies with hexachlorophene exemplify this relationship (Powell and Lampert, 1977). Twenty-five milligrams per kilogram of hexachlorophene administered in the diet of rodents for a period of 2 weeks can induce hindquarter limb weakness. Within 3 to 5 weeks of continued administration, paralysis is observed. If the investigator had recorded gross observations at a necropsy subsequent to exposure, the following observations could be noted: a reduction in muscle mass, cerebral swelling, a reduction in the size of the brain ventricles, an increase in the size of the corpus callosum, flattened and swollen optic tracts, and occasional fatty livers. These gross pathology findings in portions of the brain are, in themselves, indicative of CNS involvement.

The follow-up histopathologic examination of the nervous system tissues from these animals would show findings that include spongiform alteration of the white matter in the central nervous system as well as segmental demyelination in the peripheral nerves, with abnormalities at the nodes of Ranvier. It is important to note that even if the animal had not been properly observed while on study, the pathology support sciences can frequently detect structural changes within the CNS. Histopathological changes of this magnitude should trigger further evaluation of such a chemical in terms of its impact upon behavior and neurotoxicity. Although this discussion has focused on hexachlorophene, other chemicals such as aminoglycosides, methyl mercury, inorganic mercury, lead, alcohol, triethyl tin, and isonicotinic acid hydrazide are perhaps even better examples of chemical toxicants that effect the CNS.
2. **Mechanism of action**

The resultant effect of many neurotoxins may appear to be the same neurotoxic injury; however, the investigator may find that the mechanism of action is completely different. This is readily apparent with the effects of chemicals on the peripheral nervous system (PNS) at low doses. The net effect of toxins such as botulinum, tetrodotoxin, and saxitoxin is muscle paralysis. Botulinum toxin acts at motor endplates to prevent the release of acetylcholine from the axon terminal (Ambache, 1949; Powell and Lampert, 1977). Following exposure to this toxin, the muscle behaves as though it were denervated. In contrast, the muscle paralysis that follows exposure to tetrodotoxin or saxitoxin is the result of sodium channel blockade along the nerve axon (Hille, 1968; Kao, 1966).

These neurotoxicities are not owing to a CNS effect but can be related to an effect at the PNS level. For the untrained investigator, such effects could easily be misinterpreted as CNS effects; however, at closer evaluation, the histopathology and even the muscle weight to body weight ratios would tend to indicate otherwise.

3. **Multi-level effects**

There are examples of neurotoxic chemicals (e.g., thallium and carbon monoxide) which exert effects at more than one level of the nervous system. The multi-level effects of these chemicals could tend to confuse an interpretation for their site of action. Carbon monoxide has been well studied in a variety of animal models including man and has been demonstrated to induce behavioral effects such as confusion, disorientation, ataxia, and incoordination (Doull et al., 1980). In addition, vasodilatation and other effects on the cardiovascular system have been noted. Standard blood tests performed in support of a toxicology study will demonstrate a change in hematologic and biochemical parameters.

Changes in the color of the blood, in the carboxyhemoglobin or methemoglobin formation, which leads to anemic hypoxia, are indicative of conditions which can result in neurobehavioral changes. Measurement of cholinesterase activity in peripheral blood may be a key in evaluating convulsions, weakness, paralysis, and other signs produced by cholinesterase inhibitors such as the organophosphate insecticides. CNS stimulation produced by mercury salts may be accompanied by kidney changes detectable by measurement of blood urea nitrogen and acid/base balance. Salicylate toxicosis can produce convulsions and coma along with changes in prothrombin time and a disturbance in the acid-base balance. Logical pursuit of these findings will guide the investigator to conduct additional studies at lower doses which are designed specifically to detect effects on the CNS.
Not all studies need be performed at lower doses, but care should be taken to avoid severe toxicoses that will obscure more subtle CNS involvement. A dose-associated response is as important in neurobehavioral studies as it is in traditional toxicology.

F. DATA EVALUATION

There are several examples in the neurotoxicology literature that demonstrate the need for incorporating animal observations with a review of study data provided by supporting scientists. Unfortunately, both regulatory and industrial scientists often maintain a dependence upon statistics and the importance of the "P" value in toxicity evaluations. Emphasis is placed on the value of data that can be readily presented in some numerical relationship. The selection of study endpoints which offer ease of collection, suitability for standardization, and subsequent mathematical manipulation is not the point in question. In the course of a classical toxicologic study, a considerable bulk of the data that is of interest for evaluating CNS or behavioral effects are descriptive in nature. Behavioral data may be less convincing and more difficult to replicate than the tumor data which can be preserved as tissue specimens. In addition, behavioral data are frequently presented in the narrative portion of a report and may appear without a meaningful indication of their significance. The more specific studies needed to provide the necessary numerical data to document behavioral changes may involve both special equipment and specially trained technicians to follow test protocols that can be replicated. It is also necessary that both the regulating authority and the investigator define an endpoint at which no further testing is required. This is imperative in order to permit investigators to develop sufficient data in support of the recommendations and conclusions from the toxicologic evaluation.

G. SUMMARY

Classical/regulatory toxicology demands a well-defined endpoint when deciding the need for additional testing. There are economic constraints which play a role in defining how much testing is enough.

Test systems should be maximized to produce as much information as possible. Behavioral data collection from routine studies could be improved and several means to do so already exist. Standardization of information required would be useful in assisting the definition of what tests are to be employed as an adjunct to routine testing.
OPEN DISCUSSION

HUTCHINGS: Are there any questions about Dr. Steinberg's talk?

EMERSON: Jim Emerson, International Life Sciences Institute. In one of your tables you noted the MAFF as a Japanese agency, I think MAFF is the United Kingdom or Great Britain, Ministry of Agriculture, Fish and Food.

STEINBERG: Both the UK and the Japanese use the same abbreviation for different government agencies.

EMERSON: Detailed observation of the animals in every step of the system is an integral part of classical toxicology. More information can be gained by detailed observation, the use of appropriately trained observers, and the appropriate interpretation of the data.

STEINBERG: Yes, I think we are in violent agreement.

VORHEES: In Table II.1 which summarized the various requirements for behavioral testing in various countries (e.g., Britain and the EEC), what was meant by a negative sign next to these countries in this table? It is my understanding that the EEC has now incorporated the British language, at least in the area of behavioral teratology, and in that they require behavioral testing for new drugs. Is that correct?

STEINBERG: Yes, the EEC has incorporated the British language. The negative signs indicated whether or not guidelines exist.

VORHEES: You mean in specific tests or do you just mean that it needs to be done?

STEINBERG: Yes, what needs to be done. Several of the guidelines around the world specify that behavioral testing must be included as part of the toxicity evaluation. However, they are not always explicit in describing specifically what type of testing they want or what types of tests to include. This is very much like the in vitro tests. The regulatory agency will suggest different batteries that they would like to see reported, but the specific requirements may not exist. That is why I used the negative signs in this table.
WOOD: I am from the University of Rochester. Dr. Sette, I believe your office has issued guidelines for implementation of behavioral testing under the Toxic Substances Control Act. Could you comment on this?

SETTE: Yes, I had planned to comment on these in my presentation which is coming up next. The guidelines were published in the Federal Register last week (U.S. Environmental Protection Agency, 1985). The Office of Toxic Substances at EPA does have an explicit set of eight neurotoxicology and behavioral guidelines for the conduct of testing. What they do not have is a routine requirement for the use of that testing. So the guidelines are used on a case-by-case basis under review processes in the Agency.

HATTAN: Dave Hattan from FDA. In reviewing data produced by either the original sponsor of a compound or by contract laboratories, most of the time the information provided to us is not of sufficient depth or clarity or, as you pointed out, with insufficient quantitation. It is frequently difficult to determine whether or not something observed in the data is an actual compound-mediated effect.

From your experience, do you think that there is a way that regulatory agencies could get more information and therefore, a better indication of an actual or a subtle behavioral effect by more carefully working with the existing classical toxicological techniques and feeding studies? There is obviously a need for additional testing (e.g., to study mechanisms and things of that nature), but I wonder if there is a considerable amount of useful information that is being overlooked or, as you indicated, not being recorded. If this is true, then what might we, as regulators, do to enhance the quality of the information available?

STEINBERG: To answer your question, I am going to give you my personal opinion with which others may disagree. Part of the problem is the direction in which regulatory toxicology has been taken. For the most part, regulatory toxicology attempts to obtain data that will explain the way one may or may not use a compound. In recent years, there have probably been more meetings on different techniques for assessing hazard and assessing mathematical models for determining how one might use a chemical that may be a carcinogen or may cause a given effect. In addition, many of the agencies require the report to be written in such a way that the report may be subdivided and sent to the various disciplines for evaluation.
In this way, the process itself works against having an overview of all of the data or for individual elements of the data to be compared with one another. This is a major concern that I presented in my paper. For example, the pathologist may observe something in the histopathology, which when fed back into the system, would indicate that there is something happening in the CNS. This is not to say that petitioners do not have a comprehensive internal review of all the data before the report is presented to an agency, but the point is that the overall approach appears to lend itself to the type of review that results in the subdivision of the various scientific disciplines that make up these studies.

Several countries do a better job than others in the way they want the report presented or how they want the data reported. Some countries require that the data be written as "a story". I have heard clients ask, "please tell us the story that the data present." Conceivably, that should permit these kinds of behavioral data to come out. What these clients are looking for is an account of the comparison for the clinical pathology with the anatomic pathology and the tissue masses that were observed while the animal was on study. The whole system has been one of philosophically doing a lump and bump count and not directed the investigator to make better use of these clinical observations which are sometimes available in great abundance.

REITER: From your presentation, I sense that you believe that what currently comprises a functional observational battery in acute and subchronic testing is probably adequate and that some laboratories perform and report these observations better than others.

My concern with the current approach is that these observations are not performed in any systematic way and that there is no attempt to quantitate the results. An advantage of EPA's test guideline for a functional observational battery is that there is an attempt to impose structure and quantitation. This should help to eliminate some of interlaboratory difference and provide some continuity between laboratories.

Another problem with the current approach is that behavioral observations usually take place in the context of lethality studies where animals often show signs such as ataxia, convulsions, or coma. Because of the experimental context in which these data are collected these signs are frequently not given proper consideration. For example, Lucel-7 is a plastic foaming catalyst which was reported in acute studies to produce ataxia, tremors, and convulsions. Despite these findings, this compound was used commercially and resulted in several cases of encephalopathy and peripheral neuropathy.
This outbreak was reported by the NIOSH group and subsequently documented in the rodent by Spencer and colleagues at Albert Einstein. Perhaps if a more systematic functional observational battery had been employed during original testing, the neurotoxicity of this chemical would have been documented and the human neurotoxic outbreak averted.

STEINBERG: Yes, except that I do not think that they are adequate. The adequacy of these data depends upon the central lesion, and there are several materials (e.g., the halothanes), which unless you included some special testing (i.e., an exposure chamber at a high concentration) could not be detected though observation alone. In terms of the dose and where one would see behavioral changes, the same thing is true of those behavioral changes caused by decreased cholinesterase.

However, for the majority of compounds you are right. If we could quantitate what we are observing in those areas, we would probably have a better way of using the data.

EMERSON: Emerson from ILSI. The basic problem is that even though we frequently believe that we consider all of these things, there has been a failure by many investigators, both contract laboratories and industry, to integrate and correlate the clinical observations with the clinical chemistry and gross and microscopic pathology. In addition, however, I believe that if we take two technicians and ask them to make observations on an animal, one will find that there is no change, the other will find a change. Therein lies another basic problem.

The same thing is true if you look at a database. Some will say that there is nothing wrong with clinical chemistry when, in fact, there is something wrong. If somebody has taken the time to correlate the gross observations, clinical signs, with the chemistry, and the alterations at the tissue level, then you can make a better assessment. Frequently, I think that we are faced with data that are only mediocre in quality because there has not been enough time taken to make that correlation and integration of the facts that are known at that time the report is prepared.

REITER: We basically agree, however, I would like to emphasize that just because the data being reported are "behavioral" does not imply that the inter-observer reliability is any different than other forms of scientific data requiring observation on the part of the investigator. I would cite,
for example, that the pathologists asked to review slides from the ED01 study did not always reach the same conclusions. This illustrates that the problem of inter-observer reliability is not unique to behavior.

EMERSON: I totally agree, this is true with every study parameter.

MAILMAN: I would like to add that observation itself can be markedly improved by some new technology. Very small, hand-held, digital data collectors and microcomputers are now available for use in the laboratory. They can be programmed to prompt observers to answer questions which both time-locks the observations and also allows some degree of quantification. This method has been applied in a variety of forms and is ideal for various toxicological studies. It eliminates many of the observer comparison problems because of the fact that observations are both time-locked and independent. This is something that should be explored further in these observational studies.

STEINBERG: Depending on the degree of sophistication in the laboratory, prompting methods have been used for many years. Whether one uses a digital hand-held computer or you give the technician a piece of paper with a list of things to record. The coding system which I described in my presentation is an example of a prompting technique which is similar to the coding system currently used in many laboratories.

One problem relates to the issue of technician training. This is illustrated by the example I gave regarding sensitivity to the touch. The animal clearly responds differently to an untrained technician who grabs the animal with a death-hold because he fears a rat bite, as opposed to the technician with experience who can interface without upsetting the animal. This always comes back to the fundamental question of how well trained is the technician? Another problem that I find is the so-called "on-the-job training factor". In this situation, the person that does the training always appears to hold back one thing, whether it is done intentionally or not. There is a need for a formalized training program wherever you have an experimental protocol with one of these checklists that clearly indicates to the technician what is supposed to be observed and how it is supposed to be performed. However, these checklists have not always paid off and the problem still comes back to the question that once you have the data, what do you do with it?
DONZANTI: My name is Bruce Donzanti from the U.S. Army Chemical Defense Program. I agree that training the observer is important and can be a problem. In addition to using scoring sheets we have used computerized machines where you can lock in the information. The biggest problem that we run into is the question of what do you do with the information when you have it? The underlying problem is with the statistical analysis of this type of data. We can never seem to find a statistical model that fits observational data and it ends up being written as a description of what happened. Some form of data analysis is needed to determine how often it occurs or if it is significant. Determining the p value of something would be useful. In your personal experience have you found a statistical method that can be used for this type of observational data?

STEINBERG: While I am not a statistician, you can apply nonparametric statistics to set up a plus/minus coding system to some degree. Then you can come up with a p value based on that. Unless you had some way to really codify this well, you may not be much further along when you finished. The problem is to interject it into the report system so that it is part and has equal weight with some of the other things that are reported.

DONZANTI: Even if you were doing an initial-type of screen with six animals per group, the nonparametric tests can be misleading. These tests are not that good for this type of data. I know that it has been addressed several times in the literature but I have never seen anything that indicates whether or not the data is significant.

STEINBERG: I think we should leave that question to the statisticians and the behavioralists rather than myself.

SPENCER: Peter Spencer from Albert Einstein. There are important lessons to learn from human medicine. The clinician, in approaching a patient with neurological disease, will observe the behavior of the individual (rather crudely by Larry Reiter's standards) and also conduct a variety of functional battery tests to elicit a number of signs. On the basis of this neurological examination, and if the individual should come to autopsy, with a neuropathologic evaluation, one can learn an enormous amount about the neurological/neuropathological correlation. The neurologist is often able to predict with great accuracy the type and severity of the lesion that might be found when eventually that individual comes to autopsy.
What I am suggesting here is the correlation between use of a behavioral battery, observation of functional signs in pinpointing, and the type and severity of the pathological lesions. On observing the pathological lesion one is then able to see whether or not the particular disorder is reversible or irreversible. That might provide information on the severity of the toxicity in question.

HARRIS: Dr. Harris at the EPA. I would like to continue with that type of analysis vis-a-vis the medical conditions and the problems, not only in psychiatry and neurology, but their relationship with testing.

If you have an obvious morphological lesion in the nervous system, a correlation can often be made with behavior. In psychiatry and neurology, one finds that more often problems potentially involve biochemical abnormalities but few situations exist for correlating behavior. That is where the sensitivity of the method becomes important. The question then is, are the behavioral methods more sensitive than the biochemical methods? There is a problem inherent in this because we do not know what the biochemical abnormalities are to know whether there is a correlation with behavior. What you end up observing is a continuum of behavior.

For example in psychiatry and neurology, where you may evaluate a hyperactive child, there are degrees of hyperactivity that range from minimal (i.e., a mother reports -- "I think my child is hyperactive"), through a continuum to what we call minimal brain dysfunction, where the condition is diagnosed as hyperactivity with many other neurological problems. Ironically, and this is the case in other diseases as well, the diagnosis is confirmed after a course of treatment. For example, if amphetamines improve the condition, diagnosis is minimal brain dysfunction.

In the laboratory animal, where a technician observes a condition which is measurable (e.g., an increase in motor activity), how does that relate to the continuum and what do you correlate it with? Should we treat the animals that seem hyperactive to see if they will become less hyperactive, and determine whether there is a treatment-related disorder? This sounds ridiculous, so the problem in behavioral testing and also in psychiatry today is that there are conditions which are continuums of behavior, from a good imagination to hallucinations, from a hyperactive child to minimal brain dysfunction, from reactive to endogenous depression, from elation to mania. The distinction between the pathological disease or condition from a normal excess of emotional behavior is difficult to diagnose in humans and is uninterpretable in the laboratory animal.
It is very difficult, unless you have, as in neurology, a pathological condition where there is a known morphological lesion for which you can identify the kind of behavior caused from that lesion. But, if it is potentially a biochemical problem, which we think much of psychiatric illnesses are, it is very difficult to know what kind of behavior can result.

STEINBERG: I agree with you. That is why we in toxicology determine a dose response, in the hope that the high dose will provide clearer definition of the effect. While many investigators do attempt to include special biochemistry studies, one should be attentive to the methods used. It is important to consider what is not always taken into account for some of these data (i.e., variability in the chemical method itself, variability in the collection of the sample, and variability in the animal). In the end, one hopes that the variabilities will in some way cancel one another when the statistical analysis of data have standard deviations that run out of line. While biochemistry and clinical chemistry are used extensively, in many instances, it has not been proven to be the panacea that we hoped it would be in terms of really evaluating data.

HUTCHINGS: Let this be our last question.

GLOCKLIN: My name is Vera Glocklin at the Food and Drug Administration in human drug assessments.

I would like to comment about what Dr. Steinberg said regarding interdisciplinary communication. With human drugs, neurotoxicity or neuropharmacology is part of typical drug screening. In neuropharmacology there is a battery of acute tests that produce data which is quantitatively expressed as an ED$_{50}$ value. The LD$_{50}$ is often an extension of the ED$_{50}$ for certain of these effects.

Some years ago we met with the PMA to try to develop new toxicology guidelines. We discussed how to determine whether an acute test effect gets worse or whether accommodation or tolerance develops with chronic treatment. At that time, it was suggested by the pharmacologists that the toxicologists could, for example, use animals from a chronic study to perform a rotorod test, whereas the toxicologists suggested that if the pharmacologists wanted to know the effect of chronic treatment on the ED$_{50}$ value, that they could have their own animals separate from those in the toxicology study. The problem is that the science of risk assessment was being treated as two completely separate disciplines; toxicology at that time having a strong pathology background and pharmacology with a strong functional background.
There has always been a need for improved communication. I have seen drug applications with neuropharmacology data showing the chemical to be a potent sedative but with no indication from the toxicology data that the animals were sedated. This causes the regulatory scientist to wonder whether it was the same compound. Obviously, the toxicology investigators were not looking appropriately at all effects.
LITERATURE CITED


III. COMPLEXITY OF NEUROTOXICOLOGICAL ASSESSMENT

William F. Sette, Ph.D.
Alexandria, Virginia

ABSTRACT

One of the central objections to routine behavioral testing concerns the difficulty in interpreting behavioral effects. This paper discusses the assessment of neurotoxic hazard using data from both conventional toxicity studies and behavioral studies. It describes the breadth and complex expression of neurotoxic effects, the inference of neurotoxicity from behavioral and other data, and the validity and interpretation of behavioral effects. The basic questions asked by regulatory petition reviewers include: (1) can an effect be identified; (2) can the effect be considered to be a function of exposure; (3) can the effect be considered to be a direct effect on the nervous system; and (4) can the effect be considered to be an adverse effect? Approaches and options for improving the interpretability of current testing strategies are described. These include improvements and additions to existing studies, explicit testing, and development of generic testing requirements for specific chemical classes.
INTRODUCTION

Federal agencies and concerned neurotoxicologists have discussed the need for explicit behavioral and neurological testing of environmental chemicals for at least a decade (Anger and Johnson, 1985; Schaeppi and Hess, 1984; Tilson and Mitchell, 1983, 1984; Weiss, 1983). Numerous ad hoc groups have met to address issues in neurotoxicity for agencies such as the Environmental Protection Agency (EPA) (Weiss and Laties, 1979), Consumer Product Safety Commission (CPSC) (Laties et al., 1977), and the Food and Drug Administration (FDA) (Wurtman and Maher, 1984). Other neuroscientists have focused on the need for testing food additives (Hattan et al., 1983; Sobotka, 1986; Wurtman and Maher, 1984) and behavioral teratology (National Research Council, 1975; U.S. Environmental Protection Agency, 1984; Zbinden, 1981). The need for widespread and explicit routine testing has been generally endorsed within the scientific community, however, there is less agreement on what tests are needed (e.g., Schaeppi and Hess, 1984) or how specific testing recommendations should be pursued (e.g., Weiss and Laties, 1979). Despite these efforts, there has been little change in the testing required by federal agencies to evaluate neurotoxic effects.

One major objection raised by scientists (Hattan et al., 1983; Zbinden, 1981) to routine behavioral and neurological testing is that the biological and toxicological significance of the data are difficult to interpret. However, this is not unique to neurotoxicology. Miller and Skinner (1984) noted that "We still have difficulty in interpreting physiologic and toxicologic data, particularly that obtained from animal experiments." An example of this problem was described by Hattan et al. (1983) who cited difficulties in interpreting a behavioral study of food additives by Brunner et al. (1979). In that study, the authors found that the effects of aspartame (containing up to 50% phenylalanine) were essentially the same as those produced by phenylalanine. Hattan et al. (1983) suggested that these tests "should be able to discriminate between the toxic effects observed following administration of normal food constituents (such as phenylalanine) and the food additive (such as aspartame)." They concluded that "it is not clear that these testing paradigms are sufficiently developed at this time to warrant their routine application."

Zbinden (1981) critically reviewed methods used in behavioral teratology. He described the subtle, confusing, and inconsistent nature of many behavioral effects reported in the literature and the large number of variables that may influence these effects as a source of confusion in the interpretation of study findings. Like Hattan et al. (1983), he also concluded that "it serves no useful purpose to legislate the general adoption of behavioral test batteries whose usefulness has not been validated and which may become obsolete in a few years."
The purpose of this review will be to describe a framework for the assessment of neurobehavioral hazards and the interpretation of effects for human risk assessment. This review will focus on four questions central to risk assessment: (1) whether an effect can be identified; (2) whether the effect can be considered to be a function of exposure; (3) whether the effect can be considered to be a direct effect on the nervous system; and (4) whether the effect can be considered to be adverse. For the regulatory scientist, both the problems related to neurotoxicological assessments and the limitations in the interpretation of animal studies can be complex. The complexity of these problems, however, need not prohibit the regulatory agencies from requiring routine neurotoxicologic testing to protect the public health.

A. FEDERAL MANDATES FOR NEUROTOXICITY TESTING AND RISK ASSESSMENT

The process of evaluating environmental chemicals depends on both the enabling statute and the policy set forth by the agency responsible for carrying out risk assessments. Only EPA and FDA routinely require testing by a manufacturer prior to marketing; FDA for new drugs and food additives and the EPA for new pesticides. For these categories of chemicals, the responsibility for testing and demonstration of safety rests with the manufacturer (Fisher, 1983; Sette and Levine, 1986). The Office of Pesticides Programs (OPP) of the EPA has explicitly rejected routine specific neurotoxicity testing for all pesticides and relies on conventional acute and subchronic studies to detect "potential for persistent or permanent neuropathy" as the criterion for explicit testing (U.S. Environmental Protection Agency, 1984). However, the EPA does have two guidelines for the evaluation of organophosphorus esters that may induce paralysis. These tests are routinely required for pesticide registration (U.S. Environmental Protection Agency, 1978).

EPA and CPSC may require testing of industrial chemicals or consumer products, respectively, but the justification for testing lies with the government. In practice, little testing has actually been required by the CPSC (Fisher, 1983). The Office of Toxic Substances at EPA has prepared eight guidelines for studies of neurotoxic effects of industrial chemicals (U.S. Environmental Protection Agency, 1985), but explicit testing has only been requested, on a case by case basis, for a small number (roughly 12) of chemicals during the last 5 years.

Like the pesticide program at EPA, the FDA also requires neurotoxicity testing of food additives only if neurologic effects are seen in the conventional toxicological tests. But the Center for Food Safety and Applied Nutrition at FDA has been active in developing test methods and sponsoring symposia concerning neurotoxicity (Gryder and Frankos, 1980;
Vorhees et al., 1979). Hattan et al. (1983) have reviewed the legal basis for test requirements and described the process for regulating food additives. Since 1958, section 409 of the Food, Drug, and Cosmetic Act requires manufacturers to demonstrate "safety" prior to the marketing of any new food additive (Food and Drug Administration, 1980). Under the Code of Federal Regulations [21 CFR 170], the demonstration of "safety" requires that a manufacturer provide adequate data to assure a "reasonable certainty of no harm" from the proposed use of the food additive. The FDA is guided in their testing requirements by "the principles and procedures for establishing the safety of food additives stated in current publications of the National Academy of Sciences, National Research Council." In 1982, FDA explicitly defined the criteria to determine both the nature and extent of testing necessary for evaluating food and color additives (Food and Drug Administration, 1982).

The extent of testing required by the FDA for a food additive is a function of a toxicity assessment that is based on both the chemical structure and the estimated daily intake. There are three structural categories based on low, medium or unknown, and high probable toxicity. Each category is divided into three "concern levels" based on exposure. The extent of required testing is directly proportional to the level of exposure and probable toxicity of the food additive to be tested. Contained within each level of concern are provisions that provide for short-term and chronic carcinogenicity studies, reproductive and teratological studies, and subacute (28 day) to chronic feeding studies. Testing for neurotoxicity is based on analysis of data from these required studies. The FDA guidelines state that, "Compounds which induce neurotoxicological signs, symptoms, or effects in any of the required toxicological tests may require special testing. The type of test(s) required will depend on the review of data." Therefore, there are no predefined chemical classes of neurotoxicants that would trigger explicit testing as defined by these existing regulations.

In 1984, the National Research Council (NRC) completed a review of toxicity testing needs for all types of environmental chemicals (National Research Council, 1984). Based on a detailed evaluation of a sample of 8,627 food additives, they found that only 5% have data adequate for a complete health hazard assessment and 46% have no data at all (Figure III.1). The NRC review panel noted that part of this large gap may be due to the recent removal of many food additives from the Generally Recognized as Safe (GRAS) list (Food and Drug Administration, 1970). In their recommendations, the panel listed neurobehavioral toxicity among the top four testing priorities for food additives.

The risk assessment process involves both the qualitative and quantitative characterization of adverse health effects of chemicals. The evaluation and selection of regulatory options is defined as risk management (National Research Council, 1983).
Figure III.1. Food additives data (modified from National Research Council, 1984).
For noncarcinogenic food additives, the Center for Food Safety and Applied Nutrition applies safety factors to define "acceptable daily intakes" (ADI). This process involves both risk assessment, in defining no-effect levels and risk management, in selecting a safety factor. Typically, a no-effect level that has been derived from the animal data is divided by a safety factor to estimate an ADI level for man. This value is then compared to the estimated daily intake (EDI) to determine if the anticipated exposure to the food additive is less than the ADI, and thus, safe (Food and Drug Administration, 1982). The selection of a safety factor has historically been based on powers of ten, depending on the type of data available (Lehman and Fitzhugh, 1954). If chronic animal data are used, a safety factor of 100 is applied. This will allow a factor of 10 for species extrapolation and a factor of 10 for variation in population sensitivity.

In summary, manufacturers of food additives must provide sufficient data to adequately assure a "reasonable certainty of no harm." In addition, the extent of testing required for most adverse health effects is based on estimates of toxicity and exposure. No explicit neurotoxicity or behavioral studies are required unless such effects are observed in the required studies. The definition of ADIs is based on safety factors that are applied to no-effect levels from preclinical studies with laboratory animals.

B. HAZARD IDENTIFICATION

The first step in a risk assessment is the qualitative identification of hazards. This section will define the scope of neurotoxicity and behavior, illustrate some of the complexities of the expression of neurotoxic and behavioral effects, and derive some general principles for hazard assessment.

1. Definitions and distinctions between neurotoxicity and behavioral toxicity

Neurotoxicity is defined as any adverse effect on the nervous system. The nervous system receives and processes environmental input and regulates an organism's bodily functions, consciousness, and feelings. The scope of potential neurotoxic effects is very broad, ranging from fleeting euphoriant effects to death by respiratory depression. While neurotoxic effects in laboratory animals, as well as in man, may first be seen as changes in behavior, assessment of neurotoxicity involves many levels of analysis. Complete understanding of a neurotoxicant will include the elucidation of its action on the morphology, physiology, and biochemistry of the nervous system.
Behavioral toxicity may be defined as any adverse effect on behavior. Behavior represents the integrated output of all the organism's physiological systems including, but not limited to, the nervous system. The study of behavior is the most general level of analysis of the individual organism which reflects the organism's experience and environment, as well as its internal "milieu". The effects of toxicants on behavior are an apical indicator of toxicity in an organism (Butcher, 1976) and not all behavioral effects represent neurotoxicity. Our understanding of the action of behavioral toxicants will ultimately rely upon an understanding of the physiology, morphology, biochemistry, and the internal "milieu" of the affected organism (Bondy, 1985).

2. Types and properties of hazards

Hazard assessment involves the description of the qualitative, quantitative, and temporal characteristics of effects. Over 800 chemicals have been identified as toxicants that induce behavioral effects (Anger and Johnson, 1985). The extraordinary range of symptoms reported by humans inadvertently exposed to these chemicals encompasses sensory, motor, affective, cognitive, and physiological effects. The breadth of adverse effects poses the greatest problem in the identification of potential hazards. No conceivable test battery in animals could identify all of the potential human neurotoxic hazards. Exposure to an individual chemical agent may induce one pronounced effect or a spectrum of effects on a variety of organ systems. For example, those pesticides known to inhibit acetylcholinesterase activity have a specific site of action while other toxicants, like methyl mercury, have a broad range of effects on both the nervous system and other organ systems (Bondy, 1985). Many neurotoxicants, like nicotine, may cause an increase in the activity of affected neurons at low levels and a decrease in activity at higher levels (Goodman and Gilman, 1975). In addition, the effects seen after chronic exposure may be quite different from those seen after acute exposure. A survey of neurotoxicants found that 10 out of 23 chemicals studied had qualitatively different effects following acute and chronic exposure (Bass et al., 1985a).

Neurotoxic effects may range in severity and duration from the subtle transient disruption of vigilance performance caused by carbon monoxide (Latties and Merigan, 1979) to the fatal consequences of exposure to cyanide (Goodman and Gilman, 1975). In addition, these effects may appear rapidly following exposure as with cyanide or be delayed by weeks as in the case of organophosphorus esters like tri-ortho cresyl phosphate (Morgan, 1982).

A neurotoxic chemical may exert effects on sensitive populations that are quantitatively and qualitatively unique, as has been demonstrated in children from Iraq whose mothers were poisoned with methyl mercury (Marsh et al., 1980), in the
response of only one child to food color challenge (Weiss et al., 1980), or the response of phenylketonurics to aspartame (Rosenberg and Scriner, 1974).

Functional damage may occur in the absence of apparent morphological correlates (e.g., central nervous system (CNS) depression); whereas massive structural damage may lead to only transient changes in function (Lashley, 1929; Tilson and Mitchell, 1984). In addition, tolerance, which can be defined as a reduction in response following repeated exposure, may mask neurotoxic effects that ultimately become manifest as irreversible changes (Bignami et al., 1975; Jaffe, 1975). The tardive dyskinesias produced by many major tranquilizers are a tragic illustration that an apparent adaptation (to the acute dyskinetic effects) may in the long-term lead to irreversible effects (tardive dyskinesias).

Several general principles of neurotoxic hazard assessment are summarized in Table III.1. Testing should evaluate both the effects of acute and repeated exposures. The range of doses used should be broad enough to cover the spectrum of effects. The study should be long enough to detect delayed responses and the consequences of tolerance. In utero and special population effects may require additional testing. Additional consideration should provide testing for functional, pathological, and biochemical measurements when necessary and the spectrum of functions to be assessed should be as broad as possible.

Table III.1. Hazard Assessment Principles

- Acute and repeated exposure
- Functional and pathological effects
- Prenatal and adult exposures
- Dose-response and dose-effect
- Delayed responses
C. HAZARD ASSESSMENT

The principles stated above offer an ideal set of testing needs to fully characterize neurotoxic and behavioral hazards. In the regulatory process, however, decisions may need to be based on more limited data in order to estimate the toxic potency of a chemical based on chemical structure, to determine whether the effects seen in conventional studies should trigger explicit studies, as well as the quantitative characterization of hazards.

1. Relationships based on physical or chemical properties and chemical classification

Evidence of hazard may be inferred from physical or chemical data and from relationships based on chemical classification. Zbinden (1973) has listed 14 physical or chemical properties that may suggest routes of exposure and the extent of absorption, distribution, metabolism, and excretion for various chemicals. The lipid solubility of a compound is often considered as an index of both the ease of penetration into neurons and an indication of bioaccumulation in the lipid-rich brain. These properties are generally associated with access to the nervous system and anesthetic potency. Their consideration can facilitate general assertions concerning potential neurotoxic hazard. Two kinetic criteria for evaluation of food additives has been proposed by Wurtman and Maher (1984). They state that either the ratio or concentration of a chemical must be increased in the plasma and that the chemical must cross the blood-brain barrier through lipid solubility or as a ligand in the blood-brain transport systems. Klaassen (1980) describes the basic relationship between the speed and site of gastrointestinal absorption and lipid solubility.

Among the more than 800 chemicals listed by Anger and Johnson (1985) are classes of chemicals that are known to be neurotoxicants. For example, over 200 organophosphorus esters have been studied in hens for their potential to induce a central-peripheral axonopathy (paralysis and spasticity) in man (Johnson, 1975). Neurotoxicity testing should be required on the basis of chemical classification in a similar manner to the present testing requirements for carcinogenic and teratologic effects. The Office of Pesticide Programs of EPA requires acute delayed neurotoxicity testing for organophosphorus esters that may be present in pesticides (U.S. Environmental Protection Agency, 1978). The impetus for testing organophosphate compounds occurred in the United States 55 years ago when a ginger extract was intentionally contaminated with an orthocresol mixture that was added to help circumvent federal prohibition regulations (Morgan, 1982). This incident led to the poisoning of tens of thousands of people in the mid-southern states. Ironically, conventional studies like those in use today and neurobehavioral
screening studies would not likely identify these axonopathic effects, since rats are relatively insensitive to organo-phosphorus compounds (Johnson, 1975; Veronesi, 1984).

The major uncertainty associated with identifying concerns based on structural analogs is the extent to which an analog is appropriate. For example, Krasavage et al. (1982) have identified gamma-diketones (e.g., hexane and methyl n-butyl ketone) as a class of chemicals that induce a peripheral paralysis; however, they also found that 2,4 pentanedione, a beta-diketone, induced cerebellar lesions. At this time, the extent of concern and the ability to predict the level of concern for beta-diketones is unclear because much of the database for these compounds is incomplete. Nevertheless, the use of chemical properties and structure activity relationships should be a critical element of hazard assessment.

2. Direct evidence and validity

Direct evidence of neurotoxic hazard in laboratory animals may be obtained either from traditional toxicologic studies (see Chapter II) or from explicit neurobehavioral studies (see Chapter V). The hazard assessment process should include an evaluation of: (1) whether the incidence and severity of the adverse health effects are a function of exposure; (2) whether the effects are referable to the nervous system; (3) whether the effects are adverse; and (4) whether these effects are predictive of human effects. Validity refers to the extent to which a test measures what it is intended to measure (Anastasi, 1982). By distinguishing between the content, predictive, construct, and concurrent validity of neurotoxicity tests, we may be better able to address our concerns for the relevance of data used in hazard assessment (Table III.2).

a. Predictive validity

A fundamental assumption of modern toxicology is that the effects seen in animals are predictive of the effects that may be seen in humans. This can be called predictive validity. A cursory review of neurotoxicants (e.g., solvents, metals, pesticides) in Spencer and Schaumberg's text (1983) reveals that great similarity can be found between effects seen in experiments with laboratory animals and studies of accidental or occupational exposure in humans. This may be due to the fact that historically, as illustrated by the ginger jake episode, the effects of neurotoxicants have been discovered first in man and then followed by the development of animal models. In the most general terms, traditional or explicit toxicity studies in animals may be regarded as predictive of effects in humans.
Table III.2. Concerns for the Relevance of Data in Hazard Assessment

<table>
<thead>
<tr>
<th>Content Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Is there an exposure - related effect?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictive Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Does the effect predict activity in man?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Construct Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Is the effect neurotoxic?</td>
</tr>
<tr>
<td>• Is the effect adverse?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concurrent Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Is the effect correlated with other indices of toxicity?</td>
</tr>
</tbody>
</table>

b. **Content validity**

Evaluation of both dose-response and dose-effect relationships is essential to any hazard assessment. The public safety goal of multiple dose studies in laboratory animals is to determine a no-effect level (NOEL). The scientific objective of preclinical testing is to focus on the range of the dose-response and the dose-effect relationships. Experiments with laboratory animals that demonstrate changes in the incidence or severity of neurotoxic effects at two or more dose levels provide the most powerful evidence of toxicant action. **Content validity** can be demonstrated by a statistically significant relationship between the independent variable (the level of exposure) and the dependent variable (the incidence of observed effect). If only the highest dose causes an effect, then the effect is said to be only associated with, rather than a function of, exposure. The absence of an observable effect at lower doses may be regarded as evidence of safety but adds little or nothing to our knowledge of the relation between the chemical and the exposure. In contrast, information derived from more complete dose-response curves not only provides evidence of a functional relationship, but also allows one to consider the slope of the function in predicting hazards at lower exposure levels. NOELs are both less informative and less useful than a dose-response curve.
Failure to obtain dose-response data (Zbinden, 1981) is one of the most serious limitations of behavioral teratology studies but is also a common problem of oncogenicity studies as well. The greatest problem with data from laboratory studies that indicate an effect at only one dose level may not be whether the effect is statistically significant, but rather what the biological and toxicological significance of the data may be.

c. Construct validity: neurotoxicity

Construct validity refers to how well a test measures a construct. Both neurotoxicity and adverse toxic or behavioral effects may be regarded as hypothetical constructs. The evaluation of data from conventional toxicity studies can provide useful information about changes in behavior and the nervous system such as paralysis, convulsions, and depression if observations are systematic and fully described. For example, in acute lethality studies, the time of death can be regarded as an indication of neurotoxicity if it occurs rapidly (i.e., within 24 hours of exposure). In general, the broad classes of non-specific CNS depressants are known to kill quickly. Systematic observation may also provide reasonable quantitative data on neurotoxicity. Bass et al. (1985b) surveyed the literature of roughly two dozen chemicals with Threshold Limit Values (TLV) that were based either entirely or in part on narcosis (Anger, 1984). While the explicit behavioral testing of these chemicals was limited, and such tests provided increased sensitivity, systematic observation provided reasonable sensitivity for measuring narcosis.

d. Concurrent validity

Explicit neurotoxicity studies or conventional toxicity studies that include neuropathology or biochemistry can provide concurrent validation of functional effects to differentiate neurotoxic effects from other phenomena. These correlations are readily established for agents that produce pronounced pathology and behavioral effects (e.g., axonopathies and paralysis). Unfortunately, not all agents produce simple effects and as Bondy (1985) noted, "Correlation of chemical and behavioral data is not readily achieved." He noted as an example that changes in any of several neurotransmitter systems may underlie hyperactivity.

e. Construct validity: adverse effects

The extent to which changes in behavior or other neurologic measures are considered adverse can pose the most difficult regulatory and scientific questions. While
ascribing effects to nervous system disruption involves a construct, what is considered adverse clearly involves judgments based on social values. This is the major limitation of many behavioral tests in both the adult and teratologic testing (Evans and Weiss, 1978; Mitchell and Tilson, 1982). Anger (1984) may have referred to this when he recommended the use of tests that have some "face-validity" for nonpsychologists, that is, tests should assess some aspect of a toxicological construct. Frankos (1985) has classified developmental endpoints for FDA into two categories that are based on severity and reversibility. Both the testing requirements and safety factors used were based on this classification of effects. In a recent commentary, Dyer (1984) proposed a more elaborate, four level matrix, for classifying endpoints in animals or humans. His endpoints were defined in terms of their similarity to known human health effects. A study endpoint in laboratory animals may be the same as, apparently the same as, a model of, or unrelated to a human (toxic) endpoint whereas endpoints in man are classified on the basis of their relation to human health toxicity as well. This system of classifying neurological endpoints, as developed by Dyer, is somewhat complicated and arbitrary; nevertheless, it shows how we can classify neurological effects in animals based on the "adverse" nature and a relationship to "adverse" effects in humans.

D. SUMMARY

To consider a behavioral or neurological effect to be adverse requires a comprehensive description of all the available data, consideration of the severity and the duration of the effect, the homology between the measured effect and a toxic effect, and the social values of those making the judgment.

Hattan et al. (1983) and Zbinden (1981) have raised concerns about the use of specific types of studies. Hattan et al. (1983) were concerned with a failure of tests to discriminate between the effects of aspartame and the effects of phenylalanine, while Zbinden (1981) expressed concern for differentiating maternal, fetal, and "true behavioral effects". Their concerns focused on whether the endpoints measured were considered adverse or unique; however, these concerns do not relate to either the content or predictive validity of the tests that were used in the studies mentioned, and therefore, it does not seem reasonable to conclude that the tests need further development or that behavioral batteries are not useful. Behavioral effects are not always neurotoxic or adverse, and the determination that behavioral effects are minor is not an indictment of a test method or the usefulness of the test findings in the evaluation of a chemical. There is a need for both researchers and regulators to be attentive to both the nature of what a test measures and the interpretation of the test findings.
Routine testing for teratogenicity began after the thalidomide tragedy (Frankos, 1985) while the inexpensive screening test developed by Ames et al. (1975) initiated widespread application of mutagenicity testing. The neurosciences have already had a food additive disaster with the ginger jake episode of 1930. It is suggested that small increments to routine testing coupled with concern directed by structure can only improve our ability to make decisions and identify neurotoxic chemicals.
OPEN DISCUSSION

HUTCHINGS: Are there any questions or comments from the floor for Dr. Sette?

O'DONOGHUE: I am with Eastman Kodak. I agree that structural activity relationships are very important. However, these relationships are often very complex and when looking at a particular chemical, there is often not enough information to decide what part of the structure is important for its activity.

Bill Sette mentioned 2,4-pentanedione, but I would like to mention the problems with 2,5-hexanenedione. With the 2,5-hexanenedione problem, there is a question of which structure is most important, n-hexane, methyl n-butyl ketone, or 2,5-hexanenedione. Should we be concerned with all alkanes, all beta ketones, or all gamma diketones? There is a need to understand what is the proximate neurotoxin and how the body converts one chemical to another which results in that proximate neurotoxin. Then we can understand the importance of structure.

To be given one chemical, we can not really dissect apart that particular chemical and find out what is important unless we look at a whole series of materials.

SETTE: Yes, I would certainly agree that in cases precisely like the 2,4-pentanedione case, there is one actor that is known. Although I do not recall the details, there may be two or three others which were found to be negative. This example is hardly sufficient for what serious scientists would call structure-activity relationships.

At the same time, let me cite the Lucel 7 example. If we are to say that Lucel 7 teaches us something, one should be on the alert -- if a new chemical has only one small variation from Lucel 7 and is presented for a review. I think it is reasonable for people to expect the reviewer to question whether this chemical will cause the same kind of problem. At a minimum, we should ensure that we have examined this compound for these effects.

SPENCER: While structure-activity relationships are clearly important for a few classes of chemicals, their general applicability is presently fraught with hazard. In fact, if one had looked at the structure of Lucel 7, based on the knowledge of gamma diketone neurotoxicity, one would have concluded that this compound was probably safe, since its five methyl group was blocked.
SETTE: I did not mean to imply any similarity between Lucel 7 and the gamma diketones. I think the tert butyl area is the "hot" part; but again, that is as fuzzy a class relationship as any.

HUTCHINGS: Thank you, Dr. Sette.
LITERATURE CITED


IV. MECHANISMS OF CNS INJURY IN BEHAVIORAL DYSFUNCTIONS

Richard B. Mailman, Ph.D.
Departments of Psychiatry and Pharmacology
Toxicology and Neurobiology Curricula
Biological Sciences Research Center
University of North Carolina School of Medicine
Chapel Hill, North Carolina

ABSTRACT

Advances in the neurosciences have led to an improved understanding of the anatomical, biochemical, and molecular loci involved in injury to, and adaptation of, the central nervous system. Recent research has permitted the elucidation of the mechanisms for some neurotoxics whose actions have been studied for decades, as in the example of the pyrethroid insecticides. In contrast, the mechanism of the neurotoxicity of the organophosphate insecticides and nerve gases has been known for many years, but our understanding of the many resulting sequelae has been markedly increased by recent discoveries. Two examples illustrate the strengths and weaknesses of such methods in predicting neurotoxicity. Studies of the parkinsonian-like toxicity caused by 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) (a by-product of synthesis of illicit opiates) exemplify the best application of current methods in neurotoxicology. It has been shown that the expression of MPTP toxicity requires both metabolism of MPTP to the proximal toxicant 1-methyl-4-phenylpyridinium (MPP⁺) and active uptake into central dopamine neurons. The discovery of binding sites of MPP⁺ in these cells has clarified how dopamine neurons are destroyed, thereby causing neurological signs. This illustrates two key concepts: first, the bioconversion of compounds to proximal toxicants is often ignored, and second, these events are unlikely to be detectable by in vitro studies that focus on a few biochemical endpoints. Another useful example was that of erythrosin (FD&C Red No. 3), in which numerous in vitro studies suggested that this food color was a potential neurotoxicant. However, this was shown to be an artifact of the ability of this color to disrupt biomembranes at high in vitro concentrations, and this idea was supported by negative data from both behavioral and clinical studies. Thus,
the plethora of possible molecular and biochemical targets in the central nervous system (receptors, second messenger events, transmitter-modulator synthesis, storage and release, membrane maintenance, etc.) preclude the likelihood of developing a single test or a battery of neurochemical or biochemical tests that will be able to screen for neurotoxicants randomly or efficiently. Use of in vitro methods is likely to detect both false positives and negatives. While the availability of theoretical or phenomenological data provides the best start to the application of available biochemical and molecular techniques, predictions of neurotoxicity best can come from theoretical comparison of the structure of suspect compounds (and hypothesized metabolites) with known target sites in the CNS. Such a comprehensive structure-activity examination, while apparently an heroic task, can be made quite manageable by the availability of modern computer modeling, storage, and retrieval systems. Coupled with behavioral, pathological, and physiological screening in intact target organisms, it could provide the most efficient manner to predict neurotoxicity.
INTRODUCTION

The molecular and biochemical neurosciences are extremely productive areas of research and it is almost weekly that the advances in molecular biology and the biochemical sciences, coupled with improvements in microelectronics and computer science, lead to the announcement of breakthroughs in understanding problems that have perplexed the biomedical sciences for decades. Whether it is the structure or biology of a new virus, the neuronal architecture of a newly identified chemical messenger, or the identification of the recognition site responsible for a drug's action, these advances have both enriched and complicated our lives as scientists. It is therefore fitting to ask what may be the contribution of these disciplines to the question of the prediction of neurotoxicity.

The development of behavioral toxicology as a sub-discipline has led this field to ask questions about the biochemical and neurochemical mechanisms and sequelae of toxicant exposure. The correlation of behavioral events with alterations in neuronal function is frequently attempted and neurochemical studies provide one method for understanding "target sites" in neurotoxicity. Conversely, the richness of available in vitro methods has led to the suggestion that one or more of these methods may provide a rapid and efficient way to screen for potential neurotoxicants. The aim of this paper is to critically review the validity of both approaches and attempt to offer a strategy likely to be useful in the future. In so doing, specific examples have been selected that seem prime examples of the challenges faced by neurotoxicologists. It is the intent of this discussion to illustrate that these are the rule, rather than the exception, in the study of neurotoxicity.

OVERVIEW OF THE SITES AND MECHANISMS OF NEUROTOXICITY

The immediate issue is the potential role that molecular and biochemical studies may play in predicting neurotoxicity. The complexity underlying this matter may be gleaned in part by understanding the cellular and biochemical loci with which one must deal. Many of the examples chosen will be directly related to dopaminergic transmission in the central nervous system. These examples are similar to dozens of other examples that relate to different aspects of the central nervous system.

A. CHEMICAL MESSENGERS

It is essential for the purposes of this forum to review our present understanding of the nature of neurotransmission. Reference has been made to dopamine, a catecholamine that has been accepted as a neurotransmitter for more than
2 decades. Historically, neurobiologists have used a series of criteria to evaluate whether a compound was accepted as a neurotransmitter. These characteristics included demonstrating the compound and its biosynthetic machinery in neurons and the ability to demonstrate release of the compound upon stimulation. Mechanisms for rapid inactivation of the compound were required (e.g., uptake or hydrolysis) and it was also necessary that postsynaptic events initiated by the transmitter presence be demonstrated. The latter could include both biochemical, cellular, and physiological (or pharmacological) changes. Finally, antagonists of the action of the transmitter candidate, when directly applied to the postsynaptic terminal, should have been able to block the effects of stimulation of the presynaptic nerve terminal. Table IV.1 lists many of the compounds that were generally accepted as neurotransmitters in the mid 1970s.

When Schally and Guillemin shared the Nobel Prize for chemically identifying the hypothalamic physiotrophic agent thyrotropin releasing hormone (TRH), the final concrete evidence was provided that certain peptides had critical roles in the function of the central nervous system (Guillemin, 1978; Schally, 1978). At that time, several laboratories throughout the world had been reporting that various hormones (including peptides and steroids), when injected directly into brain, could cause profound changes in the behavior of laboratory animals that were not dependent on the presence of the appropriate endocrine target organ (Nemeroff and Prange, 1978). One decade later, there is no argument that the very definition of neurotransmission has changed. Although bearing many similarities to classical neurotransmission, the neuropeptides differ in several respects.

Table IV.1. Compounds generally accepted as neurotransmitters.

<table>
<thead>
<tr>
<th>MONOAMINES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acetylcholine</td>
</tr>
<tr>
<td>Dopamine</td>
</tr>
<tr>
<td>Norepinephrine</td>
</tr>
<tr>
<td>Histamine</td>
</tr>
<tr>
<td>5-Hydroxytryptamine (Serotonin)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AMINO ACIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>gamma-Aminobutyric Acid (GABA)</td>
</tr>
<tr>
<td>Glycine</td>
</tr>
</tbody>
</table>

IV-4
Receptors can be demonstrated for the neuropeptides and pharmacological responses can be seen to the application of miniscule amounts of the peptides into discrete brain areas containing the appropriate receptors (Kalivas et al., 1982). A host of techniques have shown that neuropeptides occur in discrete pathways. Table IV.2 lists some of the compounds which are now generally agreed to serve roles as chemical messengers, despite differing somewhat in function from classical transmitters. There has been difficulty in identifying specific mechanisms in the CNS by which neuropeptides are synthesized and how their action is terminated. In some cases, the peptides may often be synthesized elsewhere (e.g., in the soma or even in other cells) and transported to the nerve terminal (Lundberg and Hökfelt, 1983).

The difficulties in elucidating these biochemical mechanisms suggest that the neuropeptides may serve regulatory roles that differ somewhat from the classic neurotransmitters. For example, colocalization of neuropeptides with traditional chemical messengers has been shown to be an important fundamental of impulse transmission. This was most elegantly demonstrated in the interaction of acetylcholine and vasoactive intestinal peptide (VIP) in the parasympathetic nerve terminals that innervate exocrine glands (e.g., salivary or sweat glands) (Lundberg, 1981). In these systems, the application of acetylcholine or VIP alone does not result in the full pattern of physiological effects caused by electrical stimulation of the presynaptic nerve. The actions of acetylcholine are rapid and localized to the immediate area of the nerve terminal. Conversely, VIP also causes effects in the vasculature at some distance (microscopically) from the nerve terminal that are slower to offset. When applied together, both compounds produced a potentiated physiological response, similar to that caused by nerve stimulation (Lundberg, 1981). Other neuropeptides have been shown to be co-packaged with classical neurotransmitters in neurons, for example: dopamine with enkephalin; serotonin with TRH; norepinephrine with neurotensin; and acetylcholine with somatostatin (Lundberg and Hökfelt, 1983). These findings provide a level of complexity that must be considered in attempting to design in vitro screens to predict neurotoxicity.

B. THE CHEMICAL ARCHITECTURE OF THE NERVOUS SYSTEM

When discussing biochemical and neurochemical mechanisms, we frequently rely upon a concept of the nervous system as a collection of homogeneous synapses (Figure IV.1). Although a single neuron may release several transmitters, it is clear that recent advances in neuroanatomy require consideration of the differences in neurotransmission engendered by the location of both the soma and terminal field. For example, with the catecholamines dopamine and norepinephrine, Ungerstedt (1971) demonstrated that the different dopamine terminal fields in
Table IV.2. Some of the other compounds presently accepted as chemical messengers (neurotransmitters/modulators).

**PEPTIDES**

- Opioid peptides (enkephalins; dynorphins; endorphins)
- Substance P
- Angiotensin II
- Cholecystokinin
- Neurotensin
- Somatostatin
- TRH (thyrotropin-releasing hormone)
- alpha- and beta-MSH (melanocyte stimulating hormone)
- ACTH
- Cholecystokinin octapeptide
- VIP (vasoactive intestinal polypeptide)
- Glucagon
- Luteinizing hormone releasing hormones (LHRH)
- Bradykinin
- Vasopressin (ADH)
- Oxytocin
- Bombesin
- CRF (corticotrophin releasing factor)
- Calcitonin-related peptides
  "Benzodiazepine"-binding-site peptide

**AMINO ACIDS**

- Taurine
- Glutamate
- Aspartate

**MISCELLANEOUS AND/OR MINOR COMPOUNDS**

- Purines
- Epinephrine
- Tyramine
- β-Phenethyamine
Figure IV.1. Model of synapse demonstrating many of the biochemical loci where effects of neurotoxicants have been studied.
adjacent areas of the brain originated from discrete groups of nuclei (Figure IV.2). Recent studies have indicated that even within a given nuclei, individual cells tend to innervate only portions of the target terminal field (Goedert et al., 1984). There are several consequences of this for the application of a given neurochemical endpoint. At one level, some areas may have marked recurrent circuitry, whereas with others this may be noticeably lacking. For example, there is a presynaptic feedback mechanism for dopamine neurons within the nigrostriatal pathway which appears to be lacking in the prefrontal cortex (Bannon et al., 1981). Within a given terminal field, it is clear that synaptic connections may occur on a variety of target neurons, and that even within one neuron, they may be quite heterogeneous.

A second level of complexity is engendered by the fact that different chemical messengers may be utilized in neurons with dissimilar innervation patterns. For example, in the CNS, the catecholamines generally originate from a few distinct cell nuclei, but innervate large areas of the brain in diffuse, but specific fashion (Ungerstedt, 1971). Conversely, the amino acid neurotransmitter gamma-aminobutyric acid (GABA) seems to function in local circuits or interneurons that occur almost everywhere in the brain (Gale and Casu, 1981). Modern neuroanatomy has provided such a wealth of information that even specialists in a given field have difficulty with keeping updated on the architecture for single transmitters. Thus, factors such as cotransmission and the complexity of neuronal architecture significantly limit the value of studies based solely on hypotheses constructed according to a scheme as is shown in Figure IV.1.

USE OF NEUROCHEMICAL METHODS

A. NEUROCHEMICAL AND BIOCHEMICAL-STRATEGIES PROPOSED FOR EVALUATING NEUROTOXICITY

The examples of synaptic biochemistry shown in Figure IV.1 provide some of the possible loci for toxicant attack and are also systems that may be indirectly perturbed by toxic insult. Attempts to incorporate measurements of such changes for use as potential neurotoxicity screens have resulted in several testing strategies. The three most commonly used strategies include radioreceptor methods, markers of neuronal integrity, and indices of neuronal activity. The particular applications that have been made include measurements of neurotransmitter concentration, neurotransmitter uptake, neurotransmitter synthesis, and neurotransmitter release. In addition, receptor affinity and density studies, as well as the ability of toxicants to occupy receptors, either as receptor agonists or antagonists, has also been extensively studied (DeHaven and
Figure IV.2. Origin of the different dopamine terminal fields in the rat brain (modified from Ungerstedt, 1971).
Mailman, 1983). Second messenger events, concentrations of functional and/or structural proteins, have also been commonly used (Mailman and Morell, 1982).

The use of such strategies is often based on the assumption that the schematic form of synaptic transmission described is similar to that shown in Figure IV.1 and that this will be an acceptable model. It is important to note that there is little understanding of the functional interactions of most copackaged chemical messengers (e.g., acetylcholine and VIP) and it appears that the criteria used to evaluate a neurotransmitter as developed from studying the classical transmitters (e.g., acetylcholine, norepinephrine, dopamine, serotonin) may not be totally adequate for these new classes of transmitters-modulators. Neuropeptides may act more slowly, over larger distances, and often act in concert with classical transmitters. The activity sites for each of the compounds listed in Tables IV.1 and IV.2 must then be considered as possible target sites for neurotoxicity. Moreover, compensatory events in the nervous system may also involve one or more of these chemical messengers. A selection, a priori, of a specific neurotransmitter system for a neurotoxicity screen is unlikely to be productive unless toxicant-induced damage is so severe that gross and widespread damage has occurred. In the latter case, there are more expeditious ways of achieving this end. These general reservations should be kept in mind when proposing special studies to evaluate neurotoxicity.

B. CHANGES IN CONCENTRATION OR CONTENT OF NEUROTRANSMITTER

Alterations in the concentration or content of the neurotransmitter have been commonly used to detect perturbation of the specific pathways in the CNS. It has been proposed that such changes may occur for several reasons such as the massive destruction of terminals or cells due to a variety of insults (e.g., anoxia, toxic exposure, tumors). Other mechanisms might include a differential change in the rate of synthesis, storage, or degradation of the neurotransmitter. Measurements of these fluctuations have been simplified for many transmitters by the availability of high performance liquid chromatography (HPLC). The use of HPLC with electrochemical detection was first applied in neurochemistry to the measurement of catecholamines, but is now used for dozens of compounds (Caudill et al., 1982; Desiderio et al., 1981; Gruber et al., 1976; Hancock et al., 1978; Keller et al., 1976; Kilts et al., 1981; Lindroth and Mopper, 1979; Mailman and Kilts, 1985; Mönch and Dehnen, 1977; Smith et al., 1975).

Numerous investigators have postulated that changes in the total amount of neurotransmitter present in neurons may reflect either destruction of cellular elements in which neurotransmitters reside, or perturbations that alter neuronal
activity or metabolism (Mailman and Morell, 1982). HPLC measurements of neurotransmitter levels provide a means for determining how toxicants affect neuronal function, while the examination of discrete brain regions may demonstrate preferential effects on specific pathways. It must be noted that such a strategy has an implicit weakness in that these steady-state concentrations may reveal little information about how a toxicant perturbs the CNS because these measurements fail to detect major alterations not affecting total transmitter content or concentration. Some evidence indicating that measurement of steady-state transmitter concentrations alone is of limited value include data showing that drugs which markedly perturb neurotransmission may not change transmitter concentrations per se. This has been clearly demonstrated with the cholinergic (Cheney and Costa, 1978) and dopaminergic systems (Korf et al., 1976; Umez and Moore, 1979).

C. OTHER INDICES OF NEUROTRANSMITTER FUNCTION

There are numerous strategies that can assess neurochemical function and which complement and extend the measurement of steady state concentrations. Several of these are summarized in Table IV.3.

Table IV.3. Some neurochemical strategies used to assess neuronal function in the central nervous system.

<table>
<thead>
<tr>
<th>In Vitro</th>
<th>In Vivo</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Assessment of &quot;turnover&quot;</td>
<td>• Brain dialysis</td>
</tr>
<tr>
<td>(Injection of labeled precursor</td>
<td>• Push-pull perfusion</td>
</tr>
<tr>
<td>or amine, use of enzyme inhibitors,</td>
<td>• Voltammetry</td>
</tr>
<tr>
<td>metabolites, etc.)</td>
<td>• Measurement of precursors or</td>
</tr>
<tr>
<td>• Measurement of enzyme activity</td>
<td>transmitters or metabolites in</td>
</tr>
<tr>
<td>• Measurement of enzyme uptake</td>
<td>blood, urine, or CSF</td>
</tr>
<tr>
<td>or release</td>
<td></td>
</tr>
<tr>
<td>• Receptor properties</td>
<td></td>
</tr>
<tr>
<td>• Cyclic nucleotide function</td>
<td></td>
</tr>
<tr>
<td>• Protein phosphorylation</td>
<td></td>
</tr>
</tbody>
</table>
1. Synthesis and turnover

A cogent strategy to pinpoint the effects of a toxicant on CNS function which avoids some of the pitfalls described above is to examine changes in the rates of formation and degradation of the neurotransmitter in question. A review of this literature has been prepared by Weiner (1974). Measurements of these changes can be made by using several approaches such as radiometric techniques or drugs that act like enzymes. Several radiometric procedures require the injection of a radioactive precursor to permit measurements of the rate of metabolite formation. Alternately, neuronal pools can be labeled with the radioactive neurotransmitter to provide measurements of the decline in specific activity. Although these methods have the theoretical advantage of not perturbing the system being measured, the values derived from the use of these techniques are influenced by many factors which include: diffusion from the site of injection; blood-brain permeability; equilibration with, or alteration of, the size of endogenous pools of amine; metabolism to compounds other than those being studied; uptake into non-neuronal elements; and different rates of turnover for the same neurotransmitter in different brain regions. Nonsteady state methods involve the use of drugs as enzyme or transport inhibitors with concurrent measurement of the rate of accumulation or decline of the amine and its metabolites. Several factors may confound these types of studies, including diffusion out of brain, the possibility of incomplete inhibition of enzyme activity of metabolite efflux, and secondary effects of enzyme blockade on elevation of precursor concentrations and feedback inhibition.

Good design necessitates the use of two or more of these approaches in a given study because each of these methods may have a particular disadvantage. Conclusions regarding the true effects of toxicants on synthesis or turnover rate can be reinforced if similar changes are seen using several techniques.

2. Uptake and release

Another frequently used hypothesis is that a toxicant (directly or otherwise) can either alter the release of neurotransmitters, alter transmitter uptake, or interfere with a major mechanism of inactivation or recycling. These changes could be brought about through effects on packaging or storage of neurotransmitter in synaptic vesicles, transport of vesicle to the synaptic cleft, or uptake mechanisms involved in recycling of the transmitter. In release studies, an uptake inhibitor may be incorporated into either membrane or tissue slice preparations to permit activity measurements (e.g., Sanders-Bush and Martín, 1982; Schoepp and Azzaro, 1982). Neurotransmitter release can be induced by either electrical stimulation or by exposure to buffers with high concentrations.
of potassium. Electrical stimulation has an advantage in that the frequency of stimulation can be regulated, and the transmitter release represents that of vesicular neurotransmitter under appropriate conditions. Less desirable is potassium-induced stimulation, because both vesicular and nonvesicular transmitter release may occur and the amount of release cannot be adequately controlled (Vargas et al., 1977). Studies with synaptosomal preparations are technically easier to execute, but have the disadvantages of using a homogenate of disrupted nerve cells and of being unstable following prolonged incubation. Tissue slices, however, provide an intact preparation of nerve terminals with which one can study release of a given neurotransmitter, as well as the interactions of two or more transmitter systems coexisting in the same brain region. Uptake experiments generally use membrane preparations to determine the rate of accumulation of labeled amine (Baldessarini and Vogt, 1971; Ross, 1982). Although uptake is dependent upon transport across the neuronal membrane, vesicular storage, rate of release, and presynaptic monoamine oxidase (MAO) activity, the main confounding factor is that the neurotransmitter can be taken up and bound nonspecifically in other cellular compartments. However, with appropriate assay conditions and pharmacological controls, many of these problems can be overcome.

3. Receptor binding

The radioreceptor assay has become one of the most utilized techniques in neurobiology and it has been proposed to have special utility in neurotoxicology (Bondy, 1982). Theoretically, these procedures can be used to assess the effects of prior toxicant exposure on receptor characteristics (e.g., density and affinity), or the ability of the toxicant to interact directly with a receptor population and alter the binding of specific ligands (such as the endogenous transmitter). These methods are technically simple, but one must be certain that the criteria of saturability, specificity, and reversibility of binding are met and that procedures are followed precisely, since small variations in method can affect binding. The receptor binding approach involves homogenization and washing of brain tissue in buffer at physiological pH, and the incubation of tissue with radiolabeled ligand until equilibrium has been reached. The addition of an excess of unlabeled drug is used to determine nonspecific binding (i.e., binding to receptors or elements other than those under study). In either case, separation of membrane-bound ligand from free ligand is by filtration or centrifugation, thus terminating the reaction. The application of these methods to neurotoxicology has recently been reviewed with particular emphasis placed on the misuse of this technique (DeHaven and Mailman, 1983).
4. **Second messenger changes**

Second messenger changes (see Figure IV.1) may provide a sensitive method to detect perturbations in neuronal function. Two biochemical systems that are engaged by receptor interactions or changes in neuronal activity are cyclic nucleotide metabolism and protein phosphorylation. For catecholamines like dopamine, the neurotransmitter activates receptors, which indirectly stimulate the enzyme adenylate cyclase, thus causing increased conversion of ATP to cyclic AMP (Clement-Cormier et al., 1974; Kebabian and Saavedra, 1976; Kebabian et al., 1972). Cyclic AMP can activate protein kinases, resulting in the phosphorylation of numerous target proteins, ultimately resulting in changes in protein synthesis within the cell (Greengard, 1976) which then influence neuronal transmission by altering the permeability of the neuronal membrane to ions (Nathanson, 1977) or by affecting a variety of other biochemical responses. In the CNS, protein phosphorylation may influence a host of processes including synthesis of neurotransmitter receptors (Greengard, 1978), the regulation of enzymes involved in neurotransmitter synthesis (Kuhn and Lovenberg, 1983), or the function of microtubules involved in axon transport (Greengard, 1976). Finally, cyclic nucleotides are only some of the compounds for which second messenger roles have been clearly demonstrated. Other compounds critical to CNS function include the inositol phosphates and arachidonic acid metabolites. There are literally dozens of possible biochemical endpoints that may act either as sites of toxicity or may be affected by toxic insult.

5. **In vivo neurochemistry**

One of the major disadvantages to the use of pathological, biochemical, and molecular methods as a preclinical screen for neurotoxicity is that only one measurement can be made at the time that the test animal is sacrificed from the study. Technologic developments permit the detection of neurochemical changes in freely moving animals. The results from these techniques provide data consistent with hypotheses made using the in vitro methods described earlier (Ungerstedt et al., 1982). Although the push-pull perfusion techniques described by Tilson and Sparber (1970) and by Myers (1974) have been used to examine synthesis, turnover, and release of neurotransmitters and related compounds in vivo, this technique has been refined recently by the use of dialysis tubing instead of a steel cannula for implantation (Hernandez et al., 1983; Ungerstedt et al., 1982). When assessing neurotransmitter synthesis or turnover, these methods suffer from some of the same limitations described previously. In addition, appropriate controls must be included to account for possible effects resulting from the lesions caused by cannula implantations. However, brain perfusion does have several advantages. The concentration and time of administration of the compound to which the tissue is exposed can be accurately
controlled. Effects of toxicants on specific brain structures or nuclei can be examined, and this can be done in the awake animal in conjunction with behavioral toxicologic studies. With technical and theoretical improvements, this method can offer some potential for the future.

Another technique which should have important use in specific circumstances is in vivo voltammetry. This method assesses neurotransmitter function in the unanesthetized animal by the use of a microelectrode (similar to the one used in HPLC analysis) that is implanted into the brain (Adams, 1978; Kissinger et al., 1973). A small electric potential is applied to the microelectrode to measure current flow. The oxidation of certain neurochemicals, such as catecholamines, indoleamines, ascorbic acid, and others can be quantified using this technique. Some of the initial problems encountered with this method, such as interference from other endogenous compounds that oxidize at similar potentials and overlapping of peaks that oxidize at the same potential, have been resolved by technical advances (Brazell and Marsden, 1982; Ewing et al., 1982; Lane et al., 1976). The advantages of in vivo voltammetry are numerous, despite the fact that the contribution of extraneuronal compounds to the electrical signal must also be considered (Kennett and Joseph, 1982). As with brain perfusion and dialysis, one can monitor ongoing processes in the awake animal, but unlike these latter techniques, the electrode is the only exogenous element introduced into the brain region under study, thus eliminating the need for perfusion fluid or radioactive tracers. Changes seen by in vivo voltammetry can be verified by HPLC analysis, and may be correlated with behavioral changes (Yamamoto et al., 1982).

D. CAUTION IN THE USE OF THESE METHODS

From this discussion it is clear that there is a wide variety of neurochemical techniques available for the assessment of neurotoxicity. It must be emphasized that an assortment of neurochemical changes that are not specific to certain neurotransmitter systems often result from toxicant exposure and that even the so-called "model" and "specific" neurotoxicants that are targeted to single populations of neurotransmitters (e.g., catecholamine 6-hydroxydopamine) will cause alterations in other neurotransmitter systems. It has proven difficult to classify neurotoxicants based on their neurochemical sequelae and such an endpoint should not be expected due to this lack of specificity. The ideal approach to the study of neurotoxicity is to combine appropriate neurochemical techniques with tests that assess those behaviors affected by the neurochemical changes observed. At present the ideal use of in vitro methods is to study mechanistic, rather than phenomenological, endpoints. This may include the use of pharmacological challenges, neuropathological studies, and neurobehavioral toxicology. This
multifaceted strategy will provide important information in evaluating the neurotoxicity of a particular compound and can provide clues for more detailed biochemical and neurochemical studies.

E. THE APPLICATION OF BIOCHEMICAL METHODS IN NEUROTOXICOLOGY

The availability of biochemical techniques has led to advances of direct interest to neurotoxicologists. For example, the pyrethroids have been known to affect both vertebrate and invertebrate nervous systems. The behavioral effects that can be observed following the administration of pyrethroid insecticides to mammals produces seizures, tremor, and choreoathetoid movements (Ray and Cremer, 1979), as well as decreases in motor activity and differential effects on the acoustic startle response (Crofton and Reiter, 1984). The use of biochemical methods was an important element in understanding the mechanism of pyrethroid activity even though it was known that deltamethrin or fenvalerate neurotoxicity can be attenuated by diazepam (Gammon et al., 1982), and diazepam or aminoxyacetic acid attenuate the neurotoxicity of permethrin (Staatz et al., 1982a). Pyrethroids have been shown to affect dihydropicrotoxinin (Leeb-Lundberg and Olsen, 1980) and kainic acid (Staatz et al., 1982b) binding sites in the central nervous system, suggesting that pyrethroids may act centrally via GABAergic or glutaminergic mechanisms. Lawrence and Casida (1983) have reported that various pyrethroids inhibit binding of $^{35}$S-(-)-butylbicyclophosphorothionate to the picrotoxinin site of the GABA receptor complex. Radioreceptor studies such as these have extended the mechanistic findings from earlier electrophysiological studies and have given insight into the in vivo effects of toxic compounds on the nervous system.

In contrast, the cholinesterase inhibitors (e.g., organophosphates) apparently follow identical modes of action, but are known to cause different sequelae. A variety of neurochemical studies can provide an understanding of these differences, however, tolerance develops to organophosphates upon chronic administration. One mechanism for this tolerance is receptor subsensitivity (Costa et al., 1982a). Tolerance to the behavioral changes induced by these compounds also occurs, as noted by the effects of diisopropylphosphorofluoridate (DFP) on drinking water (Chippendale et al., 1972), fixed-ratio responding (Russell et al., 1975) and single alternation behavior (Overstreet et al., 1974), as well as to the antinociceptive and hypothermic effects of disulfoton (Costa et al., 1982b) and to the effects of paraoxon on two-way avoidance responding (Giardini et al., 1982). These effects may be the result of an adaptive change in receptor characteristics. Several investigators have demonstrated that compounds such as DFP (Ehlert et al., 1980; Schiller, 1979; Uchida et al., 1979), disulfoton (Costa et al.,
1981, 1982b; Schwab et al., 1981), paraxoxon (Smit et al., 1980), and Tetram (Gazit et al., 1979) caused decreases in the $B_{\text{max}}$ of muscarinic receptors. There is a correlation between the neurochemical responses in cholinergic systems and the behavioral tolerance resulting from exposure to these organophosphate compounds.

THE USE AND ABUSE OF NEUROCHEMISTRY IN PREDICTING NEUROTOXICITY

The availability of powerful in vitro and in vivo neurochemical methods has led to their application in studies of the purported neurotoxicity of numerous compounds. Two recent examples provide a cogent illustration of the advantages and disadvantages in the use of such methods to detect neurotoxicity.

A. THE CASE OF FD&C RED NO. 3

1. Background of issues

As a consequence of Feingold's 1975 book, Why Your Child Is Hyperactive, the view that our diet can have widespread toxic effects received great publicity and support. Feingold (1975) proposed that diets having no artificial flavors, colors, or certain natural ingredients can dramatically alleviate symptoms in one-third or more of hyperactive children. Although open trials have generally supported these contentions, rigorous double-blind and challenge trials have disputed Feingold's claims (Lipton et al., 1979a). This issue remained in the clinical domain for several years, until in 1979, when it was reported that the mixture of eight food colors used in pediatric clinical studies inhibited uptake of dopamine and other neurotransmitters in crude rat brain synaptosomal preparations. It was also found that this effect was due completely to one color, erythrosin, whereas the other seven dyes were ineffective (Laffermand and Silbergeld, 1979; Logan and Swanson, 1979). These data were greeted with great excitement because such results, if true, offered the potential of providing a testable mechanism, as well as providing the means of designing focused clinical trials.

2. Neurochemical investigations relevant to erythrosin

After the initial studies reporting that erythrosin would inhibit neurotransmitter uptake (Laffermand and Silbergeld, 1979; Logan and Swanson, 1979), a variety of strategies were pursued. Silbergeld attempted to define which aspect of the biochemistry of neuronal uptake was affected by erythrosin. She concluded that erythrosin inhibited the ouabain binding site linked to sodium channels, and by disrupting the transport of sodium, erythrosin inhibited neuronal uptake of
sodium into the cell membrane (Silbergeld, 1981). In addition, other groups (i.e., Augustine and Levitan, 1980; Colombini and Wu, 1981; Floyd, 1980; Levitan, 1977; Sako et al., 1980) found that erythrosin had effects on many endpoints which were essential to proper neuronal functioning.

These reports were of special interest to those mental retardation researchers led by Lipton who were particularly cognizant of the importance of dose-effect relationships, as noted in a major scholarly review on the use and effects of megavitamin therapy (Lipton et al., 1979b). Because the biological activity of erythrosin seemed to be so far ranging, they questioned why little clinical toxicity had ever been reported for a compound with such effects. Their initial hypothesis was that the many warnings of potential neurotoxicity from the in vitro tests were made without giving proper consideration to certain fundamentals of toxicokinetics at the level of the cell, organ, or organism.

They performed two types of experiments to deal with this issue. First, they explored the effects of erythrosin on the inhibition of dopamine uptake in nerve cells. They found that the ability of erythrosin to inhibit synaptosomal dopamine uptake was inversely proportional to, rather than linearly dependent upon, the concentration of tissue used in the incubation systems (Mailman et al., 1980). In the presence of 10 μM erythrosin, significant inhibition of dopamine uptake was seen at 1 mg tissue per ml, whereas no inhibition was seen at 8 mg per ml. Similar types of tissue dependency were noted using other membrane-localized biochemical parameters, such as adenylate cyclase activity or radioligand binding (Mailman et al., 1980). Mailman and DeHaven (1984) offered a model for such effects in which they proposed that the structure of erythrosin was such to permit nonspecific disordering of biological membranes. Any biological process dependent on the integrity of such membranes would be affected by exposure to such a compound. An increase in the amount of membrane added to the in vitro test system decreased the apparent potency of erythrosin.

3. Was erythrosin a neurotoxin?

The resultant hypothesis extended from these studies was that in the low doses in which erythrosin was ingested (even in high dietary use), it was extremely unlikely to have effects on the central nervous system (Mailman and DeHaven, 1984; Mailman and Lewis, 1981; Mailman et al., 1980). To verify this, it was necessary to know the toxicokinetics of this chemical. A reverse phase, HPLC method for quantifying erythrosin revealed that after administration of 50 mg/kg (i.p.) of erythrosin to adult rats, concentrations of erythrosin in kidney and liver were at least tenfold greater than in blood.
during the first 24 hours after injection. Conversely, the erythrosin concentration in unperfused brain was less than half that in blood. When brains were perfused at sacrifice, the concentration of erythrosin remaining in brain was decreased markedly (Niedzwiecki and Mailman, 1981). These data suggest that even after massive doses of erythrosin, it does not reach significant concentrations in brain, making neurotoxicity at relevant human doses unlikely. A second line of evidence in support of this hypothesis was the behavioral toxicological data which showed that erythrosin caused significant effects only after intraperitoneal doses of 50 mg/kg or more (Mailman et al., 1980).

These neurochemical studies cannot resolve all of the clinical issues (Mailman and Lewis, 1983); however, these data are relevant to one narrow issue of the neurotoxicity of erythrosin. Swanson has repeated the clinical trials with numerous food colors. He found small, but significant, effects with the mixture of eight food colors. When using erythrosin alone, no effect whatsoever was reported (Anonymous, 1982). This offers the best evidence to date in support of this hypothesis. At least until substantive and well-controlled data to the contrary can be provided, erythrosin should not be labeled as a neurotoxicant. As this example illustrates, the inappropriate application of testing methods can lead to misleading conclusions with significant secondary ramifications. The power of modern neuroscientific methods is great; however, their use is best when directed towards specific questions.

B. THE CASE OF MPTP

1. Origin of problem and clinical findings

Several years ago, Langston and coworkers (1983) were responsible for identifying the compound 1-methyl-4-phenyl-1,2,5,6-tetrahydropyridine (MPTP) as a source of idiopathic parkinsonism. MPTP was an undesired contaminant of the "designer drug" synthesis of 1-methyl-4-phenyl-4-propionoxy-piperidine (MPPP), an opiate-like drug. It became clear to numerous laboratories that this was an excellent research opportunity to develop better animal models of parkinsonism. In addition, the availability of a new cytotoxicant might provide a new and powerful research tool. This compound provides an excellent example about the inherent strengths and weaknesses of the use of biochemical and molecular methods to predict neurotoxicity.

It is important to review what is believed to be the sequence of events that ultimately causes the parkinson-like symptoms. First, MPTP apparently binds to monoamine oxidase B located in glia. The enzyme subsequently metabolizes the parent compound to 1-methyl-4-phenylpyridinium (MPP+). The MPP+ is taken up by active uptake into both dopamine and adrenergic neurons. From there it apparently binds and is concentrated
on dopamine neuromelanin, but not norepinephrine neuromelanin. The consequence of this (by a still unknown mechanism) is the destruction of the affected dopaminergic neurons. Since idiopathic parkinsonism is known to involve the death of these same cells in the substantia nigra, it is not surprising that similar symptoms result, nor that similar therapy is ameliorative (see Lewin, 1985 for review).

2. Relevance of MPTP toxicity: predictive ability?

It is interesting to speculate what the result of a battery of neurochemical methods directed to evaluate MPTP would have shown. Let us choose several of the commonly suggested methods mentioned earlier. The receptor binding assays would have probably been totally negative, since there is no evidence that MPTP binds with any affinity to pre- or postsynaptic receptors of any type. Uptake studies with a variety of tissue systems would also have been negative. It is interesting to speculate that these false negative results would have occurred even selecting chemical messenger systems in which the transmitters bear structural resemblance to MPTP. If neuropathological or behavioral studies had been done in rodents, they would have shown only a very low degree of toxicity. On the other hand, the elucidation of the mechanisms of MPTP toxicity provided information of great value. It is clear that any compound that might be metabolized to, or exists as, a structural analog of MPP+ may cause very similar effects. This provides predictive ability. Compounds such as the herbicide paraquat and its analogs are immediately suspect as neurotoxins.

An intriguing hypothesis is the suggestion that idiopathic Parkinson's disease may itself be an environmentally induced illness. If true, the human health ramifications and the potential to prevent devastating human illness is great; however, the lessons of erythrosin must be applied to this situation. To give weight to this environmental mechanism of idiopathic parkinsonism, one must a) identify the potential offending chemicals, b) determine that they are delivered to susceptible individuals at sufficient doses and for sufficient time to cause the parkinsonism, and c) confirm that the temporal and neuropathological courses of idiopathic parkinsonism and toxicant-induced parkinsonism are similar enough to support such an hypothesis. This will require not only good neuroscience, but also reliance upon analytical chemistry and epidemiology. The answers to such questions are as yet unavailable, but on the superficial examination of the data, it may be unwise to magnify this problem past the area of narrow exposure to these compounds via "designer drugs" or industrial contamination limited to a few individuals. On the other hand, many compounds structurally similar to MPP+ should certainly now be regarded as neurotoxicants.
FUTURE ROLE OF MECHANISTIC STUDIES IN PREDICTIVE NEUROTOXICOLOGY

A. IMPROVED METHODS OF PREDICTION AND DETECTION

The biochemical and molecular neurosciences have been particularly important in developing an understanding of the mechanisms involved in the injury to, and response by, the nervous system. As this information continues to be gleaned, it seems appropriate that many biochemical events can be examined, a priori, with known compounds of suspected toxicity. This includes considerations of bioconversion and of mechanism of action. This ability to predict toxicity will best come from theoretical examination of the structure of suspect compounds as well as predicted metabolites versus known target sites and their corresponding structure activity-relationships, a task that can be made quite manageable by the availability of modern computer storage and search systems. This coupled with screening in intact target organisms (e.g., using neurobehavioral or neuropathological methods) can most efficiently predict toxicity.

There are many areas of great potential interest that have been under-investigated. One of the most intriguing is the idea that chronic administration of centrally active, clinically used drugs may cause neurotoxicity. It is recognized that this may occur with some agents, such as the tardive dyskinesia caused by antipsychotic drugs. There may be other possibilities that become evident as we learn more about the mechanisms involved. This area seems to be one of particular importance and it would seem prudent to carefully examine how antipsychotic, antidepressant, and anxiolytic drugs affect the nervous system, in addition to mechanisms involved in their therapeutic effects. Since most agents have structures consistent with metabolism to potentially reactive intermediates, this possibility seems of some interest.

B. SUMMARY

Studies during the past decade have provided lessons which should be of value for the immediate future. The fundamental principles of pharmaco-toxicodynamics and kinetics while recurring as issues, are continually shown to be of great value. One important point is that care must be given to the application of pharmaco-toxicokinetics. A priori predictions of metabolism and distribution should be an integral part of in vivo toxicity studies, especially when using knowledge of the substrate specificities of metabolizing systems (e.g., various cytochromes P450, flavin-containing monoxygenase, etc.). In addition, dose-response/dose-effect relationships should be a key part of interspecies extrapolations. Conclusions based on in vitro data should be made only after relation to estimated or empirically determined tissue concentrations. Species differences are an
important consideration in terms of both functional differences of homologous neuronal systems and the species variation in biochemical characteristics.

In the future, the design of cogent neurotoxicological studies may also be greatly aided by predictions based upon chemical structural features. The power of such an approach is likely to increase rapidly with the concomitant growth in the biochemical neurosciences and the increasing availability of inexpensive and powerful computers. After demonstration or suggestion of toxicity in lower animals, or in cell cultures, biochemical and neurochemical methods can then provide the most powerful method to elucidate underlying mechanisms. These resulting data can then be of predictive value for human toxicology.

The large number of potential loci of toxicant action within the nervous system precludes the development of a useful simple battery of neurochemical screens. Despite the problems with interspecies differences in both architecture of the nervous system, and in toxicant dynamics, the use of in vivo neurotoxicity testing remains the system of choice. However, the number of in vivo toxicity studies may be decreased, and their power may be increased, by proper hypotheses formulated from considerations of chemical structure.
OPEN DISCUSSION

HUTCHINGS: Are there some questions or points of clarification for Dr. Mailman?

REITER: I would like to expand on your comment regarding structure activity relationships and suggest that we extend the use of structure activity relationships in ways other than the traditional stick molecular models. There are a number of research groups in the United States that are looking at structure activity relationships from the standpoint of charge distribution on the molecules. Perhaps some of the inconsistencies or ambiguities relative to structure activity relationships may be that we are looking at structure in the wrong way.

MAILMAN: I hoped to raise exactly that point in my remarks. All the chemical disciplines have already accepted this concept and I think the time is ripe for adapting and adopting this technology to our purposes. Although writing the programs and creating the data base will take time, it is certainly possible, and can be a powerful technique for prediction.

HARRIS: Dr. Harris, at the EPA. My question relates to MPTP. I know that the primate is more sensitive to the Parkinsonian-like experience that these animals have with this compound vis-a-vis the rodent. Since this relates to testing, could you tell me, how do the rats react to MPTP and would our classic testing (e.g., reproduction testing, which contains a behavioral component) pick up the toxicity of MPTP or is it really a hidden toxicant?

MAILMAN: Only at very high doses.

HARRIS: What do you see, then?

MAILMAN: Because you're getting disruption, you'll see neurological symptoms. Experience with the compound indicates that it is relatively selective for dopamine neurons that project to an area called the striatum. The dopamine neurons involved in endocrine regulation are not affected, at least in man and monkeys. There is little data at present in rats, because it requires much higher doses. Neurological symptoms are generally seen before anything else.
HARRIS: So in a chronic testing paradigm, we would see toxicity at high doses, since we do test very high doses in the chronic or reproduction testing.

MAILMAN: MPTP appeared to have little effect in the initial rodent studies. After the initial etiological link was discovered from the clinical data, the drug was reevaluated in rats and found toxic, but only when massive doses were used.

HARRIS: Yes. You made a comparison of the use of MPTP as a model of idiopathic Parkinson's disease. There is one major difference which makes this a limited model. Parkinson's disease is a progressive phenomenon, whereas the effect of MPTP is not. As I understand, MPTP hits and then you can treat with dopa, to reverse some effects?

MAILMAN: While we do know the etiology of these symptoms of Parkinson's disease, and while L-DOPA has benefits, there is no evidence of recovery, and these patients may debilitate with time.

One thing that we do not know, and maybe Dr. Spencer can correct me, is what will happen with time in individuals exposed to a very low dose of MPTP who may show only mild neurological effects? There are so few individuals known to have ingested this compound, and since most of these tend to be drug abusers, they very well may be taking other compounds which confound our observations.

The speculation that idiopathic Parkinson's disease is related to environmental causes is a long way from proven, and it may not even be a good hypothesis. It is interesting, and able to be tested, however.

HARRIS: With respect to the mechanism of action of this compound (i.e., the monoamine oxidase conversion), are you talking about its formation into an aldehyde and a free radical from that?

MAILMAN: It is believed that the methyl pyridinium ion, formed from MPTP by action of monoamine oxidase, is either the toxicant or its immediate precursor. An aldehyde is not believed to be involved.

HARRIS: That would be a free radical as being the toxicant.
MAILMAN: The methyl pyridinium ion is stable, although it might lead to generation of free radicals.

HARRIS: And that is the toxicant and not the free radical that's formed from it?

MAILMAN: This is not fully understood, but there is toxicity if MPP+ is administered. It could initiate many other events, such as free radical cascades. Snyder's group believes that it binds to neuromelanin concentrating within the cell and might even act as an internal detergent (Javitch et al., 1985).

HARRIS: How are you giving the MPP+ to get it into the central nervous system?

MAILMAN: Actually, it is believed the MPP+ is formed in situ from MPTP.

HARRIS: But you said when you also give the ion --

MAILMAN: I believe the accepted view is that the pyridinium ion is the compound that is a proximal toxicant. One can administer MPP+ directly into the brain where it is taken up by the active uptake system. In the brain, dopamine is also formally charged, and while it will not readily cross the blood brain barrier, it can be taken up into dopamine neurons when given centrally, much like MPP+.

HUTCHINGS: Are there some other questions?

LIPMAN: I'm at Vanderbilt University. In one of your conclusions you said that if we had pharmacokinetic data, then we could rely on these experiments instead of the traditional toxicologic animal studies. To follow such a strategy could lead to our missing an enormous number of toxic effects. I'm thinking of drugs like swensonine from loco weed, which produce neurological defects such as staggering, slobbering, and apparent hallucinatory performance in cattle and sheep. If you were to look at the pharmacokinetics of the drug alone, you would find that it accumulates in the kidneys and is water-soluble. You could end up looking at the wrong organ system, where in this case, the biological lesion persists in the absence of the drug in the brain.
MAILMAN: I did not mean to imply that there is a single way to do everything. Although there is no perfect strategy, it is important to remember that there are false negatives as well as false positives. This is one value of using an interdisciplinary approach. We need to rely upon a variety of disciplines; including neurochemistry, behavior, and analytical chemistry. A single approach would be prone to errors that would miss some hazardous compounds or mislabel innocuous compounds as toxicants.

With some common sense and various disciplines working together, the less likely we are to pursue red herrings.

LIPMAN: Right. The point that I want to make is that we are trying to reduce the number of animals used in experiments. A statement such as this could be interpreted by those who did not understand it as a reason for not doing the chronic toxicity study I gave you in my example (i.e., 6 weeks' chronic treatment of cattle). If we want to do similar experiments to see if humans can be safe, we have to do chronic neurological testing. There is no alternative.

MAILMAN: These studies are an essential part of all toxicological sciences. It might be argued, however, that on occasion, the outcome of some long-term chronic toxicity studies, a priori, could have been predicted prior to the study. I am only suggesting that, on balance, we have to weigh the merits of different approaches.

DAVIS: I'm with the Office of the Assistant Secretary for Health. I have a point of clarification on the differences between MPTP-induced Parkinsonism and idiopathic Parkinsonism. There are pathological differences that have been found. In MPTP-induced Parkinsonism brains, (although there is only one thus far), there are no Lewie bodies, which are the classic sign of idiopathic Parkinson's disease. The biochemical studies of spinal fluid collected from lumbar punctures have shown that only the dopamine system is affected in MPTP-induced Parkinson patients. With idiopathic disease, there is a spectrum of neurotransmitters that are affected; however, there is anecdotal evidence of a progression of deterioration with MPTP-induced Parkinsonism.

One of the patients in California was a bookkeeper who did a lot of writing in the course of her work before she came for treatment. It was discovered later that there was about 2 years worth of handwriting samples which showed a classic decrease in the size of her handwriting over the course of that period of time. This provides some anecdotal evidence of a similarity in terms of progression.
MAILMAN: Again, I did not mean to imply that I believe the hypothesis of chemical induction of "idiopathic" parkinsonism is valid; I feel that this hypothesis is not true. Nonetheless, the idea is something that people should think about. More directly, one of the problems with the anecdotal clinical evidence in terms of the disease progression is that we do not know what exogenous chemicals your bookkeeper, for example, may have been taking during that period of time. To me, the object lesson is not that idiopathic Parkinsonism is caused by MPP+ or related compounds, but rather, that there may be other compounds which occur at high concentrations in very localized environments (i.e., certain industrial processes, other designer drugs, etc.). Any one of these might cause drug-induced Parkinsonism that we could do something to prevent.

SPENCER: Information is available pertaining to your last comment. We have known for about 50 years that occupational exposure to carbon disulfide increases the incidence of Parkinsonism in those individuals who are occupationally exposed for prolonged periods of time. This point seems to have been forgotten by the MPTP groups, largely because this literature is hidden and rarely referred to. There is a wealth of information in the occupational toxicity literature that is of value to our neurological colleagues.

On the point as to whether MPTP or other agents might induce Parkinsonism, it is important to remember that Parkinsonism might be just one complex or one syndrome produced by a number of different agents. Consider the polyneuropathies where there can be genetic, nutritional, and toxic causations. We would be laughed at if we had suggested some years ago that all polyneuropathies were caused by some environmental toxic factor.

One of the contributions that toxicology might make is to reveal some structure activity relationships between molecules and particular effects within the nervous system. This might lead to discovery of endogenously-generated, potentially toxic metabolites, or agents which are normally present within the physiological system (e.g., cyanide), which when elevated to a particularly high level will precipitate a neurological disease. There is a tendency to overlook the possibility of endogenously-generated toxicants. The brain is not concerned whether the agent comes from outside the body or is generated from within. It's primary concern is the interface between the neuron, glia cell, and tissue fluid emanating from blood.
MAILMAN: I agree completely with your last comment. One point I emphasized was the possibility that endogenous generation of reactive species that may occur from a chronically administered drug might contribute to some of the long-term side effects that we see. This is an important point and provides opportunity for prediction in the future.

HUTCHINGS: Thank you, Richard.
LITERATURE CITED


Schwab, B.W.; Hand, H.; Costa, L.G.; Murphy, S.D. 1981. Reduced muscarinic receptor binding in tissues of rats tolerant to the insecticide disulfoton. Neurotoxicology 2:635-647.


V. BEHAVIORAL INDICES OF NEUROTOXICITY:
WHAT CAN BE MEASURED?

Hugh A. Tilson, Ph.D.
Laboratory of Behavioral and Neurological Toxicology
National Institute of Environmental Health Sciences
P.O. Box 12233
Research Triangle Park, North Carolina

ABSTRACT

The ability to identify and characterize the potential neurotoxicity of chemicals is an important and necessary function of various regulatory agencies. Behavioral assessments of toxicity may be important because of their relative sensitivity to some chemicals, their generally noninvasive characteristics, and their ability to measure toxicity in organ systems other than the nervous system. Behavioral tests can be classified by several criteria including traditional experimental definitions, their desired experimental usage, the neurobehavioral functions they are designed to assess, and the strategy chosen for their use in the evaluation of chemicals. Examples of neurobehavioral tests used to evaluate the effects of chemicals for toxicity include those that evaluate motor (spontaneous motor activity, motor coordination, weakness, abnormal movement or posture, tremor, and on-going performance), sensory (screening, reflex modification, and instrumental conditioning), learning/memory (nonassociative and associative), instrumental performance (schedules of reinforcement), and naturally occurring responses (consummatory behaviors). Behavioral procedures have also been utilized in select ways in toxicological research to detect latent damage, to study mechanisms of action, and to screen for functional dysfunction following exposure during development. Many considerations, such as the behavioral mechanism of action, definition of an adverse effect, problem of functional reserve, and several statistical questions, should be taken into account in the use and interpretation of data obtained from behavioral tests. During the last decade, there have been numerous recommendations from groups within the United States and, most recently, the World Health Organization, suggesting that systematic observational assessments may be appropriate when carried out within already existing toxicological protocols.
INTRODUCTION

A. STATUS OF NEUROBEHAVIORAL TOXICOLOGY FOR REGULATORY PURPOSES

Neurotoxicity is not routinely used as an endpoint for environmental decision-making. With the exception of certain regulations under the Toxic Substances Control Act (TSCA) of 1976 and the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) of 1972, only overt signs of neuropathy are used by regulators to evaluate nervous system injury. Under Section 5 from TSCA, the Environmental Protection Agency (EPA) has the responsibility to assess possible health hazards after a manufacturer has notified EPA of plans to produce a chemical that is not presently listed on the chemical inventory. EPA can issue an order requiring data on neurotoxicity if structure activity relationships, information from the literature, or data submitted by the manufacturer causes suspicion that there may be neurotoxicologic effects. EPA can also restrict the use of the chemical or prohibit it from entering into the market place until required data have been submitted to the agency and reviewed by regulatory scientists. Under Section 4 from TSCA, if the EPA suspects neurotoxicity for chemicals already on the market, then a test rule would be used to obtain additional data relevant to neurotoxicity (U.S. Environmental Protection Agency, 1984a).

Except for delayed neurotoxicity testing, the EPA has no standard testing requirements under FIFRA for the registration of a pesticide. The goal of EPA is to detect neurological effects from the data presented in other required toxicity studies for acute and subchronic, oral, dermal, and inhalation testing. If the agency determines that there is the potential for persistent or permanent neuropathy, then additional tests are required (U.S. Environmental Protection Agency, 1984b). With the exception of the organophosphate pesticides, there are no required guidelines as to how such tests should be conducted under FIFRA.

The Food and Drug Administration (FDA) has approached its evaluation of safety along conventional lines for toxicity testing, including mortality, pathology, carcinogenesis, growth disorders, clinical abnormalities, organ impairment or failure, reproductive disorders, or teratogenic potential (Sobotka, 1986). Although there is recognition that drugs, environmental chemicals and dietary chemicals can influence neural function and behavior, there is no clear mandate for the routine inclusion of behavioral data in required FDA safety assessment.

The responsibility of evaluating the safety of exposure to industrial chemicals is regulated by the National Institute of Occupational Safety and Health (NIOSH). The extent to which toxic effects on the nervous system can be observed has served
as part of the basis for recommended exposure limits by this agency (Anger, 1984). The American Conference of Governmental Industrial Hygienists (ACGIH) publish recommendations for Threshold Limit Values (TLVs) of exposure to many chemicals found in the workplace which are based on criteria published in the TLV documentation book (American Conference of Governmental Industrial Hygienists Inc., 1982). In the 1982 edition of the ACGIH TLV booklet, the TLV value for 167 of the 588 chemicals listed was based at least in part on behavioral and/or neurological effects.

B. RATIONALE FOR USE OF BEHAVIORAL TESTING PROCEDURES IN TOXICOLOGY

Behavioral testing procedures were derived largely from experimental and behavioral psychology and have been used in both pharmacology and toxicology for several years (Alder and Zbinden, 1983; Evans and Weiss, 1978; Norton, 1982; Tilson and Mitchell, 1984). Although it may seem apparent that those procedures used in pharmacology might also be directly applicable to toxicology, this may not necessarily be the case because in pharmacology, a specific type of response is predicted and defined in terms of efficacy relative to a therapeutic or desired effect; while in toxicology, any given effect may be unpredictable and associated with structural damage.

In toxicology, many individuals advocate the use of functional observations as one criteria for measuring the toxic effect of chemicals. Neurochemical, neurophysiological, and neurobehavioral endpoints have all been used as indicators of nervous system function. An alteration in behavior might be a relatively sensitive indicator of exposure because behavior involves an integration of several underlying processes and neurofunctions including motor, sensory, attention, motivational, and reactivity.

Behavioral testing procedures are important in both pharmacology and toxicology as they are generally noninvasive and can be used to assess subjects repeatedly during the course of chronic exposure studies. These noninvasive techniques are of particular value in those studies where the chemical effect might accelerate age-related neuronal degeneration. For example, if additional decrements in structural capacity of the brain occurred at a rate as low as 0.1% per year due to repeated exposure to a chemical, then the eventual result might be an accelerated degenerative process (Evans and Weiss, 1978). This would be particularly true if exposure began at a relatively young age and continued for many years. Some types of functional or behavioral tests might be sufficiently sensitive to detect such subtle changes.
Many behavioral testing procedures serve a dual role in toxicologic assessment studies as they provide information for specific neurotoxic effects as well as useful data for the study of general toxicity. For example, consummatory behaviors such as feeding and drinking may be affected by many toxicants and might be relatively sensitive indicators of toxic exposure.

This review defines behavior within the framework of its current use in pharmacology and toxicology. While a comprehensive review of this literature has been previously reported by Tilson and Mitchell (1984), numerous examples will be presented to illustrate specific points about the use of behavioral testing procedures and the interpretation of data from these studies. The critical issues concerning the use of behavioral tests in toxicology will be discussed.

CLASSIFICATION OF BEHAVIORAL APPROACHES

A. DEFINITION OF BEHAVIOR

Behavior may be defined as the movement of an organism or its parts within a temporal and spatial context (Tilson and Harry, 1982). From an experimental viewpoint, behavior is thought to be comprised of units called responses which covary with effective controlling variables called stimuli. The functional analysis of behavior is concerned with the relationship between stimuli, behavior, and the consequences of the behavior in the environment. Behavioral responses may be divided into two categories, respondent (elicited) and operant (emitted) as summarized in Table V.1. In addition, responses may also be unlearned (unconditioned) or learned (conditioned). Responses show specific physical properties, such as topography and effect, including that which can be measured (i.e., rate, force, or latency). These measurements are the dependent variables studied in behavioral experiments.

Respondents are elicited by a known environmental stimulus that usually has a specific temporal relationship to the occurrence of the response. The frequency of the response is dependent upon the eliciting stimulus. Examples of unconditioned respondent behaviors includes kineses, taxes, reflexes, and species-specific behaviors. Unlearned operant or emitted responses are not elicited by a single, identifiable, temporally cued stimulus in the environment. These responses occur within the context of many environmental stimuli, but there is no single eliciting stimulus, as in the case of respondent behaviors. Horizontally directed exploratory motor activity is an example of an unconditioned operant response.
Table V.1. Classification of behavior.

A. Unconditioned (unlearned) behavior

1. Respondent
   a. Elicited by known, observable stimulus
   b. Responses typically include those of smooth muscles, glandular secretions, autonomic responses, environment elicited effector responses
   c. Data are measures of response magnitude, probability, latency, or related to intensity of eliciting stimulus
   d. Taxonomy of respondents include the following:
      (1) Kinesis, environment-directed, and movement is random
      (2) Taxis, stimulus-directed, and movement is specific response of whole organism
      (3) Reflex, object-directed, and movement involves specific effect on the system
      (4) Species-specific, stimulus-specific, and movements are sequences of behaviors (fixed-action patterns)

2. Operant
   a. Emitted, with no known, observable eliciting stimulus
   b. Responses typically include those mediated by CNS, such as skeletal muscular movements that operate on and change the environment
   c. Data are measures of response probability or frequency

B. Conditioned behavior

1. Classically conditioned (respondent or type S learning): Response (CR) is elicited by a new stimulus (CS) as the result of close temporal pairing of that stimulus (CS) with another stimulus (US), which originally elicited the response (UR)

\[
\begin{align*}
\text{US} & \rightarrow \text{UR} \\
\text{CS} & \rightarrow \text{CR}
\end{align*}
\]

2. Instrumentally conditioned (operant or type R learning): Response (R) changes in frequency of occurrence as a function of the response consequence (S^R)
Most behaviors that are evaluated by researchers in pharmacology and toxicology studies are modified by a learning process. Behaviorists have identified that respondent (classical) and/or operant (instrumental) learning may take place during a study thereby exerting an effect on the study outcome.

Respondent or classical conditioning refers to a set of operational testing procedures in which there is the approximate simultaneous presentation of two stimuli. The original stimuli belong to a genetically determined stimulus-response relationship such as a shock induced withdrawal reflex, whereas the other stimuli may be neutral or unrelated. With repeated pairing of the two stimuli, there is an increase in the strength of another reflex, the conditioned reflex, that is composed of a response resembling the response from the original reflex and is elicited by the originally neutral stimulus. An example of a classically conditioned response is the conditioned withdrawal reflex. If a brief shock is applied to one limb of a restrained animal through electrodes, a reflexive withdrawal response is elicited. If the onset of a light repeatedly precedes the presentation of the shock, then the light eventually comes to elicit a conditioned escape response (limb withdrawal).

Operant or instrumental conditioning involves the pairing of a response with a stimulus. When the occurrence of an operant response is followed by the presentation of a reinforcing stimulus, the probability of recurrence of the response increases. A reinforcer stimulus is any stimulus that increases the probability of a response. An example of an instrumentally conditioned response is a lever press by a food-deprived rat for a reinforcement of a drop of milk. By the process of operant conditioning, a relatively low probability response such as pressing a lever increases in frequency following the presentation of the milk reinforcer.

B. CLASSIFICATION METHODS OF BEHAVIORAL TESTING

One method of classifying behavioral testing procedures is according to their desired use in the experimental design (Tilson and Mitchell, 1984). Behavioral testing techniques that are intended to measure the presence or absence of an effect are usually different from those used to assess the degree of toxicity or the lowest exposure level that is required to produce a behavioral effect. Screening procedures that typically permit the testing of large numbers of animals may not require extensive training of the test subjects or technical staff and are frequently simple to perform. However, these techniques may be labor intensive, provide subjective (unautomated) measures, yield quantal data, and may not be as sensitive to subtle effects as other tests.
Highly specialized test methods are more sensitive and are usually employed in studies concerning mechanisms of action or the estimation of the least effective dose. Such testing procedures are sometimes called secondary tests. These behavioral testing procedures may require special equipment, pretraining of experimental animals, and the use of motivational factors such as mild electric footshock or food deprivation. Secondary test procedures are frequently automated and generate graded or continuous data that are amenable to repetitive measurements, consistent experimental designs, and parametric data analyses.

C. TESTING FOR SIGNS OF EXPOSURE

There are many classes of chemicals that are widely used in a variety of commercial and industrial processes that affect the nervous system. The chemical structures of these known neurotoxicants can differ greatly. All structural components of the nervous system may be affected, although certain anatomical areas or cell populations may be more sensitive than others. Some agents act directly on the nervous system, whereas others affect nonneuronal processes which indirectly affect the nervous system. In this way chemicals may produce a wide range of possible neurotoxic effects that include motor, sensory, cognitive affective/personality, and general changes (Anger, 1984; Damstra, 1978; Weiss, 1985). Table V.2 summarizes behavioral and neurologic signs observed in humans that may be measured in animals. Both pharmacologists and toxicologists tend to select behavioral testing procedures that measure some aspect of these behavioral and neurological functions.

D. BEHAVIORAL TESTING SCHEMAS

Behavioral testing procedures are applied differently in pharmacology and toxicology testing protocols depending upon the experimental design. At the initial stages of toxicologic assessment, tests requiring the successful integration of intact subsystems may be appropriate (Butcher, 1976). These tests are frequently referred to as apical tests. Measuring performance on an operant schedule of reinforcement is an example of an apical test. An operant testing procedure typically uses intermittent reinforcement of a defined response and establishes a dependency between the occurrence of a specific response such as a lever press and the presentation of a specific stimulus such as food. Deficits in operant responding produced by exposure to a chemical may be due to alterations in any one or more neurobehavioral functions (i.e., sensory, motor, motivational, associative). Cabe and Eckerman (1982) have characterized that the effects of chemicals on learning is "apical" in nature. That is, an effect on learning might account for observed alterations on behavior mediated by nonassociative processes.
Table V.2. Some behavioral endpoints quantifiable in animals.

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motor</td>
<td>Activity Changes, Incoordination, Weakness and Paralysis, Abnormal Movement and Posture, Tremor, On-going Performance</td>
</tr>
<tr>
<td>Sensory</td>
<td>Primary Sensory Deficits (auditory, gustatory, olfactory, visual, somatosensory), Pain, Equilibrium Disorders</td>
</tr>
<tr>
<td>Arousal or Reactivity</td>
<td>Increased irritability or reactivity; change in CNS excitability</td>
</tr>
<tr>
<td>Cognitive</td>
<td>Nonassociative Learning, Associative Learning</td>
</tr>
<tr>
<td>General</td>
<td>Performance Changes, Reproductive Behavior, Consummatory</td>
</tr>
</tbody>
</table>
Following detection of a neurotoxic effect in a pharmacology or toxicology test screen, more sensitive and selective tests are generally used to characterize the nature of the neurotoxic effect. Behavioral procedures are most often used in a battery of tests to assess chemical-induced alterations in neurological functioning. Both general observation and the recording of neurotoxic signs have been commonly used in pharmacology. For example, Fühner (1932) listed many neurobehavioral effects in the description of frogs and mice exposed to various substances. These included insecure gait, vocalization, myoclonic bursts, and numbness.

Neurobehavioral checklists have been incorporated into acute toxicity testing for many years (Balazs, 1970; Boyd, 1959; Zbinden, 1963). Numerous behavioral test batteries have been proposed to provide a structured framework for setting testing priorities. Irwin (1968) proposed a battery of tests designed to detect chemical-induced alterations in sensorimotor function for the assessment of psychoactive, neurological, and autonomic signs. Marshall and Teitelbaum (1974) described a battery of tests to assess toxicant-induced alteration in vision, audition, pain perception, and olfaction. Evans and Weiss (1978) describe a multi-tiered approach initially consisting of observational assessments of toxicity such as ratings of locomotor impairment, the presence or absence of tremor, ptosis, and convulsions, alterations in various reflexes and autonomic dysfunction. The preliminary tier of tests were followed by a secondary tier of tests to assess specific sensory and motor functions. Gad (1982) has also proposed a series of tests utilizing simple measurements of sensorimotor function. Alder and Zbinden (1983) have recently proposed a neurobehavioral checklist for use in toxicological studies in rats. Their battery of tests are similar to those suggested by Irwin (1968) for the mouse. The Alder and Zbinden checklist includes a series of tests that were selected for use in acute and repeated-dose toxicity experiments and designed for cage-side observations during the course of exposure.

Pavlenko (1975) proposed three phases in a battery of testing. Simple methods are used initially to assess changes in orientation, aggression, and other reflexes. The second phase is intended to determine the threshold and subthreshold amounts of chemicals that are necessary to affect higher order nervous system function such as conditioned reflexes. In the last phase of assessment, functional stress tests are recommended to determine minor or latent compensatory alterations and to study mechanisms of toxicity.

Reiter and colleagues (1981) proposed the development of a behavioral toxicity index based upon the acute LD₅₀ and experimentally derived ED₅₀ values for a variety of behavioral tests such as motor activity in a figure-eight maze, schedule-controlled operant behavior, conditioned taste aversion, and
activity in a radial arm maze. Mitchell and Tilson (1982) proposed a battery of tests chosen to evaluate a wide range of neurobehavioral functions from simple reflexes to more complex processes such as sensory and motor function, changes in reactivity and associative processes. The Mitchell and Tilson test battery was used by Pryor et al. (1983) to compare the dose-related effects of eight neurotoxicants administered to rats in a chronic toxicity test protocol.

There is a general agreement that behavioral endpoints, unconditioned or conditioned respondents or operants, can be used to detect the presence of neurotoxicity in a sequential testing format. At the screening level, apical tests are useful as early indicators of neural dysfunction, while specialized tests are used to determine the nature of the observed effects as they relate to the level of the exposure at which they occur.

EXAMPLES OF NEUROBEHAVIORAL TESTS

A. TESTS OF MOTOR FUNCTION

1. Spontaneous motor activity

Reiter and MacPhail (1979) have noted in their review that spontaneous motor activity behavior has been used extensively in behavioral toxicology and pharmacology, both as a measure of motor dysfunction and as an apical test. Movement within the living space or environment occurs at a relatively high frequency and appears to be sensitive to the effects of chemicals. In the rodent, motor activity is a complex behavior consisting of a variety of motor acts, such as sniffing, grooming, rearing, and ambulation. Since activity is not a singular measure, changes in the frequency of this behavior could reflect toxicant-induced changes in any one or more sensorimotor functions, arousal, or motivational states.

Although there are many devices that have been designed to measure activity (Reiter and MacPhail, 1979), the figure-eight maze has been used extensively and proven to be an effective measure of toxicant-induced changes in behavior. This apparatus is typically used to measure effects of a chemical during a pre-determined period of time (minutes or hours). It can also be used to measure diurnal activity changes (Reiter, 1983). Reiter et al. (1981) have compared data from the figure-eight maze to other behavioral indicators of toxicity and found that the dose-effect curves compare favorably. This is true for other tests of motor function including schedule-controlled behavior, conditioned taste aversions, and activity in a radial arm maze.

Other investigators have used computer-assisted techniques to continuously monitor spontaneous locomotor patterns of activity. Elsner et al. (1979) reported that methyl mercury
lowered the activity of rats during the night portion of the diurnal cycle. The importance of naturally occurring biorhythm cycles to the expression and measurement of motor activity has been observed by many investigators. For example, Ruppert et al. (1982) measured the activity of adult rats exposed to various doses of trimethyl tin in the figure-eight maze at hourly intervals for 23 hours. After an acute exposure on days 49-51, those animals that received trimethyl tin showed a period of greater motor activity during the night portion of the cycle. Ruppert noted that these animals were markedly hyperactive during all phases of the diurnal cycle. The diurnal cycle may mask or possibly reveal the presence of toxicity. This was demonstrated in another experiment where the offspring of pregnant mice were exposed to 3,4,3',4'-tetrachlorobiphenyl during gestation. These offspring were observed to be hyperactive and exhibit a circling or spinning syndrome (Tilson et al., 1979). Analysis of the motor activity of these mice in their home cage at days 35 and 65 revealed that the affected mice were hyperactive during the nocturnal portion of the light cycle, but exhibited little or no hyperactivity during the normally quiescent day portion of the cycle.

2. Test of motor coordination

Many procedures have been used to assess chemical-induced alterations in motor coordination, including negative geotaxis (Pryor et al., 1983), rope climbing (Carlini et al., 1967), and performance on an inclined plane (Graham et al., 1957). One of the more frequently used tests to quantify motor dysfunction in rodents is to measure balance performance on a rotating rod. The rotating rod testing procedures are used widely in pharmacology and are also regarded by many investigators as a convenient technique in toxicological experiments (Alder and Zbinden, 1973).

Bogo et al. (1981) compared the accelerod and rotarod testing procedures and reviewed the primary experimental variables that affect the reliability of the data used to assess chemical agents (i.e., ethanol and acrylamide) for pharmacological and toxicological effects. They noted that the cumulative effect of rats treated with acrylamide was marked by a significant decrement in rotarod performance. Rats receiving larger doses of acrylamide lost capability to perform on the rotating rod faster than those receiving the lower dose. They concluded that the accelerod technique showed a greater sensitivity in detecting neurotoxic effects as this design provides a continuous measure of the upper limit of performance rather than the quantal or arbitrary measures obtained from the rotarod.
3. Tests for weakness and paralysis

One of the early signs of exposure to many neurotoxicants and psychopharmacological agents is muscle weakness or decrements in grip strength. Various tests have been devised to assess this effect including swimming performance (Bhagat and Wheeler, 1973; Klaus and Erdmann, 1978; Kniazuk and Molitor, 1944) and suspension from a horizontal rod (Molinengo and Orsetti, 1976). Meyer et al. (1979) described a procedure that uses mechanical strain gauges to measure strength in the fore- and hindlimbs of both rats and mice. Using this procedure, Pryor et al. (1983) measured the effects of eight neurotoxicants during 120 days of repeated dosing. The fore- and hindlimb test appears to be sensitive to chemical agents that produce axonopathies or act as muscle relaxants. For example, Pryor et al. (1983) found that repeated dosing with acrylamide produced dose-related decreases in grip strength in rodents. Hindlimb grip strength was affected to a greater extent than forelimb grip strength. Both the experimental findings of muscle weakness and decreased grip strength correlate with epidemiological evidence in humans occupationally or abusively exposed to acrylamide.

4. Measurements of abnormal movement and posture

Exposure to neurotoxicants may result in abnormal posture and gait. Jolicoeur et al. (1979) have summarized several techniques devised to quantify these neurological signs. They described a battery of tests which included tests for locomotor activity, assessment of catalepsy and rigidity, hindlimb splay, analysis of gait, and reflex analysis. Of the tests described, two tests appeared to be relatively sensitive indicators of the effects of chemical agents (e.g., acrylamide and 3-acetylpypiridine) known to produce peripheral neuropathy or ataxia.

Toxicant-induced hindlimb splay and ataxia was measured by staining the hindpaws of test animals with ink and dropping them at a constant height onto absorbent paper to measure the distance between specific digits of the hindlimbs. Another method is to quantify gait characteristics (stride width, length, and angle between steps) as rats walked across paper (Lee and Peters, 1976). Edwards and Parker (1977) have also used hindlimb splay as a measure of acrylamide-induced neurotoxicity, while Schallert et al. (1978) used gait analysis to measure motor disturbances produced by central administration of 6-hydroxydopamine.
5. Measures of tremor

Many neurodegenerative diseases, psychopharmacological agents, and neurotoxicants are characterized by signs of tremor (Stein and Lee, 1981). Several procedures have been developed to quantify tremor following exposure to environmental agents (Gerhart et al., 1985). In animals, tremor and stereotypic behaviors are frequently assessed using qualitative or semiquantitative rating scales. In some cases, attempts have been made to automate tremor produced by chemicals using transducers which generate a signal that can be analyzed using routine statistical procedures. Gerhart et al. (1985) described a procedure in which a tremulous rat is placed on a platform that is attached to a load cell. In this procedure, the analog output generated by whole body tremor is quantified by a spectral analyzer which is capable of performing a Fourier analysis of the data. Using such a procedure, Gerhart and colleagues differentiated the power spectra generated by pharmacological agents (e.g., oxotremorine and harmine) and tremorgenic insecticides (e.g., chlordecone). Recently, this procedure was used to determine the dose and time-related effects of chlordecone (Gerhart et al., 1985) and other related insecticides (Tilson et al., 1985). It has also been used to investigate the neuropharmacological mechanism of action of these chemicals.

6. On-going performance

Classical operant techniques, available from behavioral psychology, have been adapted to assess the effects of chemicals on changes in on-going performance of fine motor function in rats. For example, Falk (1970) used a lever press technique in which the rats were required to continuously hold a force-transducer lever within the limits of a 15-20 gram force band for 1.5 seconds in order to obtain food reinforcement. This procedure has been used to study fine motor control in animals that have become dependent upon ethanol (Samson and Falk, 1974).

B. BEHAVIORAL TESTS FOR ALTERATIONS IN SENSORY PROCESSES

Alterations in sensory processes, (e.g., paresthesias, visual or auditory impairments) are most frequently among the initial signs of toxicity reported by humans exposed to toxicants. The detection of sensory deficits (i.e., touch, sight, sound, taste, or smell) are an important element in preclinical assessments of neurotoxicity. Several attempts have been made to develop objective tests for sensory dysfunction in laboratory animals which include screening tests, reflex modification, and procedures based on discriminated instrumental conditioning.
1. Sensory screening test procedures

Several testing procedures have been devised to screen laboratory animals for sensory deficits. For example, the "visual cliff" procedure has been used to assess depth perception (Langman et al., 1975). This procedure measures the animals ability to choose between stepping onto a nearby platform or floor ("shallow" floor) as compared to one perceived to be farther away ("deep" floor). Another test of visual function is the optokinetic drum, which relies on the optokinetic nystagmus or optomotor response. Although this response is believed to be a measure of visual acuity (Wallman, 1975), its validity in neurotoxicologic studies has not been adequately demonstrated.

Other screening tests rely on orientation or other responses to the presentation of a stimulus. The acoustic startle response, for example, has been used to study the ototoxic effects of antibiotics (Harpur, 1974). Although the startle reflex consists of sensory and motor components, it is useful in testing the responsiveness of auditory functioning. Additional testing is needed to ascertain specificity of auditory sensory effects.

Several screening procedures that have been adopted from psychopharmacology are used to measure toxicant-induced changes in touch reactivity to noxious stimulation. Examples include both the flinch-jump technique (Evans, 1961) and the hot-plate procedure (Pryor et al., 1983). Walsh et al. (1984a) used the hot-plate procedure to study the relative toxicity of four organometal compounds. In these experiments, changes in the responsiveness to a 55°C hot-plate were measured in rats given a single dose of triethyl or trimethyl tin or lead. Both a dose- and time-related decrease in sensitivity (i.e., increased latencies to respond) to the thermal stimulus was observed in this testing procedure for those animals treated with these chemicals.

2. Reflex modification

Reflex modification procedures have been used to assess sensory dysfunction. These testing procedures are frequently termed the prepulse inhibition paradigm. In reflex modification, the presentation of an irrelevant stimulus prior to another stimulus that elicits a reflex can, under some circumstances, inhibit that reflex (Hoffman and Ison, 1980). It is important to note that stimuli capable of inhibiting the reflex are near the threshold of response and changes in intensity of the stimuli have marked effects upon reflex modification. The reflex modification observed by the presentation of the irrelevant stimulus does not depend upon prior association with the eliciting stimulus and appears to
be mediated at some level within the brainstem. This procedure has been used to evaluate auditory processing (Marsh et al., 1978) and in toxicology and pharmacology to assess sensory alterations. Young and Fechter (1983) demonstrated that neomycin, an aminoglycoside antibiotic, resulted in a shift in the auditory threshold as measured by the prepulse inhibition procedure. Fechter and Young (1983) also used the prepulse paradigm to study the effects of rats treated with triethyl tin. By determining the intensity of the pure tone necessary to produce a 15% inhibition of the reflex, they noted that there was no shift in the detection of thresholds for a 40 kHz tone administered during or after exposure to triethyl tin; while there was a highly significant loss of auditory acuity following neomycin exposure. In spite of the lack of any effect of triethyl tin on auditory acuity, tin-induced neuromuscular deficits such as hindlimb splay and weakness were observed. Wecker and Ison (1984) used the prepulse inhibition paradigm to show that alcohol does not affect loudness perception, but may disrupt the temporal relationships within the primary auditory pathway.

3. Discriminated instrumental conditioning

Mazes and similar types of apparatus have been used to test for alteration in the performance of tasks based upon discrimination of sensory cues (Winneke et al., 1977; Zenick et al., 1978). Zenick et al. (1978) trained lead exposed neonatal rats to use visual discrimination cues for escape from a T-maze containing water. Lead-exposed rats showed an increased number of errors when brightness (black versus white arms of the maze) or shapes (circles versus triangles) were used as cues to mark the position of the arm containing an escape platform. Lead-induced alterations in the motoric ability of the rats to perform the task were not evident.

Pryor et al. (1983) utilized a multisensory conditioned avoidance response to measure three sensory modalities concurrently in the same test animal. In this testing procedure, rats were initially trained to climb or pull a rope to escape and then to avoid a noxious footshock applied to the grids of the floor of the test chamber. By a process of training, the conditioned response is brought under the control of a tone (4 kHz), a low-intensity nonaversive current on the floor (.125 mA), and a change in the intensity of the house light in the chamber. A quasipsychophysical curve was established for each modality before exposure to various toxic agents. In this way, the recorded test findings provided measurements for interactive modulations to auditory, visual, and tactile sensory processes.

The most commonly used instrumental procedures are those derived from operant behavioral psychology. In such studies, animals are motivated by food or some other reinforcer to make a response only in the presence of specific stimuli. A graded
stimulus-intensity response function curve can be generated by varying some parameter of the stimulus (e.g., intensity). Toxicant-induced effects on visual (Merigan, 1979) and auditory (Chiba and Ando, 1976; Stebbins and Moody, 1979) response and reactivity to electric shock (Tilson and Burne, 1981; Weiss and Laties, 1961) have been studied using operant procedures.

An example of the use of operant procedures to study toxicant-induced changes in sensory thresholds is the acrylamide somatosensory threshold study in monkeys by Maurissen et al. (1983). These investigators trained monkeys to make a response with one hand whenever a vibratory or small electrical stimulus was applied to the fingertip of the other hand. Marked changes in vibration sensitivity were noted during the course of repeated exposure to acrylamide, but there were no effects on sensitivity to the electric shock. These studies are important in that they demonstrate, in a laboratory animal species, the same characteristic sensory deficits (loss of vibration sense) that can be seen in humans exposed to neurotoxicants such as acrylamide.

C. TESTS FOR AROUSAL OR REACTIVITY

Nervousness, irritability, emotionality, and altered reactivity to environmental stimulation are important indicators of neurotoxicity. Laboratory testing procedures have been developed to access alterations in the arousal and reactivity responsiveness of laboratory animals.

1. Startle reflex

One method used by neuroscientists to quantify changes in responsiveness to external stimulation such as noise or movement is to measure the acoustic startle reflex. In this test procedure, the laboratory animal is subjected to a startle stimulus (i.e., a loud noise or sudden movement) and the distance traversed by the animal in response to this stimulus is recorded. The effect of various chemicals can be compared. For example, Crofton and Reiter (1984) studied the effects of deltamethrin and cismethrin, Type II and Type I pyrethrins, respectively, on the acoustic startle response in rats. These investigators found that the effects of these two agents on startle responsiveness were different (i.e., cismethrin increased, while deltamethrin decreased, acoustic startle responsiveness). They also studied the effect of these two agents on the augmentation of the startle reflex as a function of background noise (i.e., sensitization). If sensitization of the startle response is calculated as the difference in response amplitude between 80 and 50 dB background noise, the two pyrethrins had significantly different effects on this behavioral process.
2. Change in central nervous system excitability

Exposure to toxicants can frequently alter the excitability of the nervous system. This can be quantified by measuring changes in seizure thresholds.

Fox et al. (1979) have demonstrated that changes can be observed in the threshold level needed to produce maximal electroshock seizure (MES) in rats exposed to lead during development. Dyer et al. (1982) have reported that rats exposed to trimethyl tin were more sensitive to the seizure induced effects of metrazol. This suggests that a general increase in seizure susceptibility can occur if there is an additive effect between two chemicals.

Another procedure that has been used to study changes in nervous system excitability is kindling behavior. This type of behavior has been used by many investigators to study temporal lobe epilepsy (Joy et al., 1982). Some toxicants, such as lindane and dieldrin, have been shown to be proconvulsants, that is, repeated exposure to low, nonconvulsive doses of lindane, decreases the number of amygdaloid stimulations necessary to produce kindling (Joy et al., 1982).

3. Emotionality

Procedures ostensibly measuring emotionality of animals exposed to toxicants are many and varied (Alder and Zbinden, 1983; Archer, 1973). As pointed out by Alder and Zbinden (1983), none of these techniques have been properly validated for use in toxicology, and there is some question as to the actual "emotion" that is measured by such procedures (Barnett and Cowan, 1976).

D. TESTS FOR LEARNING AND MEMORY

Behavioral toxicologists and pharmacologists have employed a wide variety of tests to assess chemical effects on both associative and cognitive function in laboratory animals (Heise, 1984). The effects of chemicals on learning and memory must be inferred from the change in behavior following exposure and compared relative to that prior to exposure. Acquisition may be defined as an enduring change in behavior, while memory can be defined as the preservation of the learned behavior over time. Measurement of changes in learning and memory must be separated from other changes in performance that involve nonassociative processes. Before the investigator can record that changes in learning and memory have been affected by the experimental treatment, other factors such as motivation, sensory, or motor changes have to be experimentally controlled. In addition, any apparent toxicant-induced changes in learning and memory should be demonstrated over a range of stimulus and response conditions.
1. **Nonassociative learning**

The most simple form of learning is habituation, which can be measured as a gradual decrease in the magnitude or frequency of a response following repeated presentations of a stimulus. Habituation is characterized by: a) the recovery of a response over time, if the stimulus is withheld; b) the potentiation across repeated habituation-recovery cycles; c) a correlation of the stimulus repetition rate and speed of habituation; d) an inverse relation between habituation speed and stimulus strength; e) a stimulus generalization of habituation; and f) the habituation of the response to a dishabituating stimulus (Thompson and Spencer, 1966).

Overstreet (1977) demonstrated that alterations in habituation can be chemically induced. He conducted studies with rats exposed to diisopropylfluorophosphate (DFP), an irreversible inhibitor of cholinesterase. The rats were placed in a small cage to which a small strain gauge was attached. A 1 kHz, 100 dB tone was presented for 1 sec for 42 trials per session, each with trial separated by 90 seconds. The response of each test animal was determined prior to, as well as after, drug administration. In some experiments, multiple sessions were run using the same animals. Exposure to DFP had no apparent effect on the magnitude of the initial startle response, but DFP-exposed rats showed a decreased habituation in that they responded more than controls on subsequent trials. These investigators also showed that physostigmine, a reversible inhibitor of cholinesterase, had a similar effect. These studies suggest that the cholinergic system may be involved in the modulation of habituation processes.

2. **Tests for associative learning using classical conditioning procedures**

Most testing procedures that are used to study associative learning have been adapted from the classical conditioning studies of psychology. One classically conditioned response used in behavioral pharmacology and toxicology is the conditioned nictitating membrane response in the rabbit. In this procedure, a restrained rabbit is presented with a mild electric shock (unconditioned stimulus) to the skin of the paraorbital region after presentation of a tone or light conditional stimulus. The unconditioned response, movement of the nictitating membrane, eventually occurs following presentation of the conditional stimulus alone. Yokel (1983) dosed rabbits repeatedly with aluminum and found that treated rabbits learned the conditioned response at a lower rate than controls. One important feature of this experiment was the fact that neither the baseline rates of the nictitating membrane extension (unconditioned stimulus-response reflex) was affected prior
to conditioning nor was the sensitivity to the electric shock affected. These control manipulations tended to rule out aluminum-induced changes in sensory or motor function.

Another example of the classically conditioned response used in toxicology is the conditioned taste aversion. If a rat ingests a novel substance and becomes ill, it will tend to ingest less of that substance in the future. This form of learning consists of pairing a novel stimulus, such as a sweet taste, with a toxic effect (the unconditioned stimulus). A single pairing of the conditioned with the unconditioned stimuli frequently is sufficient to produce the conditioned response (i.e., long-lasting aversion of the conditioned stimulus). This conditioned flavor aversion testing procedure has been suggested by several investigators as a screen for toxicity (MacPhail, 1982; Riley and Tuck, 1985). In their review on the use of flavor aversion testing procedures, Riley and Tuck (1985) conclude that most known toxins produce conditioned taste aversions. However, their conclusion remains controversial as there may be a possibility for a high incidence of false positives in toxicology assessments. For example, many psychoactive drugs that are not considered to be toxicants are known to produce flavor aversions (Riley and Tuck, 1985).

The conditioned suppression method involves the presentation of a stimulus such as a mild electric shock which is intended to disrupt or suppress ongoing behavior. Disruption of responding might be regarded as an unconditioned response and pairing a previously neutral stimulus with the shock eventually elicits pausing by a process of classical conditioning. Although the procedure seems to be useful in assessing the effects of chemicals on acquisition and/or retention of classical conditioning, it has not been used in this context. In behavioral pharmacology, it has been used to evaluate the psychoactive properties of anxiolytics (Cook and Davidson, 1973) and in toxicology to assess the effects of chemicals on sensory processing (Chiba and Ando, 1976).

3. Instrumental conditioning

As discussed in a previous section, instrumental or operant learning involves pairing of a response with a reinforcing event; learning involves making a response to obtain positive reinforcement such as food or the removal or termination of an aversive stimuli such as electric footshock. Table V.3 summarizes the distinction between positive and negative reinforcers and the experimental operation of positive and negative reinforcement.
Table V.3. Effects of presentation or withdrawal of a positive or negative reinforcer on behavior.

<table>
<thead>
<tr>
<th>Type of Reinforcer</th>
<th>Presentation</th>
<th>Withdrawal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Positive Reinforcement: Presentation of reinforcer follows the response which increases in probability.</td>
<td>Punishment: Removal of reinforcer follows the response which decreases in probability.</td>
</tr>
<tr>
<td>Negative</td>
<td>Punishment: Presentation of reinforcer follows the response which decreases in probability.</td>
<td>Negative Reinforcement: Removal of reinforcer follows the response which increases in probability.</td>
</tr>
</tbody>
</table>

a. Procedures using negative reinforcement

(i) Passive avoidance

In the passive avoidance testing procedure, the experimental animal initially receives one or more training trials in which it is placed on a raised platform or in a lighted chamber. Within a period of time, the animal will step down from the platform or leave the lighted chamber to enter another chamber at which time a noxious stimulus is presented. At some later time, the animal is replaced upon the platform or in the lighted chamber and retested to make a comparison with the initial training session. A longer step-down or step-through latency on the second test trial is taken as an indication of learning. Walsh et al. (1982a) dosed rats with trimethyl tin and 21 days later trained them in a passive avoidance step-through procedure. The tin-exposed rats showed significant impairment of retention when tested 24 hours later. This observation is significant in that trimethyl tin affects hippocampal morphology and produced an effect on passive avoidance similar to that observed after electrolytic lesioning of that area (O'Keefe and Nadel, 1978).
A significant improvement upon the typical passive avoidance paradigm was developed by Maclutus et al. (1982), who reported a multiple measure step-through passive avoidance procedure to determine retention deficits in animals that had been exposed neonatally to chlordecone. By using several measures of responding (i.e., initial step-through latencies, frequency of vacillatory behaviors, or the usual latency to reenter the chamber after being shocked), it is possible to gain some understanding for the role of emotional or reactive influences in the mediation of the passive avoidance response.

(ii) Active avoidance

In contrast to passive avoidance tasks in which an animal withholds a response to avoid presentation of a negative reinforcer, active avoidance tasks require that the animal perform a specific response to avoid negative reinforcement. One-way shock avoidance tasks require that the animal move unidirectionally from one chamber to another to avoid or escape negative reinforcement. The impending onset of shock is signaled by a conditional stimulus which is extinguished if a conditioned response (e.g., movement from one compartment to another) is made by the test animal. If the correct response is not made, then a shock is presented to initiate the expected conditioned response. Once an escape or avoidance is made, the animal is replaced into the original chamber and the process is repeated. The one-way procedure can be differentiated from the two-way shuttle box in that the animal learns to shuttle from one compartment to another in order to escape or avoid negative reinforcement. Unlike one-way avoidance, the animals must learn to return to a compartment where they have been negatively reinforced. Sobotka et al. (1975) reported that rats exposed neonatally to lead acetate performed as well as controls on a one-way shock avoidance task, but displayed significant deficits when required to learn a two-way task.

Acquisition of avoidance responding is dependent upon many variables present in the test environment such as shock intensity, the characteristics of the conditional stimuli, and the configuration of the test chamber. For example, Tilson et al. (1982) varied the sizes of the two testing compartments used in the conditioning of a two-way shuttle box response in rats previously exposed to triethyl lead. In this experiment, the sizes of the two compartments were different in that one compartment was larger than the other. Rats having received triethyl lead performed significantly better over a 60 trial acquisition session. Subsequent analysis found that the lead-exposed rats adapted a strategy of retracing. Before the start of the next trial, these rats were observed to retrace their movements starting from the smaller box, avoiding to the larger box, and then retracing back to the smaller box. This observation suggests that nonassociative factors may have contributed
to the observed enhanced rate of acquisition in the lead-exposed rats. However, a subsequent experiment indicated that triethyl lead did not change sensitivity to the electric footshock as measured by a flinch-jump procedure.

As in the case of an impaired passive avoidance response, facilitation of a two-way shuttle response is sometimes taken as an indication of brain damage. The facilitatory effects of septal and hippocampal lesions on active avoidance have been described by King (1958).

The symmetrical Y-maze is a more complex learning task than either the one- or two-way shuttle avoidance procedure. In this procedure, a conditional stimulus is presented in one of two arms of the maze which are unoccupied by the study animal. If the animal does not enter the arm with the conditioned stimulus, negative reinforcement is presented. The animal can terminate the negative reinforcement by entering the cued arm. Unlike the shuttle box, the animal in the Y-maze must learn when, as well as where, to go to avoid negative reinforcement. The Y-maze has been applied to both behavioral pharmacology (Ray and Barrett, 1975) and toxicology (Vorhees, 1974).

b. Procedures using positive reinforcement

(i) Mazes

A variety of maze designs have been developed to study the effects of chemicals on learning. The Hebb-Williams maze consists of a series of problems (usually a 12-problem sequence), which the animal must learn in order to receive positive reinforcement (Hebb and Williams, 1946). Swartzwelder et al. (1982) dosed rats with trimethyl tin and trained them to perform for food reinforcement in the Hebb-Williams maze. They observed significant deficits in maze performance (i.e., increases in number of total and perseverative errors) which were associated with marked thinning of the pyramidal cell fields in the hippocampus.

The radial arm maze (RAM) is frequently used in behavioral toxicology. The RAM is a spatial learning task in which animals are required to recall the location of previously entered and nonentered feeding sites during a free-choice test session (Olton et al., 1979). In this test, the most effective response strategy is to not enter those arms of the maze from which the food has been removed during a previous entry or in which food has never been present. The maze usually consists of a circular starting arena from which arms radiate outward. Performance in this maze is thought to demonstrate the existence of both working and reference memory components. Walsh et al. (1982b) reported that trimethyl tin-exposed rats were impaired in the performance of this task and that this effect was associated
with damage to the hippocampus. Recently, Walsh et al. (1984b) reported that bilateral administration of AF64A, a cholinergic neurotoxin, into the lateral cerebroventricles impaired RAM performance and that this effect was associated with cholinotoxicity in the hippocampus.

(ii) Operant discrimination procedures

(a) Repeated acquisition

The procedure known as repeated acquisition of response chains (Thompson, 1973) requires that a study animal must solve a series of problems which may vary from session to session by responding in a particular order. The procedure generates a pattern of within sessions acquisition which can remain stable for relatively long periods of time. For example, Schrot et al. (1984) trained rats on a repeated acquisition of behavioral chains procedure in which food reinforcement depended on the correct completion of a four member response sequence on three separate response levers. The animals were exposed to carbon monoxide periodically during the tests. These investigators found that carbon monoxide exposure resulted in pausing following completion of a sequence, but did not affect accuracy. They concluded that the exposure to carbon monoxide appeared to disrupt baseline performance. These findings correlate with similar effects of carbon monoxide on other schedules of reinforcement. Other experiments have shown that pharmacological agents and microwave exposure can disrupt the accuracy of responding on the repeated acquisition paradigm (Schrot and Thomas, 1983; Schrot et al., 1980).

(b) Matched-to-sample (MTS)

This procedure involves the presentation of a sample stimulus for a brief time. After some delay the subject must identify the sample from among one or more comparison stimuli. Retention is indicated by the percentage of correct choices as a function of the delay interval (D'Amato, 1973; Wasserman, 1976). The MTS procedure has been regarded as a model of short-term memory (Hogan et al., 1981; Roberts and Grant, 1976), although others (Thompson, 1978) regard the procedure to reflect stimulus control by the sample stimulus. Variations on the MTS include the go/no-go MTS and the continuous nonmatching-to-sample MTS. In the go/no-go MTS testing procedure, a decision to make a response is made on the presentation of a sample stimulus (Konorski, 1957); whereas in the continuous nonmatching-to-sample MTS testing procedure (Pontecorvo, 1983), a variable number of trials with one stimulus alternates with a variable number of trials of a second stimulus and the initial response on a trial following a stimulus change is reinforced. The effects of various psychoactive drugs on MTS performance has been summarized by Heise (1984) and McMillan (1981).
E. SCHEDULE-CONTROLLED OPERANT BEHAVIOR

Responses are not usually reinforced on a one-to-one basis inside or outside the laboratory. Systematic arrangement of reinforcement contingencies following responses has evolved into four descriptive classes called schedules of reinforcement. The four types of schedules used to define the relationship between response and reinforcement are summarized in Table V.4. These include: simple, compound, complex, and higher order. A full description of the schedules has been described by Ferster and Skinner (1957).

Schedule-controlled operant responding has been frequently used in behavioral pharmacology and toxicology because it is sensitive to a wide range of chemicals, including methyl mercury (Armstrong et al., 1963; Evans et al., 1975; Laties and Evans, 1980), solvents (Colotla et al., 1979; Glowa, 1985), pesticides (Anger and Wilson, 1980; Bloom et al., 1983; Dietz and McMillan, 1979a,b; Leander and MacPhail, 1980; MacPhail, 1985), acrylamide (Daniel and Evans, 1985; Rafales et al., 1982; Tilson et al., 1980), carbon monoxide (Geller et al., 1979; McMillan and Miller, 1974), carbon disulfide (Levine, 1976), both organic (Tilson et al., 1982) and inorganic lead (Barthalmus et al., 1977; Cory-Slechta and Weiss, 1985; Rice, 1985), and triethyl and trimethyl tin (DeHaven et al., 1982; Swartzwelder et al., 1981; Tilson and Burne, 1981; Wenger et al., 1984). Schedule-controlled performance is useful in that the experimental animal frequently serves as its own control, and the procedure provides an opportunity to study a few animals extensively over relatively long periods of time. This testing strategy can be of value in determining the onset and recovery of an effect (Weiss et al., 1985). In addition, a testing procedure that permits a complex behavioral analysis may provide a more dependable basis for extrapolation to humans.

The multiple fixed ratio (FR), fixed interval (FI) schedule of reinforcement has been widely used in pharmacology and toxicology because it generates different rates and patterns of responding within the same animal during the same testing session. For instance, responding under the FR component of the schedule is typified by a relatively continuous rate of responding which is punctuated by short pauses following the reinforcement. Responding under the FI component is characterized by a low rate of responding after reward, which is followed by an increase in the response rate until the reinforcement has taken place. Pharmacological and toxicological agents have been shown to selectively affect both rates and patterns of responding. Swartzwelder et al. (1981) reported that a single dose of trimethyl tin can have long-lasting suppressant effects on FI schedules of reinforcement, while triethyl tin produces progressive decreases in the number of responses in rats trained on a FR or FI schedule of reinforcement (DeHaven et al., 1982). Wenger et al. (1984) found that mice
Table V.4. The four descriptive classes for reinforcement contingencies (schedules of reinforcement).

<table>
<thead>
<tr>
<th>SIMPLE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Nonintermittent</td>
<td></td>
</tr>
<tr>
<td>(1) Operant Level</td>
<td></td>
</tr>
<tr>
<td>(2) Continuous Reinforcement</td>
<td></td>
</tr>
<tr>
<td>(3) Extinction</td>
<td></td>
</tr>
<tr>
<td>B. Intermittent</td>
<td></td>
</tr>
<tr>
<td>(1) Ratio</td>
<td></td>
</tr>
<tr>
<td>a. Fixed Ratio</td>
<td></td>
</tr>
<tr>
<td>b. Variable Ratio</td>
<td></td>
</tr>
<tr>
<td>c. Progressive Ratio</td>
<td></td>
</tr>
<tr>
<td>(2) Interval</td>
<td></td>
</tr>
<tr>
<td>a. Fixed Interval</td>
<td></td>
</tr>
<tr>
<td>b. Variable Interval</td>
<td></td>
</tr>
<tr>
<td>(3) IRT (Interresponse Time)</td>
<td></td>
</tr>
<tr>
<td>a. IRT&gt;t sec</td>
<td></td>
</tr>
<tr>
<td>b. IRT&lt;t sec</td>
<td></td>
</tr>
<tr>
<td>c. IRT schedules with Limited Hold</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COMPOUND</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Sequential</td>
<td></td>
</tr>
<tr>
<td>(1) Chained</td>
<td></td>
</tr>
<tr>
<td>(2) Tandem</td>
<td></td>
</tr>
<tr>
<td>(3) Multiple</td>
<td></td>
</tr>
<tr>
<td>(4) Mixed</td>
<td></td>
</tr>
<tr>
<td>B. Concurrent</td>
<td></td>
</tr>
<tr>
<td>(1) Reversible</td>
<td></td>
</tr>
<tr>
<td>(2) Nonreversible</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>COMPLEX</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Fixed contingency</td>
<td></td>
</tr>
<tr>
<td>(1) Conjunctive</td>
<td></td>
</tr>
<tr>
<td>(2) Alternative</td>
<td></td>
</tr>
<tr>
<td>B. Schedules in which contingency varies with responding</td>
<td></td>
</tr>
<tr>
<td>(1) Adjusting</td>
<td></td>
</tr>
<tr>
<td>(2) Interlocking</td>
<td></td>
</tr>
</tbody>
</table>

| HIGHER ORDER SCHEDULES |       |
trained to respond under a multiple FR 30, FI 600 second schedule of milk reward showed disrupted responding for up to 6 to 7 weeks after a single injection of trimethyl tin. Trimethyl tin was found to disrupt both components of the multiple schedule. These investigators also observed acute effects of trimethyl tin on operant behavior and that animals sacrificed 51 hours after dosing showed significant neuronal necrosis in the fascia dentata of the hippocampus.

F. NATURALLY OCCURRING BEHAVIORS

The most frequent approach to study naturally occurring behaviors is to measure the repertoire of responses exhibited by animals in their homecage environment. Bushnell and Evans (1985) quantified the frequency and patterning of both motor activity and consummatory responses by a computer monitoring system. These investigators assessed diurnal patterns of feeding, drinking, locomotor activity, and rearing in rats for up to 2 weeks after a single dose of trimethyl tin. Immediately after dosing, both food and water consumption decreased and then increased and the diurnal patterns of drinking and rearing were also disrupted.

Silverman et al. (1981) recorded species-specific behaviors (e.g., exploration, sex-related activities, aggression, and submission) of rats in an observation cage and measured toxicant-induced changes in the frequency of these behaviors. They reported that a dietary exposure to methyl mercury significantly affected these behaviors.

The naturally occurring behavior, aggression, is frequently mentioned in toxicology studies as an anecdotal finding. Although aggression has operationally been defined as the actual inflicting of, or the threat to, inflict damage (Sheard, 1977), the quantification of this response can be somewhat subjective. Predatory aggression usually occurs without autonomic signs and is typically a spontaneous attack on some other species. Elicited aggression usually occurs with autonomic signs and is facilitated by environmental manipulations such as isolation or electric footshock. Dominance/submissive postures are frequently mentioned by investigators as indicators of aggression. Michalski (1974) has quantified this behavior in rats using a multi-channel event recorder.

G. SPECIAL APPLICATION OF BEHAVIORAL PROCEDURES

1. Pharmacological challenges

Pharmacological challenges are used with increasing frequency in neurobehavioral toxicology as a method of assessing neurotoxicant-induced alterations in nervous system function.
The pharmacological challenge is used to reveal the effect of a chemical exposure that has in some way altered the dynamic equilibrium of the nervous system which has adaptively compensated for a toxicant exposure by the use of existing homeostatic mechanisms. These adaptive changes may be revealed following the administration of a pharmacological agent that acts on the same system(s) affected by the neurotoxicant.

Neurotoxicologists have benefited from the use of pharmacological agents to address different questions in neurology. These agents have proven effective in determining the functional significance of the observed effect in those circumstances for which previous experiments have provided information concerning the neurogenic effects of a neurotoxicant. For example, neurochemical experiments showed that exposure to acrylamide resulted in an increase in dopamine receptor binding in the neostriatum of rats (Agrawal et al., 1981). Tilson and Squibb (1982) showed that rats exposed to a seemingly ineffective dose of acrylamide (i.e., no behavioral abnormality could be detected) were more responsive to the effects of apomorphine (a dopamine receptor agonist) and d-amphetamine (a releaser of catecholamines). This suggests that the dopamine system in acrylamide-exposed animals was functionally hyperresponsive.

Pharmacological challenges are also used to study the mechanism of action of neurotoxic agents. In this application, antagonism or exacerbation of a drug-interaction study method is used to compare the interaction between a psychoactive drug with a known mechanism of action to the effects of a neurotoxicant with an unknown or suspected mechanism of action. For example, the insecticides permethrin and p,p'-DDT are believed to produce tremor and behavioral hyperexcitability in the whole animal by blocking the sodium channels in the open position. This results in increased membrane excitability and repetitive firing of the nerve fiber. This interpretation was supported by an experiment (Tilson et al., 1985) which showed that pretreatment with hydantoin markedly attenuated the effects of these agents but had no effect or exacerbation of the effects of agents believed to have other mechanisms. Hydantoin is believed to block repetitive firing by binding to the inactivation gates of sodium channels.

Pharmacological agents have also been used to detect the presence of toxicity. For example, Harry and Tilson (1982) found that adult rats which had been neonatally exposed to tri-ethyl tin showed little if any signs of neurological disturbance. However, the tin-treated rats were found to be more sensitive to the stereotyped effects of apomorphine. Subsequent work found increased dopamine receptor binding in the corpus striatum of the tin-treated rats.
2. **Irritability**

Many toxicants are strong irritants which frequently affect the skin, eyes, respiratory pathways, and gastrointestinal tract. It has been proposed by Wood (1979, 1981), that behavioral procedures could be used to estimate the potential for chemicals to produce irritation. In such procedures, experimental animals are trained to make a response (e.g., a nosepoke to interrupt a photobeam) to terminate the presence of irritating agents which are introduced into the test environment. This bioassay was used by Wood to determine the relative irritability of several chemicals, including ozone, chlorine, toluene, acetic acid, and ammonia.

3. **Developmental neurotoxicology**

During the last decade, much attention has been drawn to the fact that exposure to environmental agents during development can result in long-lasting alterations in neurobehavioral function in the absence of physical malformations (Adams and Buelke-Sam, 1981; Barlow and Sullivan, 1975; Rodier, 1978; Spyker, 1975; Zbinden, 1981). Buelke-Sam and Kimmel (1979) have shown that behavioral tests can be used in developmental neurotoxicity to detect alterations that are manifest as a change in one or more neurobehavioral functions (i.e., sensorimotor or cognitive processes). There have been several tests or batteries of tests which have been suggested to assess the consequences of toxic exposure during development (Adams and Buelke-Sam, 1981; Barlow and Sullivan, 1975; Buelke-Sam and Kimmel, 1979; Butcher and Vorhees, 1979; Dews and Wenger, 1979; Grant, 1976; Mactutus and Tilson, 1985; Mactutus et al., 1984; Nelson, 1978; Reiter et al., 1980; Rodier, 1978; Spyker, 1975; Tesh, 1977; Zbinden, 1981). For example, the Psychotерatogenicity Screening Test Battery for rats (Butcher and Vorhees, 1979; Vorhees, 1983), includes the measurement of body weights of the dams and offspring in addition to several preweaning and postweaning behavioral tests procedures. The Collaborative Behavioral Teratology Study (Buelke-Sam et al., 1985) was initiated in response to a need for evaluation of standardized test procedures. This battery provides general toxicity measurements of both the dams and the offspring and includes testing procedures for both preweaning (eye opening, incisor eruption, negative geotaxis, olfactory orientation, startle response) and postweaning (startle response, motor activity, and pharmacological challenge).

Neurobehavioral function may be assessed during all stages of development, including maturity and senescence as determined in a longitudinal study strategy. Changes in neurobehavioral function might be expressed by retardation in the occurrence or rate of development of specific functions,
alterations in adult neurobehavioral capacity and/or a premature decline in functional capacity, or induction of premature senescence (Grant, 1976).

In routine acute and chronic toxicologic evaluations, the effects of toxicants in mature animals may be very different from those seen in the developing animal. Table V.5 summarizes a scheme formulated by Wilson (1973) that shows the developmental pathogenesis resulting from interference with fundamental aspects of cellular physiology. For example, protein synthesis or mitosis can lead to cell death or failure of some critical cellular function. The events at the cellular level lead to an abnormal developmental sequence which may be expressed as lethality or as a more subtle deficit such as behavioral or neurologic dysfunction. An important feature of Wilson's scheme is that agents with different toxic mechanisms might induce similar outcomes because they converge into a final common pathway of abnormal development. The nature of the functional change might depend upon which portion of the brain is developing in the CNS at the time of exposure. Rodier (1976) has clearly illustrated this principle by showing that the type of behavioral effect produced by a specific agent such as x-radiation or exposure to an antimitotic drug can depend upon the embryonic stage of development.

CONSIDERATIONS IN THE USE OF BEHAVIORAL PROCEDURES

A. MECHANISM OF ACTION

It is sometimes difficult to isolate the relative contributions of the various sensory, motor, arousal, or cognitive factors that contribute to an observed change in study findings from behavioral tests in pharmacology and toxicology. For example, if a toxicant induces an effect such as a decrease in the magnitude of the startle reflex to an acoustic stimulus, it may be inappropriate to conclude on the basis of this data alone that this toxic chemical has affected hearing. Other interpretations of the data may include that the neurotoxicant may have affected the capacity of the animal to respond to the stimulus as a consequence of altered motor function or that the toxic chemical may have interfered with the acoustic startle reflex circuit lying in the brainstem. Investigators have found that it is frequently necessary to conduct additional experiments or include appropriate control groups in order to define the behavioral mechanism of action.
Table V.5. Schematic summary of initial cellular reactions and different types of pathogenesis into a final common pathway.

<table>
<thead>
<tr>
<th>MECHANISMS</th>
<th>PATHOGENESIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Initial cellular reaction(s) within the germ cell, embryo, or fetus]</td>
<td>[Abnormal developmental sequence]</td>
</tr>
</tbody>
</table>

**DEMONSTRABLE EVENT**
Noted as one or more of the following:
- Mitotic interference
- Altered nucleic acid functions
- Mutation
- Chromosomal nondisjunction
- Lack of substrates, precursors, etc.
- Lack of energy sources
- Enzyme inhibition
- Changed membrane characteristics
- Osmolar imbalance

**EARLY PATHOGENESIS**
Initiated as one or more of the following:
- Cell death
- Failed cellular interaction (induction)
- Reduced biosynthesis
- Impaired morphogenic movement
- Mechanical disruption
- Tissue disruption
- Altered differentiation schedules

**FINAL COMMON PATHWAY**
(Insufficient cells or cell products to carry out morphogenesis or to carry on function.)

Resulting in final defect(s) expressed as one or more of the following:
- Intrauterine death
- Malformation
- Growth retardation
- Functional deficit (e.g., neurobehavioral, hormonal)

---

1 Modified from Wilson, 1973.
2 Final common pathway means that the many types of toxic effects may eventually cause common effects due to a common underlying substrate.
B. DEFINING AN ADVERSE EFFECT

In toxicology, as in many other sciences, there is considerable disagreement among scientists concerning the definition of an adverse effect when the study findings are based solely on behavioral endpoints. Many scientists have advocated that any evidence of a behavioral change constitutes an adverse effect, while others require evidence of an irreversible decrease in performance or enhanced susceptibility to the deleterious effects of other environmental influences. The Office of Toxic Substances (OTS) of the U.S. Environmental Protection Agency (1982) has defined neurotoxicity as any adverse effect on the structure or function of the central and/or peripheral nervous system that is related to an exposure to a chemical substance; whereas a toxic effect has been defined by OTS as an adverse change in the structure or function of an experimental animal that results from the exposure to a chemical substance.

The inherent variability of biological data and an imprecise definition of "normal" limits of functioning are two factors which complicate assessments of the adverse effects that can be detected by behavioral testing procedures. Discrimination between adverse and nonadverse effects requires an understanding of the importance of reversible changes by investigators and the sensitivity of the testing procedure to detect and quantify deviations from "normal".

C. FUNCTIONAL RESERVE

The functional reserve capacity of various compensatory mechanisms which are inherent in a biological system may obfuscate the presence of toxicant-induced damage as assessed by functional test procedures. One testing approach that is frequently used by investigators to focus on the toxicant-induced effect is to incorporate into the testing procedures one or more conditions in which the system or organism is subjected to an environmental challenge. The combination of the test substance plus the challenge might reveal the presence of latent toxicity or may result in a greater deficit in performance than might otherwise be seen without the stressor. For example, Chance (1946) showed that the lethal doses of d,l-amphetamine were much smaller for group-housed mice than for individually-housed mice.

D. STATISTICAL CONSIDERATIONS

The statistical considerations that are confronted by both behavioral and other functional tests are similar to those of any biological endpoint. For example, when it is desirable to test for the presence of a specific functional change, it is important for investigators to use tests that are valid in the sense that the test is both sufficiently sensitive and selective
to provide a measurement of the functional change in question. If the test findings are used in toxicologic assessments to predict adverse health effects in humans, then it is also important that tests have predictive validity. Behavioral test findings should be consistent with the scientific literature and be readily reproduced in repeated studies under the same testing conditions.

SUMMARY

Various national and international groups have indicated that behavioral tests are useful in the provisional assessment of chemicals. Toxicological protocols that are in current use frequently include observations for the presence of toxic signs which include indicators of neurotoxicity; however, a systematic and standardized battery of tests has not been widely accepted by the scientific community at this time. The systematic assessment of neurologic function could be enhanced by including a standardized battery of behavioral testing procedures into already existing toxicological protocols. This review has shown that there are sufficient data regarding available behavioral testing procedures to support the use of a standardized battery of tests in the context of a sequential toxicology screening format. Indications of neurotoxicity detected at this level could be further characterized by any number of secondary tests, the selection of which would depend upon the nature of the suspected functional defect and the demonstrated validity, reliability, sensitivity, and reproducibility of the testing methods included in the test protocol.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the technical assistance of Ms. Loretta Moore in the preparation of this manuscript and the constructive comments received from Drs. L.W. Reiter and T.J. Sobotka.
OPEN DISCUSSION

HUTCHINGS: Questions for Dr. Tilson?

WENK: I'm from Johns Hopkins. With regards to the interpretation of the behavioral data that you have outlined and the extension to humans, you and other speakers have already commented on how difficult it is to make that extension. I wonder if it might be easier to interpret the data and extend it to the human condition if less emphasis was given to the sensitivity of the task and more to the selectivity of those tasks? For example, of the tasks that you mentioned, passive avoidance is a very sensitive task. It is sensitive to many manipulations in the brain that are very difficult to interpret. Many lesions in the brain can impair an animal's performance in a passive avoidance task. We have often felt that yelling at the animal can make it become impaired in the passive avoidance task. I would like to point out that maybe the selectivity should be emphasized and a little less emphasis put on sensitivity.

TILSON: Many of the tests that I talked about do afford some cross-species extrapolation. Of course, when you get into tests of memory and learning, it's always a problem. Effects on learning and memory are basically inferred from observable behaviors.

There is enough theoretical literature to show that tests like the passive avoidance procedure measure certain components of memory and retention that might generalize to processes seen in humans. In the case of the radial arm maze test for animals, there are processes of working and reference memory that are clearly similar to processes seen in humans.

With regard to the passive avoidance task, I believe that this task must be conducted under controlled conditions. Five years ago, I argued against the use of the passive avoidance test. However, over the last 3 or 4 years, it has been demonstrated by Charles Mactutus (Mactutus et al., 1982) and others that, if you control the conditions appropriately, you can detect subtle changes in animal behavior. The passive avoidance task can be an appropriate tool, if used correctly.

WOOD: Ronald Wood, University of Rochester. As most of what you have described so far this morning has been in the context of hazard identification, could you talk a little more about the
relationship to dose and its use in risk estimation? Specifically are these procedures useful for describing dose effect relationships in risk estimations?

TILSON: In behavioral pharmacology, such tests are useful in describing dose effect relationships, and I believe that they might be equally effective in toxicology.

BOAST: Dr. Boast, CIBA-GEIGY Pharmaceuticals. You mentioned the subtlety of some of the behavioral tests that assess learning and memory. It is my perception that you are proposing a sequence of tests that would first determine some abnormality of gross motor activity or sensory dysfunction. It seems to me that in many cases we are very interested in those subtle cognitive losses and that perhaps these kinds of tests should be used as a primary battery. Could you comment on that?

TILSON: What I proposed was that, in the provisional assessment of chemicals, some sort of standardized checklist be devised that would include measurements or observations that are already made. Presently, these measures are not usually documented in a systematic way. Changes seen in the functional battery might be the trigger to perform other kinds of tests. I am not sure if it is appropriate to put learning and retention measurements at the initial stage of toxicological assessment. What we are trying to do is to standardize the observational measurements that are currently being made as part of toxicological testing.

LATIES: I think we are in a real box here, especially in dealing with new chemicals. At one and the same time, you want something that is cheap and you want something that is enormously sensitive and specific. I suspect that no observational battery is ever going to do that. It may be that you will need something very sophisticated and expensive as the first pass at a chemical. It is not possible to get a strong notion of what is happening with an observational battery. You will always be finding false negatives. I do not know how to get out of that except by spending the money and that is a difficult decision to make.

TILSON: I am not sure that I agree with the statement that we expect these tests to be overly sensitive or selective. We have already been given examples where, if people would have recorded their observations in a systematic way, then neurotoxicity would have been detected. I believe that this would have led to further characterization of that neurotoxicity.
We need the chance to evaluate the data that we already collect and look at it in a systematic way. Then we can make decisions about what other tests to do afterward. Rather than spending more money and putting a burden on those people that have to use these tests all the time, we should collect the data and look at it now. We should extend the data base in a systematic way.

LATIES: So long as you are recommending a systematic observational battery, plus some sort of apical test, I certainly would go along with that.

TILSON: Some might consider the functional observational battery as an apical test because many of those tests rely on sensorimotor integration. Nonetheless, I am advocating that we use a systematic battery which should provide us with the basis for appropriate test selection at the next level of analysis.

ZENICK: Hal Zenick, EPA. I would like to comment on consummatory behaviors and the aspect of whether or not we design artificial tasks for study animals (i.e., ask the animal to do something that it does not normally do).

Over the past several years we have used a protocol that requires us to monitor copulatory behavior in the male animal before we recover a semen sample. One of the things that we have observed is that with several neurotoxicants there are disruptions. I am specifically thinking about acrylamide, where we try with difficulty to evaluate splayed gait or hindlimb strength measurements. Yet with copulatory behaviors we have a very simple behavior. These animals have difficulty with mounts and intromissions. It is readily observable in a natural setting.

I wonder if there are roles for certain types of consummatory behaviors that can readily be quantified in a fairly inexpensive manner. With the acrylamide example, we observed that the offshoot of this disruption is that a lot of sperm-positive females never become pregnant. We found that the acrylamide exposed male can not coordinate copulatory behavior sufficiently, and there are no sperm in the uterus. There can be consequences far reaching beyond just the basic motor impairment. I would make an argument for at least the imposition of certain types of consummatory behaviors being evaluated.
SPENCER: Just a small comment on the last point. Before it becomes a generalized point enshrined in the neurotoxicology literature, we should remember that acrylamide not only induces a sensorimotor neuropathy, but also causes substantial changes in the autonomic nervous system. These have been well described by an Australian group (Post and McLeod, 1977). One wonders whether or not rats on study are undergoing retrograde ejaculation because of autonomic dysfunction. I realize you were trying to use it for illustrative purposes, but perhaps that was a rather specific effect of acrylamide, because of its adverse effects on the autonomic nervous system.

SINGER: My name is Raymond Singer. I am from New York City. I would like to say a few words about the choice of tasks for screening purposes. My experience has been working with humans at the Environmental Sciences Laboratory, where one of our tasks was to develop a battery of tests that would screen human exposure to a variety of toxic chemicals in a short period of time -- 20 minutes.

One of the criteria that we ended up using was premature aging. The idea was that many of the indications of normal aging are also seen in people with toxic exposures, but these effects are seen at an earlier age. This was one of our criteria for selection of tests and it offered a good correlation with the tests for normal aging.

The second characteristic was the complex reaction time, which is something that the person or the test subject had to think about and process centrally. Psychomotor speed was an important factor that was also assessed. This may be some assistance in selecting a short battery of tests for screening animal toxicity.

TILSON: It is good to know that such tests may be useful in humans, since it might provide the basis for devising tests in animals. However, accelerated aging is a difficult thing to assess in animals such as rodents. The other tests you mentioned such as tests for changes in reaction time could be studied in animals.

VOICE: How do you assess the aging?

SINGER: There is what would be considered normal aging, in an epidemiologic sense. There is a decrease in the ability to learn new material, to concentrate, and the ability to remember. In the example that was given of the maze for testing memory or retention in the rat, one would expect
an improvement over some of the simple reaction times because in that test, the rat has to learn something new, retain it, and reproduce it.

YOUNG: John Young, Johns Hopkins University. I would like to express a concern with the hierarchical testing scheme with specific reference to sensory function, because I can not think of any quick screening tests that have any adequacy for sensory function at all. These tend to give you nothing but false negatives.

TILSON: Or false positives. I think some of Marshall Steinberg's data are probably prototypic of this approach. Such tests certainly do not detect changes in acuity. They are intended to determine whether or not the animal can see, feel, hear, touch, or whatever. I think that some of these tests could be used to detect the presence or absence of a deficit.

YOUNG: To be specific, in the auditory system, the common test is Pryor's reflex, which is an ear flick in rodents. This reflex has been shown in several studies to be intact after hearing losses up to 70 db, which is functional deafness.

TILSON: That is right, however, I did not necessarily recommend that test. But I am glad you pointed it out.

REITER: I think there is a point here that needs clarification, as it relates to an attempt in these discussions to reduce testing to a single situation or a single battery of tests. A problem with this approach is that not all chemicals come to the testing table with the same amount of available background information on toxicity. When presented with a chemical for which no a priori information exists, the first requirement is to screen (detect) for neurotoxicity associated with exposure to that chemical. The next step in the process is to characterize the nature and extent of neurotoxic effects. Finally, there is a need to define the dose- and time-effect relationships aimed at determining no-observable-effects levels which may be useful for standard setting. The point here is that the questions being asked about the neurotoxicity of a chemical are often quite different and therefore, the methods utilized will be different.

A question was also raised about behavioral teratology testing which was similar to an issue raised at a recent meeting in Cincinnati on the collaborative behavioral teratology study. Developmental behavioral testing is
often performed in conjunction with reproduction studies. In the pharmaceutical industry, reproductive studies are usually performed in the late stages of drug development. At this point in time, the company has already invested considerable resources into the development of a chemical and is not particularly interested in pursuing an apical primary-secondary testing approach. I have been told they want to perform the needed testing so that they can proceed with the drug development process. For industrial chemicals, on the other hand, the tier-testing approach may be more practical. Again, the point is that the testing strategy will depend on the questions being asked.

Under the premanufacturing notification process, EPA is faced with a large number of chemicals (1300-1600 new chemicals introduced each year) and can not arbitrarily require that every chemical be tested for neurotoxicity. Regulatory agencies are limited with respect to the number of tests that they can reasonably request industry to apply to a new chemical. If one examines the literature on neurotoxic chemicals, there are very few which produce a single neuro-behavioral effect. Therefore, in terms of acute and sub-chronic testing, a hierarchical approach should provide for the detection and characterization of neurotoxic chemicals. I guess the real issue is whether the tests used at the primary screening level comprise a "fishnet" which is so loose that many chemicals slip through.

LEVINE: Tina Levine from the Office of Toxic Substances of the Environmental Protection Agency. I heard you talk about the World Health Organization suggesting that it was time to validate the functional observational battery. The word "validation" has been a thorn in our side for a long time. I wonder whether you feel that such validation is necessary and how you would go about it?

TILSON: There is some degree of validity that is already taking place, that is whether or not these tests can detect the presence of toxicity and whether or not certain kinds of compounds will produce positive effects. Some data exist for that already. I think if we started collecting the data base for compounds that we are already testing, then we should begin to see how valid some of these tests really are.

VORHEES: I agree with the idea that when you are talking about screening large numbers of environmental chemicals, you have to approach this problem in terms of tiers of testing. However, I think that I might take issue with you in terms of what might be included in that first tier? If you set the first tier so that it becomes too gross a measure, then you
are not going to detect anything and you will never get to the second tier. This is the potential dilemma and the reason why tests such as the modification of the acoustic startle response and measures of instrumental learning are important to be considered in the first tier of analysis and not deferred to the second tier where they will never be reached. These tests are important to detect the facts which are specific to those situations.

TILSON: Perhaps it all depends on your perspective whether it be from a practical point of view, from an objective point of view, from the viewpoint of those that are regulated, or the regulators. You still have to deal with what can be done. The question is, are any of these tests going to be useful at all? The answer, at least in my opinion, is that they are useful and that the data are already being collected, but not necessarily in a systematic way. The critical issue is, what will happen when you get some positive effects in some of these tests? Is the mechanism that triggers the next level going to be there, or are the data going to be ignored?

VORHEES: I think that we are in agreement that the functional observational battery may have some usefulness, however, I question whether or not it is adequate. The second thing is that I do not think our purpose here is to define what can be done or what kind of regulatory changes can be made; I think that our purpose here is to suggest what should be done. It ends up being a political decision about what can be effected. What I think I hear you saying is because the functional observational battery can be invoked, that is all we should really ask for, whether it is completely adequate or not.

Maybe I misunderstand you, but it seems to me that what we want to define is what ought to be done under the best of circumstances, and then the political decision may have to be that that has to be modified in some way. Fundamentally, I do not think that is a scientific decision.

TILSON: What should and should not be done is, by and large, a qualitative decision, anyway. We have to look at these tests in a practical way and ask what can be accomplished by applying certain types of tests under certain circumstances. I think the time has come to really consider using these tests at this level of provisional assessment.

WOOD: In addressing this issue, would it be helpful not to discuss individual tests but, rather, classes of functions that you might evaluate? The National Academy has recently
indicated that tests of unlearned and learned performance, as well as a neuropathologic examination, would be some minimal first step towards assessing the neurotoxicity of environmental chemicals.

TILSON: Right.

HUTCHINGS: I hate to interrupt this discussion at this point but we need to move along to the next paper. There will be a general discussion this afternoon, so maybe we can continue this discussion at that time.
LITERATURE CITED


American Conference of Governmental Industrial Hygienists Inc. 1982. Threshold limit values for chemical substances and physical agents in the work environment, with intended changes for 1982. Available from: American Conference of Governmental Industrial Hygienists Inc., Cincinnati.


VI. RELIABILITY, SENSITIVITY AND VALIDITY OF BEHAVIORAL INDICES OF NEUROTOXICITY

Charles V. Vorhees, Ph.D.
Institute for Developmental Research
Children's Hospital Research Foundation and
Department of Pediatrics
University of Cincinnati College of Medicine
Cincinnati, Ohio

ABSTRACT

Adequate sensitivity, reliability, and validity are features of scientific testing in all fields. Accordingly, these have been central concerns in the development of behavioral methods for assessing neurotoxicity. Advances during the last 10 years have established behavioral tests that show high degrees of sensitivity and reliability both within and across laboratories. This was demonstrated in the recently completed National Center for Toxico logical Research (NCTR) Collaborative Behavioral Teratology Study and other collaborative studies. Validity of behavioral testing has been based on: a) the use of known neurotoxins; b) comparison of agents causing behavioral dysfunction in humans; and c) theoretical considerations of domains of central nervous system (CNS) function to be evaluated. These approaches have worked well, however, the neurochemical/neuropathologic to behavior correlations have been effective only under circumstances where a neurotoxic compound, as recognized by an identifiable neurobiologic signature, has been available. Behavioral methods are important in neurotoxicity evaluations because they provide an assessment of overall CNS functional integrity, whereas other methods in the neurosciences are directed towards specific subsystems.

This presentation discusses the validity of measurements in: 1) behavioral testing; 2) neuropathologic evaluations; 3) neurochemical tests; and 4) electrophysiological techniques for detecting neurotoxicity.
RELIABILITY OF BEHAVIORAL INDICES OF NEUROTOXICITY

Logic dictates that health protection guidelines be arranged in a hierarchical order of importance starting with mortality, followed by a downward progression in severity to lessening degrees of morbidity. Testing to determine the absence of brain damage by such an approach should logically precede that of many other organ systems because the brain functions as the nucleus of human health and well-being. No balanced view could place the brain low on the list of priorities for protection. However, many scientists recognize that there is an imbalance in the current regulatory practice of preclinical toxicity testing in that the nervous system is less thoroughly monitored for safety than virtually any other organ system.

It is perplexing that in 1985 when most aspects of human health have been protected via federal regulation that there remains a need to further regulate safety testing to protect the brain and its function (behavior). Advances in the neurobehavioral sciences in the last 10-20 years have overcome many problems associated with measurement of CNS structure and function. Elegant methods now exist for measuring a wide variety of CNS functions with remarkable precision, whether they be electrophysiological, neurochemical, neuroanatomical, or behavioral events. While knowledge of other organ systems has established correlation between structural and functional pathology, such correlations for neurotoxicity as demonstrated by behavior and brain function are substantially more complex and difficult to establish.

Measurements of CNS functional integrity are important to developing a comprehensive assessment of health and toxic potential. It is difficult to imagine an adequate safety system that does not include the appraisal of crucial aspects of CNS functional intactness. Current toxicologic assessments for detecting neurotoxic effects from standard toxicologic examinations have frequently proven inadequate from the perspective of measuring those aspects of CNS injury that can occur from some chemical exposures. Although the traditional testing methods may sometimes detect an effect on the CNS (i.e., a correct positive for neurotoxicity), the false negative rate is a serious and incalculable problem that hinders good hazard assessment. The current federal regulatory requirements for observations of behavior are too vague and the pathologic examination of three sections of brain and two sections of spinal cord are too rudimentary to be of any real value in detecting neurotoxicity. Reliable and valid methods are available for evaluating CNS integrity and these should be one of the cornerstones of toxicologic assessment.

This paper will describe the reliability, sensitivity, and validity of behavioral indices of neurotoxicity. In discussing validity, a comparison of behavioral to neurochemical,
neuropathological, and electrophysiological methods will be made from representative examples. The focus of this discussion will be upon behavioral teratogenesis (i.e., developmental neurobehavioral toxicity) rather than on adult behavioral toxicity although the underlying points to be made apply equally well to both. The emphasis on developmental neurotoxicity is appropriate for this discussion because there has been a considerable amount of research on methods development in behavioral teratology. Unfortunately, these methods have no direct counterpart to permit literature comparisons with behavioral toxicology of adult animals at this time.

Pertinent to this discussion are the findings from two research projects aimed at developing methods for detecting behavioral teratogenicity. This includes both the set of data from the National Collaborative Behavioral Teratology Study, sponsored by the National Center for Toxicological Research (NCTR) of the Food and Drug Administration (FDA) and cosponsored by the Environmental Protection Agency (EPA), National Toxicology Program-National Institutes of Environmental Health Science (NTP-NIEHS), and National Institute of Occupational Safety and Health (NIOSH) (Adams et al., 1985a,b,c; Buelke-Sam et al., 1985; Holson et al., 1985; Kimmel and Buelke-Sam, 1985; Kimmel et al., 1985; Nelson et al., 1985) and the results of methods research funded by the former Bureau of Foods of the FDA (Vorhees et al., 1979a, 1981a,b, 1983a,b,c, 1984a,b). Both these projects studied albino rats. It should be recognized that standardization of the animals selected for preclinical studies is equally important as the standardization of the behavioral tests and the procedures outlined in the test protocol. The standardization of study animals should be taken into consideration at least to the level of species and stocks (or strains), although in certain studies it may be necessary to take substocks into consideration. The studies cited above have used Sprague-Dawley CD rats from Charles River (Portage, Michigan) or Sprague-Dawley rats from the Laboratory Supply Company (Indianapolis, Indiana). The generality of future toxicologic and neurobehavioral study findings may be further enhanced by the use of heterogeneously bred stocks of rats originated through systematic crossing of a number of defined inbred strains (Wilson, 1986).

A. RELIABILITY OF BEHAVIORAL METHODS

For a test to be reliable it should produce both a consistent pattern of results within the same laboratory concurrently and over time (intralaboratory reliability) and across different laboratories (interlaboratory reliability). Reliability will be first discussed with respect to stability over time.
1. **Intralaboratory reliability**

Under ideal circumstances, test stability would occur regardless of noise in the testing system. Noise can be described as any change in the test stemming from changes in irrelevant background factors, such as random reshuffling of the gene pool in the stock of animals being tested, because of changes in season, food or water constituents, or fluctuations in temperature and humidity. On the other hand, the test should reflect real changes which occur in all populations over time because of factors such as genetic drift, changes in housing and handling conditions by the animal supplier. Although beyond the control or knowledge of the experimenter, such changes can influence measurements of behavior and can almost never be completely specified even by the most attentive researcher. For these reasons it is incorrect to assume that a test is unreliable if the data indicate a slightly different profile in control subjects in 1985 than the same test produced when conducted 10 years earlier. Change in the baseline level of performance alone is not the arbiter of reliability unless all significant influences on the test behaviors can be controlled.

An examination of some negative control data from rat studies conducted over a period of years in the Cincinnati FDA project on methods development and food additive behavioral toxicity illustrates the drift in these background factors. In the data that follow (Table VI.1), comparisons have been made for groups of control Sprague-Dawley rats obtained from the same supplier (Laboratory Supply Company, Indianapolis, Indiana) between June 1975 and June 1980. Although the experiments from which these control groups were drawn overlapped, each experiment was entirely separate from those preceding and following it in this series. With the exception of the C9 group (a saline injected control), all the experiments were food additive studies in which the test compound was administered in the diet. The experimental controls from these Cincinnati FDA studies represent groups of test animals which are equivalent to the "untreated controls" in the NCTR Collaborative Behavioral Teratology Study. The group designations used in Table VI.1 are arranged chronologically to refer to the dates when each study was conducted. Individual project protocols and behavioral methods have been described elsewhere (Vorhees et al., 1979a, 1981a,b, 1983a,b,c, 1984a,b). In subsequent tables, N represents the number of dams or litters, while n represents the number of individual offspring examined in those few cases where litter was not the unit of analysis. In general, four offspring were tested per litter, however, two offspring per litter were used in a few time-consuming tests. The mean litter value was used as the datum for litter in all group analyses.

These data illustrate several noteworthy points. Note that the mean values change over time even for basic parameters such as length of gestation, offspring body weight, and vaginal
Table VI.l. Control Group Codes and Performance Dates for the Cincinnati Study

<table>
<thead>
<tr>
<th>Group</th>
<th>Study Dates</th>
<th>Study Length&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Date Begun</td>
<td>Date Ended</td>
</tr>
<tr>
<td>C1</td>
<td>8/75</td>
<td>8/76</td>
</tr>
<tr>
<td>C2</td>
<td>4/76</td>
<td>3/77</td>
</tr>
<tr>
<td>C3</td>
<td>12/76</td>
<td>8/77</td>
</tr>
<tr>
<td>C4</td>
<td>6/77</td>
<td>3/78</td>
</tr>
<tr>
<td>C5</td>
<td>1/78</td>
<td>8/78</td>
</tr>
<tr>
<td>C6</td>
<td>7/78</td>
<td>5/79</td>
</tr>
<tr>
<td>C7</td>
<td>6/79</td>
<td>1/80</td>
</tr>
<tr>
<td>C8</td>
<td>6/79</td>
<td>1/80</td>
</tr>
<tr>
<td>C9</td>
<td>7/79</td>
<td>4/80</td>
</tr>
</tbody>
</table>

<sup>a</sup> Study length is given in months.
patency development (Tables VI.2, VI.3, and VI.4). If fluctuations of these standard toxicologic measurements are considered to be within the normal range, then it should be expected that examination of behavioral measures should also produce some fluctuations over time. Table VI.5 illustrates an example of this fluctuation for a simple behavioral reflex such as surface righting. Note that the amount of statistically significant fluctuation among controls over a 5-year period is modest. This is also true for measures such as development of the auditory startle reflex (Table VI.6). Some changes can be accounted for, as in the C9 saline injection group, by the fact that the treatment procedure was actually changed. This change in treatment procedure visibly affected the dependent variable. Not all procedural changes alter the outcome, as seen by the example of the test procedure change introduced for the C3 group (Table VI.6).

Procedural differences can account for virtually all of the significant variations over time on a measure such as pivoting locomotion (Table VI.7). Few significant fluctuations were seen in negative geotaxis (Table VI.8) and swimming ontogeny over time (Tables VI.9, VI.10, and VI.11) other than those that may be accounted for by material changes in the test procedure or in the definition of the response being measured.

Manually conducted tests such as the open-field test of locomotor activity show good stability over time in this experimental series (Tables VI.11, VI.12, VI.13, VI.14, and VI.15) even though this test is known to have high variability.

Measurements of passive avoidance have been reported in the literature to have some replicability problems, although as indicated in Table VI.16, this test can show moderate within-laboratory stability over a 5-year period. Improved testing procedures are needed, however, to increase the sensitivity of this measurement.

Table VI.17 shows the variation over time in 90 day body and brain weight. These data indicate a significant fluctuation for a measure thought to be as stable as brain weight over the 5-year time period. Significant variation occurred among the control groups in this study even though all brain measurements over the 5-year period were performed by one individual highly experienced in rat brain removal and dissection. In contrast, some of the behavioral measures that contained higher intrinsic variability (hence detection power is somewhat different) and were conducted by more than 15 different research assistants over the 5-year life of study, often showed few or no significant fluctuations among the control groups. Demonstration that a test is stable over time is not an indication that it is optimal to reflect the parameter being measured. Sensitivity, inter-laboratory reliability, and validity must also be established.
<table>
<thead>
<tr>
<th>Group</th>
<th>Mean ± SE(^a)</th>
<th>No. of Dams</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>22.2 ± 0.2</td>
<td>14</td>
</tr>
<tr>
<td>C2</td>
<td>22.5 ± 0.3</td>
<td>17</td>
</tr>
<tr>
<td>C3</td>
<td>22.6 ± 0.1</td>
<td>19</td>
</tr>
<tr>
<td>C4</td>
<td>21.8 ± 0.2</td>
<td>14</td>
</tr>
<tr>
<td>C5</td>
<td>21.6 ± 0.2</td>
<td>5</td>
</tr>
<tr>
<td>C6</td>
<td>22.2 ± 0.2</td>
<td>26</td>
</tr>
<tr>
<td>C7</td>
<td>22.1 ± 0.1</td>
<td>12</td>
</tr>
<tr>
<td>C8</td>
<td>21.7 ± 0.2(^b)</td>
<td>14</td>
</tr>
<tr>
<td>C9</td>
<td>22.1 ± 0.1</td>
<td>21</td>
</tr>
</tbody>
</table>

\(^a\) Group main-effect, p <0.01; values are in days.

\(^b\) p <0.05 compared to group C3 by Duncan's multiple range \textit{a posteriori} test.
Table VI.3. Preweaning Body Weights for Males\textsuperscript{a} Using Mean (SE)

<table>
<thead>
<tr>
<th>Group</th>
<th>Day 1</th>
<th>Day 7</th>
<th>Day 14</th>
<th>Day 21</th>
<th>No. of Litters</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>7.4 (0.2)</td>
<td>14.6 (0.4)</td>
<td>25.9 (0.6)</td>
<td>37.8 (1.5)</td>
<td>20</td>
</tr>
<tr>
<td>C2</td>
<td>7.5 (0.2)</td>
<td>15.1 (0.4)</td>
<td>25.4 (0.6)</td>
<td>35.8 (0.9)</td>
<td>17</td>
</tr>
<tr>
<td>C3</td>
<td>7.7 (0.2)</td>
<td>14.8 (0.4)</td>
<td>25.6 (0.9)</td>
<td>34.6 (1.4)</td>
<td>18</td>
</tr>
<tr>
<td>C4</td>
<td>7.3 (0.2)</td>
<td>14.8 (0.4)</td>
<td>26.4 (0.8)</td>
<td>37.7 (1.5)</td>
<td>15</td>
</tr>
<tr>
<td>C5</td>
<td>7.2 (0.3)</td>
<td>15.2 (0.6)</td>
<td>26.0 (0.7)</td>
<td>37.8 (1.8)</td>
<td>6</td>
</tr>
<tr>
<td>C6</td>
<td>7.2 (0.2)</td>
<td>15.2 (0.4)</td>
<td>27.3 (0.7)</td>
<td>37.7 (1.0)</td>
<td>25</td>
</tr>
<tr>
<td>C7</td>
<td>7.5 (0.2)</td>
<td>15.3 (0.6)</td>
<td>27.4 (0.9)</td>
<td>39.2 (1.9)</td>
<td>19</td>
</tr>
<tr>
<td>C8</td>
<td>7.2 (0.2)</td>
<td>14.1 (0.8)</td>
<td>24.5 (1.7)</td>
<td>39.4 (1.9)</td>
<td>12</td>
</tr>
<tr>
<td>C9</td>
<td>7.0 (0.2)</td>
<td>13.0 (0.8)</td>
<td>24.0 (1.1)</td>
<td>34.8 (1.6)</td>
<td>11\textsuperscript{b}</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Group main-effect, \( p < 0.02 \).
\textsuperscript{b} \( p < 0.05 \) compared to groups C4, C6, or C8 across sexes and days by Duncan's test.
Table VI.4. Vaginal Patency

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SE) Age at Patency&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. of Dams</th>
</tr>
</thead>
<tbody>
<tr>
<td>C3</td>
<td>39.2 (0.7)</td>
<td>18</td>
</tr>
<tr>
<td>C4</td>
<td>39.3 (0.8)</td>
<td>15</td>
</tr>
<tr>
<td>C5</td>
<td>43.0 (0.6)</td>
<td>5</td>
</tr>
<tr>
<td>C6</td>
<td>44.8 (1.2)</td>
<td>23&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>C7</td>
<td>43.6 (1.9)</td>
<td>13&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>C8</td>
<td>39.2 (3.3)</td>
<td>12</td>
</tr>
<tr>
<td>C9</td>
<td>46.2 (1.8)</td>
<td>10&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup> Group effect, p = 0.001.
<sup>b</sup> p < 0.05 compared to groups C3, C4, C8 by Duncan's test.
Table VI.5. Surface Righting Development

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SE) Day To Right In ≤2.0 Sec.</th>
<th>No. of Litters</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>9.0 (0.4)</td>
<td>20</td>
</tr>
<tr>
<td>C2</td>
<td>8.0 (0.2)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>19</td>
</tr>
<tr>
<td>C3</td>
<td>8.7 (0.4)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>18</td>
</tr>
<tr>
<td>C4</td>
<td>9.6 (0.4)</td>
<td>15</td>
</tr>
<tr>
<td>C5</td>
<td>9.8 (0.6)</td>
<td>6</td>
</tr>
<tr>
<td>C6</td>
<td>9.3 (0.4)</td>
<td>25</td>
</tr>
<tr>
<td>C7</td>
<td>10.1 (0.3)</td>
<td>19</td>
</tr>
<tr>
<td>C8</td>
<td>9.2 (0.8)</td>
<td>12</td>
</tr>
<tr>
<td>C9</td>
<td>9.9 (0.4)</td>
<td>12</td>
</tr>
</tbody>
</table>

<sup>a</sup> Group main-effect, p <0.01.
<sup>b</sup> p <0.05 compared to groups C4, C5, C6, C7, C9 by Duncan's test.
<sup>c</sup> p <0.05 compared to group C7 by Duncan's test.
Table VI.6. Auditory Startle Response Development

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SE) Day To Startle&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. of Litters</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>12.3 (0.2)</td>
<td>21</td>
</tr>
<tr>
<td>C2</td>
<td>12.7 (0.3)</td>
<td>19</td>
</tr>
<tr>
<td>C3&lt;sup&gt;b&lt;/sup&gt;</td>
<td>12.2 (0.3)</td>
<td>19</td>
</tr>
<tr>
<td>C4</td>
<td>13.1 (0.2)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>12</td>
</tr>
<tr>
<td>C5</td>
<td>12.2 (0.3)</td>
<td>6</td>
</tr>
<tr>
<td>C6</td>
<td>12.4 (0.2)</td>
<td>26</td>
</tr>
<tr>
<td>C7&lt;sup&gt;d&lt;/sup&gt;</td>
<td>12.7 (0.2)</td>
<td>18</td>
</tr>
<tr>
<td>C8</td>
<td>13.0 (0.4)</td>
<td>12</td>
</tr>
<tr>
<td>C9</td>
<td>13.2 (0.3)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>11</td>
</tr>
</tbody>
</table>

<sup>a</sup> Group effect, p < 0.07.
<sup>b</sup> Procedural change: Testing began on day 10 from hereafter rather than day 8.
<sup>c</sup> p < 0.05 compared to C3 by Duncan's test.
<sup>d</sup> Procedural change: One trial/day from hereafter rather than 2.
<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SE) Time Spent Pivoting (sec) out of 1 Min.</th>
<th>No. of Litters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 7</td>
<td>Day 9</td>
</tr>
<tr>
<td>C3</td>
<td>12.6 (1.8)</td>
<td>14.5 (1.6)</td>
</tr>
<tr>
<td>C4</td>
<td>13.6 (3.1)</td>
<td>12.8 (1.7)</td>
</tr>
<tr>
<td>C5b</td>
<td>3.7 (0.9)</td>
<td>5.1 (1.5)</td>
</tr>
<tr>
<td>C6b</td>
<td>6.7 (1.2)</td>
<td>9.0 (0.9)</td>
</tr>
<tr>
<td>C7b</td>
<td>3.3 (0.5)</td>
<td>6.1 (0.8)</td>
</tr>
<tr>
<td>C8b</td>
<td>7.2 (2.0)</td>
<td>5.2 (1.1)</td>
</tr>
<tr>
<td>C9b</td>
<td>5.1 (2.0)</td>
<td>7.7 (2.1)</td>
</tr>
</tbody>
</table>

a Group main-effect, p < 0.001; group x days, ns.
b Changed in definition of pivoting compared to groups C3 and C4.
c p < 0.01 compared to group C3 across sexes and days by Duncan's test.
Table VI.8. Negative Geotaxis

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SE) Time\textsuperscript{a} to Rotate Through 180° on a 25° Plane</th>
<th>No. of Litters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 6</td>
<td>Day 8</td>
</tr>
<tr>
<td>C3\textsuperscript{b}</td>
<td>29.8 (2.9)</td>
<td>19.6 (2.9)</td>
</tr>
<tr>
<td>C4</td>
<td>40.9 (3.7)</td>
<td>21.7 (3.4)</td>
</tr>
<tr>
<td>C5</td>
<td>35.9 (2.5)</td>
<td>20.5 (2.4)</td>
</tr>
<tr>
<td>C6</td>
<td>37.4 (2.4)</td>
<td>24.5 (2.2)</td>
</tr>
<tr>
<td>C7</td>
<td>36.8 (3.3)</td>
<td>22.8 (2.9)</td>
</tr>
<tr>
<td>C8</td>
<td>42.9 (3.9)</td>
<td>30.2 (4.3)</td>
</tr>
<tr>
<td>C9</td>
<td>40.2 (4.0)</td>
<td>31.9 (5.0)</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Group main-effect, p < 0.05; group x days, ns.
\textsuperscript{b} Procedural difference: Angle of plane set at approximately 27°.
\textsuperscript{c} p < 0.05 compared to group C3 across sexes and days by Duncan's test.
Table VI.9. Directional Swimming Ontogeny in Males

<table>
<thead>
<tr>
<th>Group</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>6-10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>12-20</th>
<th>No. of Litters</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>1.4(.1)</td>
<td>1.7(.1)</td>
<td>1.9(.1)</td>
<td>-</td>
<td>2.1(.1)</td>
<td>2.3(.1)</td>
<td>2.8(.1)</td>
<td>3.0(.0)</td>
<td>3.0(.0)</td>
<td>-</td>
<td>19</td>
</tr>
<tr>
<td>C2</td>
<td>1.5(.1)</td>
<td>1.8(.1)</td>
<td>2.0(.1)</td>
<td>1.8</td>
<td>2.2(.1)</td>
<td>2.4(.1)</td>
<td>2.9(.1)</td>
<td>3.0(.0)</td>
<td>3.0(.0)</td>
<td>2.7</td>
<td>19</td>
</tr>
<tr>
<td>C3</td>
<td>1.8(.1)</td>
<td>1.9(.1)</td>
<td>2.0(.1)</td>
<td>1.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>17</td>
</tr>
<tr>
<td>C4</td>
<td>1.8(.1)</td>
<td>2.0(.0)</td>
<td>2.1(.1)</td>
<td>2.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>15</td>
</tr>
<tr>
<td>C5</td>
<td>1.8(.2)</td>
<td>2.0(.0)</td>
<td>2.0(.0)</td>
<td>2.0</td>
<td>2.0</td>
<td>2.3(.2)</td>
<td>2.5(.2)</td>
<td>2.8(.2)</td>
<td>2.9(.1)</td>
<td>3.0(.0)</td>
<td>2.7</td>
</tr>
<tr>
<td>C6b</td>
<td>2.1(.1)</td>
<td>2.4(.1)</td>
<td>2.6(.1)</td>
<td>2.3c</td>
<td>2.8(.1)</td>
<td>2.9(.1)</td>
<td>3.0(.0)</td>
<td>3.0(.0)</td>
<td>3.0(.0)</td>
<td>2.9e</td>
<td>25</td>
</tr>
<tr>
<td>C7</td>
<td>1.8(.2)</td>
<td>2.7(.1)</td>
<td>2.9(.1)</td>
<td>2.5c, d</td>
<td>3.0(.1)</td>
<td>3.0(.0)</td>
<td>3.0(.0)</td>
<td>3.0(.0)</td>
<td>3.0(.0)</td>
<td>3.0e</td>
<td>19</td>
</tr>
<tr>
<td>C8</td>
<td>2.5(.2)</td>
<td>2.7(.1)</td>
<td>2.9(.1)</td>
<td>2.8c, d,f</td>
<td>2.9(.1)</td>
<td>3.0(.0)</td>
<td>2.9(.1)</td>
<td>3.0(.0)</td>
<td>3.0(.0)</td>
<td>3.0e</td>
<td>11</td>
</tr>
<tr>
<td>C9</td>
<td>2.1(.2)</td>
<td>2.4(.2)</td>
<td>2.9(.1)</td>
<td>2.4c</td>
<td>3.0(.1)</td>
<td>3.0(.1)</td>
<td>2.8(.1)</td>
<td>3.0(.0)</td>
<td>3.0(.0)</td>
<td>2.9e</td>
<td>11</td>
</tr>
</tbody>
</table>

a Group main-effect and group x days (days 6-10 and 12-20), all p < 0.001; group x sex and group x sex x days, ns.

b Procedural change: Definition of circling behavior redefined.

c p < 0.01 compared to groups C2-C5 across days and sexes by Duncan's test.

d p < 0.05 compared to group C6 across days and sexes by Duncan's test.

e p < 0.01 compared to groups C2, C5 across days and sexes by Duncan's test.

f p < 0.01 compared to groups C6, C7, C9 across days and sexes by Duncan's test.
Table VI.10. Swimming Angle Ontogeny in Males

<table>
<thead>
<tr>
<th>Group</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>6-10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>12-20</th>
<th>No. of Litters</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0.2 (.1)</td>
<td>0.7 (.1)</td>
<td>1.3 (.2)</td>
<td>-</td>
<td>1.9 (.2)</td>
<td>2.9 (.1)</td>
<td>2.9 (.1)</td>
<td>3.0 (.0)</td>
<td>3.0 (.0)</td>
<td>-</td>
<td>20</td>
</tr>
<tr>
<td>C2</td>
<td>0.1 (.1)</td>
<td>0.7 (.1)</td>
<td>1.6 (.2)</td>
<td>0.8 (^b)</td>
<td>2.4 (.2)</td>
<td>3.4 (.1)</td>
<td>3.6 (.1)</td>
<td>3.9 (.1)</td>
<td>4.0 (.0)</td>
<td>3.4 (^c)</td>
<td>19</td>
</tr>
<tr>
<td>C3</td>
<td>0.2 (.1)</td>
<td>1.1 (.1)</td>
<td>2.0 (.1)</td>
<td>1.1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>17</td>
</tr>
<tr>
<td>C4</td>
<td>0.4 (.2)</td>
<td>1.3 (.2)</td>
<td>2.6 (.1)</td>
<td>1.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>15</td>
</tr>
<tr>
<td>C5</td>
<td>0.3 (.1)</td>
<td>0.9 (.2)</td>
<td>2.7 (.2)</td>
<td>1.3</td>
<td>3.1 (.3)</td>
<td>3.7 (.2)</td>
<td>3.9 (.1)</td>
<td>4.0 (.0)</td>
<td>4.0 (.0)</td>
<td>3.8</td>
<td>6</td>
</tr>
<tr>
<td>C6</td>
<td>0.4 (.1)</td>
<td>1.6 (.1)</td>
<td>2.2 (.1)</td>
<td>1.4</td>
<td>2.9 (.1)</td>
<td>3.8 (.1)</td>
<td>4.0 (.0)</td>
<td>4.0 (.0)</td>
<td>4.0 (.0)</td>
<td>3.8</td>
<td>25</td>
</tr>
<tr>
<td>C7</td>
<td>0.2 (.1)</td>
<td>1.2 (.2)</td>
<td>2.1 (.1)</td>
<td>1.2</td>
<td>3.0 (.2)</td>
<td>3.6 (.1)</td>
<td>3.9 (.6)</td>
<td>3.9 (.1)</td>
<td>4.0 (.0)</td>
<td>3.7</td>
<td>19</td>
</tr>
<tr>
<td>C8</td>
<td>0.3 (.1)</td>
<td>0.8 (.2)</td>
<td>1.9 (.2)</td>
<td>1.0</td>
<td>2.7 (.3)</td>
<td>3.6 (.2)</td>
<td>3.7 (.3)</td>
<td>4.0 (.0)</td>
<td>4.0 (.0)</td>
<td>3.6</td>
<td>11</td>
</tr>
<tr>
<td>C9(^d)</td>
<td>0.2 (.1)</td>
<td>0.6 (.2)</td>
<td>1.5 (.2)</td>
<td>0.7 (^b)</td>
<td>2.2 (.3)</td>
<td>3.1 (.2)</td>
<td>3.4 (.3)</td>
<td>3.9 (.1)</td>
<td>4.0 (.0)</td>
<td>3.3 (^c)</td>
<td>11</td>
</tr>
</tbody>
</table>

\(^a\) Group main-effect and group x days (days 6-10 and 12-20), all p < 0.001; group x sex and group x sex x days, ns.

\(^b\) p < 0.01 compared to groups C4-C7 across days and sexes by Duncan's test.

\(^c\) p < 0.05 to 0.01 compared to any other group across days and sexes by Duncan's test.

\(^d\) Note that unlike all other groups, the C9 group was injected with saline during gestation.
Table VI.11. Swimming Paddling Ontogeny in Males

<table>
<thead>
<tr>
<th>Group</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>6-10</th>
<th>12</th>
<th>14</th>
<th>16</th>
<th>18</th>
<th>20</th>
<th>12-20</th>
<th>No. of Litters</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0.8(.1)</td>
<td>1.0(.0)</td>
<td>1.0(.0)</td>
<td>-</td>
<td>1.0(.0)</td>
<td>1.0(.0)</td>
<td>1.3(.1)</td>
<td>1.6(.1)</td>
<td>1.8(.1)</td>
<td>-</td>
<td>19</td>
</tr>
<tr>
<td>C2</td>
<td>0.8(.1)</td>
<td>0.9(.0)</td>
<td>1.0(.0)</td>
<td>0.9\textsuperscript{b}</td>
<td>1.0(.0)</td>
<td>1.0(.0)</td>
<td>1.3(.1)</td>
<td>1.8(.1)</td>
<td>2.0(.0)</td>
<td>1.8</td>
<td>19</td>
</tr>
<tr>
<td>C3</td>
<td>0.9(.1)</td>
<td>1.0(.0)</td>
<td>1.0(.0)</td>
<td>1.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>18</td>
</tr>
<tr>
<td>C4</td>
<td>0.9(.1)</td>
<td>1.0(.0)</td>
<td>1.0(.0)</td>
<td>1.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>13</td>
</tr>
<tr>
<td>C5</td>
<td>1.0(.0)</td>
<td>1.0(.0)</td>
<td>1.0(.0)</td>
<td>1.0</td>
<td>1.0(.0)</td>
<td>1.1(.1)</td>
<td>1.5(.1)</td>
<td>1.8(.1)</td>
<td>1.8(.1)</td>
<td>1.9</td>
<td>6</td>
</tr>
<tr>
<td>C6\textsuperscript{c}</td>
<td>0.9(.0)</td>
<td>1.0(.0)</td>
<td>1.0(.0)</td>
<td>1.0</td>
<td>1.1(.0)</td>
<td>1.2(.1)</td>
<td>1.7(.1)</td>
<td>2.1(.1)</td>
<td>2.3(.1)</td>
<td>1.7\textsuperscript{d}</td>
<td>25</td>
</tr>
<tr>
<td>C7</td>
<td>0.9(.0)</td>
<td>1.0(.0)</td>
<td>1.0(.0)</td>
<td>1.0</td>
<td>1.0(.0)</td>
<td>1.2(.1)</td>
<td>1.6(.1)</td>
<td>2.1(.1)</td>
<td>2.3(.1)</td>
<td>1.7\textsuperscript{d}</td>
<td>19</td>
</tr>
<tr>
<td>C8</td>
<td>1.0(.1)</td>
<td>1.0(.0)</td>
<td>1.0(.0)</td>
<td>1.0</td>
<td>1.1(.1)</td>
<td>1.1(.1)</td>
<td>1.8(.2)</td>
<td>1.9(.1)</td>
<td>2.2(.1)</td>
<td>1.6\textsuperscript{d}</td>
<td>11</td>
</tr>
<tr>
<td>C9</td>
<td>0.9(.1)</td>
<td>0.9(.1)</td>
<td>1.0(.0)</td>
<td>0.9</td>
<td>1.0(.0)</td>
<td>1.1(.1)</td>
<td>1.4(.2)</td>
<td>1.8(.2)</td>
<td>2.5(.2)</td>
<td>1.6\textsuperscript{d}</td>
<td>11</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Group main-effect (days 6-10 and 12-20), both \(p < 0.01\); group \(\times\) days (days 6-10), \(p < 0.05\); (days 12-20), \(p < 0.001\); group \(\times\) sex and group \(\times\) sex \(\times\) days, ns.

\textsuperscript{b} \(p < 0.05\) compared to groups C4, C6 across days and sexes by Duncan's test.

\textsuperscript{c} Procedural change: Paddling scale expanded with intermediate level of performance between scores 1-2 defined as 2, and original 2 redefined as a score of 3.

\textsuperscript{d} \(p < 0.05\) compared to group C5 across days and sexes by Duncan's test.
Table VI.12. Preweaning Open-Field Ambulation in Males

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SE) Section Entries&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. of Litters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 1</td>
<td>Day 2</td>
</tr>
<tr>
<td>C2</td>
<td>39.5 (5.3)</td>
<td>42.2 (6.2)</td>
</tr>
<tr>
<td>C3</td>
<td>45.2 (6.7)</td>
<td>45.9 (4.2)</td>
</tr>
<tr>
<td>C4</td>
<td>36.4 (7.1)</td>
<td>49.9 (7.7)</td>
</tr>
<tr>
<td>C5</td>
<td>37.7 (11.4)</td>
<td>45.3 (13.2)</td>
</tr>
<tr>
<td>C6</td>
<td>24.5 (4.0)</td>
<td>27.5 (3.0)</td>
</tr>
<tr>
<td>C7&lt;sup&gt;c&lt;/sup&gt;</td>
<td>35.9 (4.3)</td>
<td>54.3 (7.3)</td>
</tr>
<tr>
<td>C8</td>
<td>32.0 (6.9)</td>
<td>45.6 (7.3)</td>
</tr>
<tr>
<td>C9</td>
<td>28.2 (8.5)</td>
<td>40.7 (10.1)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Group main-effect, p <0.01; group x days, p <0.05; group x sex and group x sex x days, ns.
<sup>b</sup> p <0.05 compared to groups C3, C4 across days and sexes by Duncan's test.
<sup>c</sup> Procedural change: Rats placed in a cylinder at start of test beginning with this group and thereafter.
Table VI.13. Preweaning Open-Field Rearing Frequency in Males

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SE) Rears(^a)</th>
<th>No. of Litters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 1</td>
<td>Day 2</td>
</tr>
<tr>
<td>C2</td>
<td>8.1 (1.0)</td>
<td>8.2 (1.8)</td>
</tr>
<tr>
<td>C3</td>
<td>4.5 (0.9)</td>
<td>6.4 (1.2)</td>
</tr>
<tr>
<td>C4</td>
<td>5.7 (1.1)</td>
<td>6.1 (0.9)</td>
</tr>
<tr>
<td>C5</td>
<td>5.0 (2.1)</td>
<td>2.7 (1.2)</td>
</tr>
<tr>
<td>C6</td>
<td>5.9 (0.9)</td>
<td>6.9 (0.8)</td>
</tr>
<tr>
<td>C7(^b)</td>
<td>5.6 (1.2)</td>
<td>8.0 (1.1)</td>
</tr>
<tr>
<td>C8</td>
<td>4.0 (1.2)</td>
<td>5.8 (1.9)</td>
</tr>
<tr>
<td>C9</td>
<td>5.1 (1.1)</td>
<td>4.0 (1.3)</td>
</tr>
</tbody>
</table>

\(^a\) Group main-effect, p < 0.01; all group-related interaction, ns.
\(^b\) Procedural change: Rats placed in cylinder at start of test beginning with this group and thereafter.
\(^c\) p < 0.01 compared to group C2 across days and sexes by Duncan's test.
### Table VI.14. Postweaning Open-Field Ambulation in Males

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SE) Section Entries&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. of Litters</th>
<th>Day 1 and 2 M and F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 1</td>
<td>Day 2</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>66.9 (3.9)</td>
<td>57.6 (6.0)</td>
<td>18</td>
</tr>
<tr>
<td>C2</td>
<td>56.4 (5.2)</td>
<td>46.3 (4.8)</td>
<td>16</td>
</tr>
<tr>
<td>C3</td>
<td>57.3 (5.8)</td>
<td>53.3 (4.2)</td>
<td>17</td>
</tr>
<tr>
<td>C4</td>
<td>68.2 (4.6)</td>
<td>62.3 (5.7)</td>
<td>14</td>
</tr>
<tr>
<td>C5</td>
<td>74.8 (12.0)</td>
<td>58.8 (6.0)</td>
<td>6</td>
</tr>
<tr>
<td>C6</td>
<td>63.4 (3.1)</td>
<td>61.1 (4.5)</td>
<td>25</td>
</tr>
<tr>
<td>C7</td>
<td>72.9 (4.5)</td>
<td>69.0 (6.6)</td>
<td>15</td>
</tr>
<tr>
<td>C8</td>
<td>67.2 (5.1)</td>
<td>64.9 (7.1)</td>
<td>12</td>
</tr>
<tr>
<td>C9</td>
<td>69.9 (5.0)</td>
<td>72.4 (7.6)</td>
<td>8</td>
</tr>
</tbody>
</table>

<sup>a</sup> Group main-effect, p < 0.001; all group-related interactions, ns.

<sup>b</sup> p < 0.05 compared to group C9 by Duncan's test.

<sup>c</sup> p < 0.05 compared to groups C4, C6, C7, C9 by Duncan's test.
Table VI.15. Postweaning Open-Field Rearing Frequency in Males

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean (SE) Rears&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. of Litters</th>
<th>Day 1 and 2 M and F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Day 1</td>
<td>Day 2</td>
<td></td>
</tr>
<tr>
<td>C1</td>
<td>18.8 (1.0)</td>
<td>16.6 (2.1)</td>
<td>18</td>
</tr>
<tr>
<td>C2</td>
<td>16.5 (2.0)</td>
<td>10.4 (1.9)</td>
<td>16</td>
</tr>
<tr>
<td>C3</td>
<td>13.5 (1.5)</td>
<td>10.8 (1.2)</td>
<td>17</td>
</tr>
<tr>
<td>C4</td>
<td>16.4 (1.9)</td>
<td>14.5 (2.4)</td>
<td>14</td>
</tr>
<tr>
<td>C5</td>
<td>12.8 (4.4)</td>
<td>8.3 (1.2)</td>
<td>6</td>
</tr>
<tr>
<td>C6</td>
<td>21.5 (2.4)</td>
<td>16.8 (2.5)</td>
<td>25</td>
</tr>
<tr>
<td>C7</td>
<td>22.7 (3.0)</td>
<td>20.6 (3.9)</td>
<td>15</td>
</tr>
<tr>
<td>C8</td>
<td>16.3 (1.8)</td>
<td>17.3 (3.7)</td>
<td>12</td>
</tr>
<tr>
<td>C9</td>
<td>16.3 (2.4)</td>
<td>14.0 (2.5)</td>
<td>8</td>
</tr>
</tbody>
</table>

<sup>a</sup> Group main-effect, p < 0.001; all group-related interactions, ns.

<sup>b</sup> p < 0.05 compared to group C6 by Duncan's test.

<sup>c</sup> p < 0.01 compared to groups C6, C7 by Duncan's test.
Table VI.16. Passive Avoidance

<table>
<thead>
<tr>
<th>Group</th>
<th>Males</th>
<th>No. of Litters</th>
<th>Females</th>
<th>No. of Litters</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>16.9 (4.3)</td>
<td>11</td>
<td>21.3 (6.7)</td>
<td>13</td>
</tr>
<tr>
<td>C2</td>
<td>19.2 (2.8)</td>
<td>25</td>
<td>19.9 (4.0)</td>
<td>27</td>
</tr>
<tr>
<td>C3</td>
<td>10.4 (3.3)</td>
<td>10</td>
<td>8.8 (2.2)</td>
<td>10</td>
</tr>
<tr>
<td>C4</td>
<td>12.0 (2.5)</td>
<td>15</td>
<td>11.2 (2.1)</td>
<td>12</td>
</tr>
<tr>
<td>C5</td>
<td>14.5 (3.6)</td>
<td>6</td>
<td>17.5 (5.6)</td>
<td>6</td>
</tr>
<tr>
<td>C6</td>
<td>17.5 (7.0)</td>
<td>25</td>
<td>14.7 (2.7)</td>
<td>21</td>
</tr>
<tr>
<td>C7</td>
<td>18.1 (4.7)</td>
<td>14</td>
<td>9.8 (1.9)</td>
<td>14</td>
</tr>
<tr>
<td>C8</td>
<td>10.6 (3.1)</td>
<td>12</td>
<td>12.4 (1.7)</td>
<td>10</td>
</tr>
<tr>
<td>C9</td>
<td>9.4 (2.3)</td>
<td>8</td>
<td>11.5 (2.8)</td>
<td>9</td>
</tr>
</tbody>
</table>

*a* All effects, ns.
Table VI.17. Day 90 Body and Brain Weights for Males (Mean and SE in Grams)

<table>
<thead>
<tr>
<th>Group</th>
<th>Body Weight&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Cerebellum&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Brain Stem&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Cerebral Cortex&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Eyes&lt;sup&gt;b&lt;/sup&gt;</th>
<th>No. of Offspring</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>347.7 (10.5)</td>
<td>.299 (.006)</td>
<td>.205 (.004)</td>
<td>1.442 (.025)</td>
<td>.255 (.003)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>15</td>
</tr>
<tr>
<td>C2</td>
<td>342.5 (5.2)</td>
<td>.294 (.006)</td>
<td>.214 (.004)</td>
<td>1.449 (.020)</td>
<td>.259 (.002)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>15</td>
</tr>
<tr>
<td>C3</td>
<td>355.8 (8.1)</td>
<td>.286 (.005)</td>
<td>.215 (.005)</td>
<td>1.471 (.016)</td>
<td>.268 (.003)</td>
<td>16</td>
</tr>
<tr>
<td>C4</td>
<td>349.3 (5.5)</td>
<td>.292 (.004)</td>
<td>.224 (.005)</td>
<td>1.449 (.016)</td>
<td>.270 (.003)</td>
<td>15</td>
</tr>
<tr>
<td>C5</td>
<td>359.5 (13.7)</td>
<td>.271 (.008)</td>
<td>.196 (.009)&lt;sup&gt;e&lt;/sup&gt;</td>
<td>1.453 (.025)</td>
<td>.262 (.004)</td>
<td>16</td>
</tr>
<tr>
<td>C6</td>
<td>337.9 (8.0)</td>
<td>.275 (.005)</td>
<td>.208 (.005)</td>
<td>1.415 (.020)</td>
<td>.257 (.003)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>18</td>
</tr>
<tr>
<td>C7</td>
<td>371.5 (12.5)</td>
<td>.281 (.008)</td>
<td>.223 (.006)</td>
<td>1.420 (.030)</td>
<td>.268 (.003)</td>
<td>14</td>
</tr>
<tr>
<td>C8</td>
<td>374.2 (14.9)</td>
<td>.264 (.007)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.200 (.007)</td>
<td>1.432 (.028)</td>
<td>.271 (.009)</td>
<td>12</td>
</tr>
<tr>
<td>C9</td>
<td>363.0 (10.9)</td>
<td>.278 (.004)</td>
<td>.207 (.009)</td>
<td>1.455 (.014)</td>
<td>.275 (.004)</td>
<td>7</td>
</tr>
</tbody>
</table>

<sup>a</sup> Group effect, ns.
<sup>b</sup> Group effect, p < 0.001.
<sup>c</sup> Group effect, p < 0.01.
<sup>d</sup> Group effect, p < 0.05 compared to groups C3, C4, and C7-C9 by Duncan's test.
<sup>e</sup> p < 0.05 compared to groups C2-C4 and C7 by Duncan's test.
<sup>f</sup> p < 0.05 compared to groups C1-C4 by Duncan's test.
in order for a test to be of value for detecting neurotoxicity. The intralaboratory reliability data clearly show that there is stability in behavioral measurements equal to that which is characteristic of other types of bioassays. Therefore, behavioral data should not be viewed as inherently less reliable than measures of physical characteristics.

2. Interlaboratory reliability

One of the prime goals of the NCTR Collaborative Behavioral Teratology Study was to evaluate interlaboratory reliability. Some representative data are presented here to highlight the important findings of this complex study.*

The Collaborative Project involved two experiments per laboratory using prenatal exposure to the putative positive control agents amphetamine (Study 1) and methylmercury (Study 2). Six laboratories within the United States participated in the project. Figure VI.1 shows intralaboratory and interlaboratory data for the untreated control groups of both studies from the 1-hour tests of locomotor activity as measured in a figure-8 shaped apparatus. This figure shows only females, but the data for males were equally reliable. The high degree of comparability between all laboratories for a given study and the remarkable degree of comparability within laboratories from Study 1 to Study 2 can be noted in this graphic presentation.

Preliminary data from a laboratory in Germany produced comparable data using the same activity testing apparatus, but with a modification of the study protocol (Ulbrich et al., 1985). These data are of interest because they challenge the belief that tests of spontaneous locomotor activity are among those most subject to variation due to differences across laboratories, personnel, and animal suppliers. The German data offer a unique opportunity for comparison as they not only followed the Collaborative Study using the exact duplicate apparatus, but also kept constant the same testing session length, data recording intervals, animal age at time of testing, and basic animal handling histories. With attention to these features, an excellent degree of between laboratory comparability was obtained which refutes the notion that almost any procedural variation has major effects on measures of behavior.

Figure VI.2 shows the results of the amphetamine study for a test of auditory startle. In this procedure, test sessions of 50 trials were given at each of several different ages.

* I am particularly indebted to Judy Buelke-Sam for her invaluable assistance in providing this information.
Figure VI.1. One-hour figure-8 monitor locomotor activity in female control rats in the Collaborative Behavioral Teratology Study. Data are mean photocell interruptions for the entire hour of testing at one of four postnatal ages. Animals are from both studies in the project and illustrate the high degree of consistency within and across laboratories during the 2 years of the live animal phase of the experiments. Reprinted with permission from Buelke-Sam et al., 1985.
Figure VI.2. Results from the Collaborative Behavioral Teratology Study for the auditory startle tests in adult female rats from the d-amphetamine experiment. These results show that despite small differences between laboratories in baselines, all labs found no effects attributable to prenatal d-amphetamine. The pattern of startle results was also similar across laboratories. Reprinted with permission from Buelke-Sam et al., 1985.
Responses were measured on each trial to the onset of a 120 dB tone. All laboratories reported no significant effects of prenatal amphetamine. While this finding does not prove that amphetamine may not produce adverse effects on behavior at higher doses and/or other periods of exposure than used in this study, it is reassuring to note that all laboratories correctly identified lack of effects. It is important to note that the Collaborative Study was carefully conducted with attention to blinding, so that no experimenter bias could affect the outcome. Although the baseline among controls was slightly different between laboratories, all laboratories were internally consistent. In addition, all laboratories had comparable degrees of variation around the mean values for each parameter measured regardless of baseline, so that the detection power of the test was similar for all laboratories.

Figure VI.3 depicts the auditory startle test for the methylmercury study. Not only did all laboratories detect a significant methylmercury-induced change in the evoked startle response, but they also found the change in the methylmercury treated animals to be increased and of similar magnitude. While there was some baseline variation between laboratories, this did not prevent the same pattern of effects from being seen by all. The observed startle response effect was clearly indicated at the higher dose of methylmercury, but was not observed at the lower dose in most laboratories. The low dose of methylmercury was not sufficient to induce overt toxicity as it was explicitly selected to meet a minimal toxicity specification defined by project design (Adams et al., 1985a; Kimmel and Buelke-Sam, 1985). The terms high and low dose are relative only to this study since both doses are not generally thought to be neurotoxic. Preliminary analyses have shown that the startle response data from the German laboratory was consistent with those of the Collaborative Study laboratories in that startle response facilitation was observed from prenatal methylmercury exposure.

Examination of the Collaborative Study data analyses offers sufficient evidence to demonstrate that behavioral measures can be conducted with a high degree of interlaboratory reliability when attention is paid to standardization of testing procedures and training of personnel. This conclusion is in contrast to the fact that 4 years ago many regulatory authorities and investigators who were not familiar with behavioral teratology methods questioned whether behavioral assessments could meet such rigorous standards.

A note at this point about control groups is in order. The data just presented on reliability demonstrate the importance of having baseline data from historically tested controls. Nevertheless, as with all bioassays, tests of behavioral dysfunction must include groups of contemporaneous controls for comparison.
Figure VI.3. Results from the Collaborative Behavioral Teratology Study for the auditory startle test for female rats from the methylmercury experiment. The results show that all laboratories correctly detected that the prenatal 6 mg/kg dose of methylmercury facilitated startle responses. Reprinted with permission from Buelke-Sam et al., 1985.
B. SENSITIVITY OF BEHAVIORAL METHODS

Within the context of developmental toxicity, the sensitivity of behavioral test methods may relate to (1) CNS integrity in relation to other methods of detecting developmental toxicity, (2) the ability of the test to respond to particular types of dysfunction, or (3) the detection sensitivity of measurements based on error variance (i.e., how small a change can the test detect as significant).

1. Sensitivity of behavior in relation to other measures of developmental toxicity

A review of literature sources for measures of developmental toxicity has been prepared by Vorhees (1986a). This review indicated that the basic relationship between the major indices of developmental toxicity (embryo-fetal death, teratogenic malformations, growth abnormality, and behavioral dysfunction) appear as illustrated in Figure VI.4. Both the shape and spacing of the curves are strictly hypothetical and should not be taken to be representative of a particular test agent. The four curves presented are drawn as idealized dose-response curves, arbitrarily depicted as symmetrical. The relationships represented are for a situation in which a toxic agent is able to produce the full range of adverse effects; however, the depiction is not intended to encompass every possible situation. There are numerous examples of chemical agents where the graphic presentations for behavioral and growth abnormality appear to be superimposed. The neuroleptic agent prochlorperazine is exemplary of this phenomenon (Vorhees et al., 1979b), while other drugs such as acetazolamide, may not produce behavioral dysfunction at any reasonable dose (Butcher et al., 1975a,b). The fundamental principle illustrated in Figure VI.4 is that drugs and other chemicals frequently produce all four types of damage. When this occurs it is usually the case that behavioral dysfunction has proven to be the most sensitive index of developmental toxicity in comparison to growth abnormality, teratologic malformation, or embryo-fetal death.

At this time there is insufficient data to permit a graphic depiction of the more subtle neurochemical, neuropathological, or neurophysiological indices in comparison to behavioral data. The reader is referred to the paper by Mailman in the previous chapter for a detailed discussion of the limits to neurochemical screening for neurotoxicity.

A list of those classes of compounds which should routinely be considered for possible behavioral teratogenicity was developed by participants at the recent Collaborative Behavioral Teratology Study Symposium. The chemical classes of concern include: (1) central nervous system (CNS) teratogens;
Figure VI.4. Theoretical dose-response curves for the four major manifestations of developmental toxicity. Note that if a chemical produces all four categories of effects, then the curve for behavioral dysfunction is expected to be farthest to the left in comparison with other curves based on a review of the current experimental literature (Vorhees, 1986a).
(2) adult neurotoxic agents; (3) psychoactive agents; (4) pesticides; (5) solvents; and (6) peptides. While this list constitutes a reasonable beginning, it is hoped that eventually structure-activity relationships may aid in the analysis of new agents that should be screened for their potential to cause behavioral and neurologic dysfunction. The breadth of the list attests to the widespread potential of many commonly used chemicals to adversely affect CNS function.

2. Test sensitivity for specific toxic effects

The selectivity of tests used for behavioral and neurologic dysfunction have been discussed primarily in terms of apical or nonapical functional coverage (Butcher, 1976; Laties et al., 1977; Vorhees, 1986a; Vorhees and Butcher, 1982). An apical test is one that relies on the use of higher order cognitive functions as a measure of successful performance. In contrast to nonapical test methods, an apical test is frequently more complex for the test subject to perform and involves more information processing. It is generally accepted that apical tests require increased use of CNS subsystems and greater integration of processes in order to be successfully completed. In practice, this dichotomization into apical and nonapical tests becomes blurred because the behavior of the organism is the final common output of CNS and sensorimotor integration and all tests of behavior are apical to one degree or another. It is more appropriate to refer to the varying degree to which a particular test is considered to be apical.

Simple reflex testing methods require the coordination of fewer CNS subsystems than tests such as learning. For this reason, tests of learning are considered to be more apical than tests of simple reflexes. It has been suggested that the more apical a test, the better suited it is for toxicologic screening purposes. This concept is based on the belief that damage to any one subsystem upon which an apical behavior relies will disrupt the apical behavior pattern. Since the most apical testing methods reflect the integrative functioning of multiple subsystems, fewer tests would presumably suffice to monitor a major portion of CNS function. This reasoning may be flawed because the CNS possesses both functional reserve and plasticity. Even after substantial damage, the CNS may be able to accommodate to injury in a manner that precludes the detection of readily revealed underlying losses (Finger and Stein, 1982). Such compensation or accommodation is most likely to occur for those functions measured by apical tests because these tests draw upon a greater number of overlapping subsystems each of which may have its own reserve capacity and plasticity. The collective effect may be a false negative test result in that the apical behavior can be maintained even though the brain has sustained substantial tissue damage. An optimal test battery would be one which includes a mixture of highly apical and less
apical tests. Such a battery would include measures of complex behaviors such as spontaneous locomotor activity and learning as well as measures of simple behaviors such as reflexes.

It is difficult to choose test methods for high specificity (i.e., tests that are less apical in nature) because the specificity of many behavioral tests are not completely understood. Experience has shown that some tests are selectively more sensitive than others. For example, it has been shown that complex swimming maze tests, such as the Biel water maze, have correctly identified a number of behavioral teratogens and have correctly detected some nonbehavioral teratogens as negative. This test appears to be selective for certain types of behavioral teratogens and has proven to be accurate as a test for detecting behavioral teratogenicity (Vorhees, 1985a,b).

3. Detection sensitivity of behavioral measurements

As this subject has been extensively reviewed for the Collaborative Behavioral Teratology Study, the reader is referred to the discussion of sensitivity in that project (Buelke-Sam et al., 1985). The focus of this discussion will be on two statistical methods for numerically describing test sensitivity.

The previous index most widely used to reflect the relationship of the variability of a test to its central tendency has been the coefficient of variation (CV):

\[ CV = \frac{sd}{X} \times 100 \]

This value, expressed as a percentage, is a means of describing the extent of variance associated with a measurement as reflected by the inverse relationship between the standard deviation (sd), and the mean (\( X \)) (Butcher et al., 1980).

The Collaborative Behavioral Teratology Study used a different index for describing sensitivity. This value has been termed the coefficient of detection (CD) (Vorhees, 1985c) and indicates the percent change required to detect a significant change using a given alpha (e.g., 0.05). It is defined as:

\[ CD = \frac{(SE \times t_\alpha)}{X} \times 100 \]

For this equation, the standard error of the mean (SE) is used as the measure of variation and the interval is determined by \( t_\alpha \) with known or assumed N (number of subjects) for a two-tailed distribution in relation to the mean (\( X \)) and expressed as a percentage. The numerator forms the basis for computing a confidence interval, but unlike a descriptive confidence interval
from sample data, which uses the standard deviation, this confidence interval is intended for predicting the next expected group mean and therefore uses the standard error. A CV value can be converted to a CD as follows:

\[
CD = (\frac{t_{\alpha}}{N}) \cdot CV
\]

Table VI.18 shows the CV and CD values using the ANOVA standard deviation and grand mean for all the measures presented in Tables VI.2-VI.17. The CD values for behavioral measures in Table VI.18 range from a low of 4.4 to a high of 59.5.

A comparison of these values for other test methods is also shown in Tables VI.19-VI.21. These tables (VI.19-VI.21) compare results from the Collaborative Study with an alternative test system (the Cincinnati Test Battery) which was run as a parallel study (Vorhees, 1985a,b,c). Both studies used animals derived from the same breeding colony. The CD values for auditory startle ranged from 4.4 to 38.4 (Table VI.19), while the figure-8 maze activity ranged from 5.6 to 36.4 (Table VI.20), and the tests of instrumental learning ranged from 2.5 to 25.7 (Table VI.21). Note that most of the values obtained were well below the upper bounds of these ranges, indicating that these tests are generally less variable than the high end of the range implies. For example, a typical value for startle development from studies conducted over a 5-year period in the Cincinnati laboratory was 4.4%. The Collaborative Study found values around 15% for startle habituation across all participating laboratories and test chemicals studied. For postweaning, figure-8 activity using different protocols, both the Cincinnati Study and the Collaborative Study found values for ambulation scores of 10-15%. The Collaborative Study operant conditioning test of learning found CD values around 10-20%, while the water maze test had CD values of around 16%. Such values are significant because they imply that it would take a mean value within an experimental group that is different by these percentages to fall outside the 95% confidence interval based on the control data and assuming a group size of 16. The group size of 16 was the mean control group size across all 5 years of the Cincinnati Study and was the mandated group size in the Collaborative Study. Experience has shown that although these group sizes are below those required by FDA for most preclinical toxicity testing, they are sufficient to detect reasonable degrees of behavioral and neurologic dysfunction. The detection sensitivity of these tests would be improved if the number of animals for each group were increased to the FDA minimum of 20.

If one were considering a multigroup study then the percentage change required to detect an effect as significant would be slightly larger than these CD values suggest. Nevertheless, these values provide an index of test sensitivity suggesting that behavioral tests can exhibit good sensitivity when attention is given to standardization of the testing
Table VI.18. Summary of Study Findings Across All Control Groups Presented in Tables VI.2-VI.17

<table>
<thead>
<tr>
<th>BEHAVIORAL MEASURE</th>
<th>CV&lt;sup&gt;a&lt;/sup&gt;</th>
<th>CD&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of Gestation</td>
<td>3.5</td>
<td>1.8</td>
</tr>
<tr>
<td>Preweaning Body Weight</td>
<td>25.1</td>
<td>13.3</td>
</tr>
<tr>
<td>Vaginal Patency</td>
<td>11.4</td>
<td>6.0</td>
</tr>
<tr>
<td>Surface Righting</td>
<td>17.8</td>
<td>9.4</td>
</tr>
<tr>
<td>Auditory Startle Development</td>
<td>8.3</td>
<td>4.4</td>
</tr>
<tr>
<td>Pivoting Locomotion</td>
<td>79.8</td>
<td>42.3</td>
</tr>
<tr>
<td>Negative Geotaxis</td>
<td>64.6</td>
<td>34.2</td>
</tr>
<tr>
<td>Swimming Ontogeny--Direction</td>
<td>24.3</td>
<td>12.9</td>
</tr>
<tr>
<td>Swimming Ontogeny--Angle</td>
<td>62.1</td>
<td>32.9</td>
</tr>
<tr>
<td>Swimming Ontogeny--Paddling</td>
<td>14.4</td>
<td>7.6</td>
</tr>
<tr>
<td>Preweaning Open-Field Ambulation</td>
<td>65.7</td>
<td>34.8</td>
</tr>
<tr>
<td>Preweaning Open-Field Rearing</td>
<td>91.8</td>
<td>48.6</td>
</tr>
<tr>
<td>Postweaning Open-Field Ambulation</td>
<td>36.2</td>
<td>19.2</td>
</tr>
<tr>
<td>Postweaning Open-Field Rearing</td>
<td>64.9</td>
<td>34.4</td>
</tr>
<tr>
<td>Passive Avoidance</td>
<td>112.2</td>
<td>59.5</td>
</tr>
<tr>
<td>Day 90 Body Weight--Males</td>
<td>10.3</td>
<td>5.4</td>
</tr>
<tr>
<td>Day 90 Cerebellum Weight--Males</td>
<td>7.6</td>
<td>4.0</td>
</tr>
<tr>
<td>Day 90 Brain Stem Weight--Males</td>
<td>9.7</td>
<td>5.1</td>
</tr>
<tr>
<td>Day 90 Cortex Weight--Males</td>
<td>8.3</td>
<td>4.4</td>
</tr>
<tr>
<td>Day 90 Eye Weight--Males</td>
<td>4.3</td>
<td>2.3</td>
</tr>
</tbody>
</table>

<sup>a</sup> CV is the ANOVA standard deviation for all groups divided by the grand mean and multiplied times 100.

<sup>b</sup> CD is the CV times t<sub>α</sub> at .05 two-tailed or 2.131 with df = 15, divided by the square root of N = 16.
Table VI.19. Comparison of Coefficients of Detection Between Test Batteries for Tests of Reactivity

<table>
<thead>
<tr>
<th>TEST</th>
<th>d-amphetamine</th>
<th>methylmercury</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NCTR</td>
<td>Cincinnati</td>
</tr>
<tr>
<td>Auditory Startle Habituation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Days 18-19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>13.1-17.4</td>
<td>-</td>
</tr>
<tr>
<td>Female</td>
<td>9.9-22.0</td>
<td>-</td>
</tr>
<tr>
<td>Days 57-58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>12.6-22.6</td>
<td>-</td>
</tr>
<tr>
<td>Female</td>
<td>12.8-19.5</td>
<td>-</td>
</tr>
<tr>
<td>Auditory Startle Emergence</td>
<td>-</td>
<td>4.4</td>
</tr>
</tbody>
</table>

\(^a\) Collaborative Behavioral Teratology Study test battery.
\(^b\) Cincinnati Study test battery.
<table>
<thead>
<tr>
<th>Test Battery</th>
<th>Period</th>
<th>Length of Test</th>
<th>Dependent Measure</th>
<th>Test Condition</th>
<th>CD&lt;sup&gt;a&lt;/sup&gt;</th>
<th>methylmercury</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Age&lt;sup&gt;b&lt;/sup&gt;</td>
<td>M and F</td>
<td></td>
</tr>
<tr>
<td>Cinn.</td>
<td>Preweaning</td>
<td>15 min/d</td>
<td>Ambulation</td>
<td>15:</td>
<td>28.9</td>
<td>36.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16:</td>
<td>20.7</td>
<td>30.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>17:</td>
<td>16.4</td>
<td>21.2</td>
</tr>
<tr>
<td>Cinn.</td>
<td>Adolescence</td>
<td>15 min/d</td>
<td>Ambulation</td>
<td>40:</td>
<td>11.3</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>41:</td>
<td>14.6</td>
<td>11.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>42:</td>
<td>15.3</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Rearing</td>
<td>40:</td>
<td>29.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>41:</td>
<td>22.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>42:</td>
<td>24.1</td>
</tr>
<tr>
<td>NCTR</td>
<td>day 21 and 60</td>
<td>1 hr</td>
<td>Ambulation</td>
<td></td>
<td>5.8-10.5</td>
<td>5.6-9.2</td>
</tr>
<tr>
<td>NCTR</td>
<td>Adult (~ day 100)</td>
<td>23 hr</td>
<td>Ambulation</td>
<td>Light&lt;sup&gt;e&lt;/sup&gt;:</td>
<td>10.6-18.7</td>
<td>13.2-19.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Dark:</td>
<td>6.7-15.1</td>
<td>11.2-19.7</td>
</tr>
<tr>
<td>NCTR</td>
<td>Adult (~ day 125)</td>
<td>4 hr</td>
<td>Ambulation</td>
<td>Pre&lt;sup&gt;f&lt;/sup&gt; D1&lt;sup&gt;g&lt;/sup&gt;:</td>
<td>7.4-12.9</td>
<td>6.1-11.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D2:</td>
<td>10.4-22.0</td>
<td>12.2-28.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Post D1:</td>
<td>10.1-14.3</td>
<td>6.5-12.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D2:</td>
<td>14.1-17.9</td>
<td>9.4-19.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>F</th>
<th></th>
<th>M</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6.2-7.7</td>
<td>6.6-8.0</td>
<td>8.9-14.8</td>
<td>9.1-14.6</td>
<td></td>
</tr>
</tbody>
</table>

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Post D1:</td>
<td>16.4-27.8</td>
<td>15.1-26.0</td>
</tr>
<tr>
<td>D2:</td>
<td>15.1-26.0</td>
<td>12.9-19.6</td>
</tr>
</tbody>
</table>

<sup>a</sup> Coefficient of detection.
<sup>b</sup> Age at testing in weeks.
<sup>c</sup> Cincinnati Study test battery.
<sup>d</sup> Collaborative Behavioral Teratology Study battery.
<sup>e</sup> Light/dark refers to diurnal or nocturnal phase of the light cycle.
<sup>f</sup> Pre/post refers to whether testing was before or after injection with a pharmacological challenge dose of d-amphetamine.
<sup>g</sup> D1/D2 refers to whether the rats received the low (D1) or high (D2) dose of the d-amphetamine challenge.
Table VI.21. Comparison of Coefficients of Detection Between Test Batteries for Tests of Learning and Memory

<table>
<thead>
<tr>
<th>Task</th>
<th>Test Battery</th>
<th>Test Phase</th>
<th>Response(^a)</th>
<th>Study 1(^b)</th>
<th>Study 2(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operant Conditioning</td>
<td>NCTR(^d)</td>
<td>Acquisition</td>
<td>Cor. Res. 4.1-14.0</td>
<td>3.9-8.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ITI Res. 8.8-14.9</td>
<td>6.3-13.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Discrimination</td>
<td>Cor. Res. 2.5-15.6</td>
<td>2.8-12.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ITI Res. 10.7-19.2</td>
<td>11.8-14.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reversal</td>
<td>Cor. Res. 12.7-21.4</td>
<td>5.9-21.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ITI Res. 11.9-25.7</td>
<td>11.5-21.0</td>
<td></td>
</tr>
<tr>
<td>Complex Water Maze</td>
<td>Cinn.(^e)</td>
<td>Straight Channel</td>
<td>Time 13.0</td>
<td>20.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maze Path A+B</td>
<td>Errors 15.4</td>
<td>19.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Maze Path A+B</td>
<td>Time 13.5</td>
<td>18.7</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Responses labeled Cor. Res. refer to total correct responses, and ITI Res. refer to responses occurring during the intertrial interval.

\(^b\) d-amphetamine.

\(^c\) Methylmercury.

\(^d\) Collaborative Behavioral Teratology Study test battery.

\(^e\) Cincinnati Study test battery.
apparatus, protocol design, and expertise of personnel. These are the same factors that influence test performance in any area of scientific investigation.

Behavioral tests should not be chosen solely on the basis of reliability and sensitivity as these criteria are not sufficient in and of themselves. There is information which indicates that some tests may be reliable and sensitive, but are not responsive to the influence of neurotoxic agents. Some behavioral tests possess small variances because they restrict the range of response that is measured and assess functions having high intrinsic or biological redundancy, such that considerable damage may be withstood without alteration of the observed behaviors (e.g., some reflex responses). A small variance does not imply that the test is optimally sensitive for detecting a dysfunction since it has been observed that animals with major CNS injury can show basic reflexes which remain unaffected (Vorhees, 1983a). This may occur because some reflexes appear to be essential to survival and abnormalities in the brain regions that control these basic functions would have to be so large in order to be detected, that survival of the organism would be seriously compromised. In developing tests with sufficient sensitivity it is important to have tests which (1) measure the performance of important CNS functional domains (i.e., learning, memory, activity, etc.); (2) provide an adequate range of graded response options (dose-response sensitivity); and (3) exhibit low variability. By adhering to these criteria, a good test would have a large treatment or effect-related variance and a small error variance. Such stringent criteria have been demonstrated for the behavioral tests devised in the Collaborative Study. Another example is shown in Figure VI.5 which illustrates data from a water maze test in adult rats prenatally exposed to varying doses of the anticonvulsant phenytoin (Vorhees, 1986b). As discussed elsewhere (Vorhees, 1986a), the lack of widely reported dose-response relationships in the behavioral teratology literature is an artifact of the study design and response measures utilized. Figure VI.5 shows that this is not a genuine reflection of some anomalous dose-response pattern for this phenomenon.

C. VALIDITY OF BEHAVIORAL METHODS

Anastasi (1982) and Cronbach (1970) describe three basic types of validity which are common to test development in all fields. These include: (1) construct validity; (2) criterion validity; and (3) predictive validity. The term validity refers to two aspects of a test: (a) the parameter that the test measures and (b) how well the test measures this parameter. These critical issues determine the relevance of specific test methods as measures of neurotoxicity. Both the importance of neurotoxicity testing and the application of specialized test methods for detecting changes in behavior have been accepted
* p<.05 by Duncan's test
** p<.01 by Duncan's test
+ p<.10 by Duncan's test

Figure VI.5. Adult maze performance in rats prenatally exposed to diphenylhydantoin (DPH). The abscissa shows the different groups (0, 100, 150, or 200 mg/kg/day on days 7-18 of gestation) and the ordinate shows the mean number of errors (left) and time (right) in the maze summed across all 11 maze trials of testing. Note that clear dose-response effects are observed (Vorhees, 1986b). The group effect was significant by ANOVA, p<.01.
by the scientific community. At question is which behavioral test methods can be used to measure a particular aspect of neurotoxicity. Often the correlation between the underlying neurophysiological and behavioral defect is lacking. However, once a particular test has been proven to measure an important element of CNS dysfunction, it can then be redirected for use in applied toxicology research. The scientific debate over test validity in behavioral toxicology and teratology has generally focused on the question of what an individual test measures, rather than on the accuracy of the measurement in reflecting a one-to-one correspondence to underlying pathology. Tilson (see p.V-1) has reviewed what can be measured by behavioral test methods and documents that behavioral tests are able to reflect important aspects of neurotoxicity.

1. Construct validity of behavioral testing methods

Construct validity is a term that refers to what biological functions can be measured by a given test method. Without consideration for what aspect of CNS function a given behavioral toxicity test is intended to measure, the selection of an appropriate test can only be made empirically. Empirical test selection may be illustrated in relation to human psychological testing, as for example, in developing tests for neuropsychological injury (e.g., Halsted-Raitan battery) or for disorders of personality and affect (e.g., the Minnesota Multiphasic Personality Inventory). When using the empirical approach the researcher tests the response of a group of individuals with an established neurologic or psychiatric disorder and then compares the test findings to those of a control population not having the characteristics in question. Test elements are selected from those groups which answer the test items differently. These differential response items are then assembled in a test battery to be used for evaluating individuals with unknown diagnoses. Using this approach it is not important for the researcher to know exactly what characteristics the test measures, because it has already been determined that the test correctly discriminates among the individuals being examined for those that possess the target trait and those that do not.

While effective, the empirical approach is seldom followed because it involves the difficult and complex task of first evaluating massive numbers of test procedures of unknown utility. Test development has proven more efficient when there is a concept for what should be measured and when there is some understanding about which test methods will provide accurate measurement of the observed trait. This approach has been used in behavioral toxicology and teratology to establish construct validity. Based on current knowledge of the numerous CNS functional domains that require monitoring and the test methods that are presently available for measuring these functions, there are at least five areas which should be included in a reasonable
test battery for behavioral and neurologic dysfunction. These include: (1) reflex development; (2) activity; (3) reactivity; (4) cognition and learning (including memory and problem-solving); and (5) sensorimotor functions. It has been suggested that measures of reproductive function and social behavior be included in this list (Chapman and Cutler, 1984; Grant, 1976).

2. **Criterion validity of behavioral testing methods**

The ability of a test to measure a characteristic that can be independently defined is referred to as criterion validity. An example is the Scholastic Aptitude Test (SAT) which was developed to assess the ability of high school students to do well in college. The criterion selected for the SAT exam was college grades. Since both the SAT results and college grades can be quantified, a correlation can be calculated to determine how well the SAT performs at predicting college grades in selected test groups after they have completed their college education. A validity coefficient is obtained from this process which can be used to predict college grades in future high school students taking the SAT prior to college entry. The criterion selected in behavioral toxicology and teratology has often been based on the ability of tests to detect behavioral dysfunction after exposure to known neurotoxic agents (Buelke-Sam and Kimmel, 1979; Vorhees, 1983b). This method takes advantage of the use of positive controls which is a common practice in many fields of toxicology and is the primary method for establishing the validity of most testing procedures in toxicology (i.e., teratogenicity, carcinogenicity, and mutagenicity). The use of this approach in behavioral teratology has shown that behavioral tests can detect the adverse neurologic and behavioral effects related to prenatal exposure to aspirin, vitamin A, phenytoin, polychlorinated biphenals (PCBs), pesticides, methylmercury, lead, alcohol, trimethadione, methadone, carbon monoxide, and others (Vorhees and Mollnow, 1986). The success of this approach is related to the availability of a significant database of chemical agents known to be neurotoxic. This database has been established on the basis of widely replicated findings from studies that measured behavioral, neuroanatomical, neurophysiological, and neurochemical indices of brain injury. In practice validity coefficients are rarely calculated in toxicology, but they may be determined under certain special circumstances.

3. **Predictive validity**

Predictive validity refers to the ability of a test to predict effects from an incomplete or partial data set. The SAT is said to have fairly good predictive validity because it projects how well most high school students do in college in terms of grades. The determining factor for such predictions is the correlation between the predictor test and the outcome.
variable. For the SAT this is relatively straightforward; however, for preclinical animal toxicity tests it is important to determine whether the tests predict adverse outcomes in humans (i.e., cross-species predictive validity). This is a major problem encountered by researchers who must interpret study findings from mutagenicity, carcinogenicity, and teratogenicity studies in laboratory animals. There are numerous examples in the literature where animal bioassays have failed to predict adverse effects in humans. A classic example would include the teratogenicity associated with thalidomide (Brent, 1964; Wilson, 1973). In behavioral teratology, problems of cross-species prediction have been studied for those special agents with known human developmental neurotoxicity. Some of the chemical agents that have been studied include alcohol, methylmercury, lead, methadone, pesticides, and some anticonvulsants (Vorhees and Mollnow, 1986). The chemical agents listed above are frequently used as positive controls because they provide a strong correlation with human clinical evidence. A dilemma of toxicity testing is that human morbidity is sometimes demonstrated before an appropriate test model is developed. Positive controls are being used to build a database to verify that certain tests for behavioral teratogenesis can correctly identify these positive test agents. By implication, tests developed in this way should be able to correctly identify other neurotoxicants in the future.

It is equally important that well-chosen tests correctly label as negative those agents that do not produce CNS damage. In behavioral teratology and toxicology, well-chosen tests should correctly identify as safe those chemicals previously determined not to be neurotoxic. Although there are few examples of this in the literature, examples would include moderate doses of prenatally administered amphetamine, (Buelke-Sam, 1986), acetazolamide (Butcher et al., 1975a,b), and some food additives (Butcher and Vorhees, 1984). It is important to note that this statement must be viewed with caution, as the null hypothesis can never be proven true. Other examples could be cited but remain unpublished because of the logistical problems encountered when publishing negative outcome studies in the scientific literature.

Data from the Collaborative Study has provided a rich source of information from which general conclusions can be drawn regarding sensitivity and validity. A significant finding of this study is that all six testing laboratories agreed that the negative test compound (amphetamine) was negative and the positive test compound (methylymercury) was positive. This agreement is an important step in establishing the validity of behavioral testing procedures. It is important to comment on the negative findings associated with amphetamine in the Collaborative Study, as the original study design included amphetamine as a positive control. When the Collaborative Study was proposed in 1979, the published literature for the effects on behavior associated with prenatal amphetamine exposure were contradictory (Buelke-Sam, 1986). As a consequence, the National Center for Toxicologic
Research (NCTR) conducted pilot studies which offered findings that implicated amphetamine as a positive control for certain behavioral tests (Adams et al., 1982, 1985a). Time, definitional, and other constraints were imposed on the Collaborative Study which led to changes in several aspects of the study design between the pilot experiments and the interlaboratory trials. This situation resulted in no-observable-effects for amphetamine in the Collaborative Study. In retrospect, it is clear that amphetamine may be either positive or negative depending upon the dose level administered, the exposure period, and the behaviors measured within the study. Amphetamine illustrates the importance of dose and exposure period selection as crucial factors in the design of behavioral teratogenicity studies.

D. RELIABILITY OF OTHER INDICES OF BEHAVIORAL NEUROTOXICITY

In addition to measures of behavioral and neurologic dysfunction, neurotoxicity may also be determined based on neurophysiological, neurochemical, and neuroanatomical test methods.

1. Behavior and neuroanatomical measures of CNS injury

The relationship between behavioral dysfunction and neuroanatomical changes has been well illustrated (Rodier and Reynolds, 1977; Rodier et al., 1979). In these experiments, Rodier administered the antiproliferative agent 5-azacytidine to mice prenatally and tested the offspring for behavioral dysfunction. Table VI.22 summarizes the many dysfunctions which were found. The abnormality observed was dependent on the day of gestation when the drug was given. At the time of sacrifice, a comprehensive neuroanatomical examination of the adult brains from these animals showed no detectable general pathology. Rodier then conducted quantitative neurohistology on matched sections measuring the distances shown in Figure VI.6. She found significant and highly replicable reductions in the thickness of certain regions depending on the timing of the drug dose given during neurogenesis. On the basis of autoradiographic data on the sequence of neuron formation in different brain regions (Rodier, 1980), Rodier was able to predict which regions would be affected by 5-azacytidine when given on specific days of neurogenesis. Rodier was then able to correlate the affected region of the brain with the type of behavioral outcome observed. Rodier discovered systematic relationships between the brain regions affected and the type and pattern of behavioral abnormalities. Some of these paralleled effects on behavior seen in adult rats with electrolytically-induced focal brain lesions to the same areas as affected by 5-azacytidine when given prenatally.
Table VI.22. Summary of the Behavioral Effects of Interference with Cell Proliferation During CNS Development of Mice Exposed Prenatally to 5-azacytidine (modified from Rodier et al., 1979)

<table>
<thead>
<tr>
<th>Behavioral measures</th>
<th>Day of Embryonic treatment</th>
<th>Day of Postnatal Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>1. Righting tasks</td>
<td>*</td>
<td>-</td>
</tr>
<tr>
<td>2. Gait rating</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3. Grid walking</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>4. Adult activity</td>
<td>*↓</td>
<td>*↑</td>
</tr>
<tr>
<td>5. Bolus counts</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6. Passive avoidance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate</td>
<td>-</td>
<td>*↓</td>
</tr>
<tr>
<td>24-hour retest</td>
<td>-</td>
<td>*↓</td>
</tr>
<tr>
<td>7. Active avoidance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light-to-dark</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Dark-to-light</td>
<td>-</td>
<td>*↑</td>
</tr>
<tr>
<td>8. Spatial observations</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9. Spatial maze</td>
<td>*</td>
<td>-</td>
</tr>
</tbody>
</table>

Key: 0 Observational data difference.
* Different from controls at p <0.02.
†↓ † Direction of difference on tasks where treated animals may differ in two directions.
- No difference between treated animals and controls.
I. ARTERIOR TO CORPUS CALLOSUM

II. ANTERIOR COMMISSURE

III. HABENULA

IV. MEDIAL GENICULATE

V. MIDSAGITTAL CEREBELLUM

Figure VI.6. Illustration of brain regions evaluated by Rodier and Reynolds of mice exposed to 5-azacytidine prenatally. Letters a-i shown above refer to standardized measurements of cortical and subcortical thickness of matched brain sections. Different segment lengths were reduced depending upon which day prenatally the mice were exposed to the drug. Although the tissues were not grossly abnormal, the effects of 5-azacytidine were detectable only by these quantitative analyses. Modified from Rodier and Reynolds, 1977.
Rodier's work has proven that behavioral teratogenicity can be verified using neuroanatomical methods to detect neurotoxicity morphologically by quantitative neurohistology. Some types of qualitative neuropathologic screening can also be used to detect such effects. While pregnant women would rarely be exposed to a drug such as 5-azacytidine, Rodier's work is prima facie evidence indicating the necessity for screening new agents for behavioral dysfunction.

2. Behavior and neurochemical measures of CNS injury

Neurochemical techniques can also be used as measures of behavioral dysfunction and CNS injury (Kellogg et al., 1980, 1983a,b; Simmons et al., 1984). The research by Kellogg and colleagues showed that prenatal exposure of rats to diazepam on days 13-20 of gestation produced a reduction in measurements of activity and auditory startle responsiveness during the third week of postnatal life (Figure VI.7) in pups from treated dams. Since the benzodiazepine receptor is thought to be part of the gamma-aminobutyric acid (GABA) receptor complex, Kellogg and colleagues studied changes in the GABA synthetic pathway by measuring the enzyme glutamic acid decarboxylase (GAD). They also studied stimulated GABA uptake and measured $^3$H-flunitrazepam binding. No differences were found in the diazepam treated offspring for any of these neurochemical parameters. Diazepam treated animals were then examined for norepinephrine (NE) and dopamine (DA) levels and turnover. Figure VI.8 illustrates that no changes were found up to at least 4 postnatal weeks of age for NE, but reductions in hypothalamic NE were observed after week 4. No changes were found in NE in other regions and no changes were found in DA in any region of the brain that was examined. NE turnover in the hypothalamus was also reduced in the older animals, but neither NE nor DA turnover were affected in any other brain regions.

Figure VI.9 illustrates the data for gap-induced auditory startle response inhibition at different postnatal ages after prenatal treatment with diazepam. Differences were found at 28 and 70 days post treatment, but the response changes at these ages were in opposite directions. It seems unlikely that NE reductions can account for these behavioral effects because the behavioral changes emerge before there are any detectable changes in NE. Furthermore, after the NE changes emerge, measurements of noradrenergic changes show a unidirectional decrease with advancing age, while the gap-induced startle response inhibition changes are bidirectional, first showing inhibition and later facilitation when compared to the response of control animals.

Data from Buelke-Sam and Ali (1985) show postnatal changes in activity, startle response, and neurochemical measurements of both dopamine and serotonin after prenatal
Figure VI.7. Effects of prenatal diazepam on the auditory startle reflex from the work of Kellogg et al. Data show that prenatal diazepam suppressed the rise in startle amplitude that normally occurs at background noise levels of 70 dB at 12, 16, 18, and 20 days of postnatal life after prenatal exposure to diazepam on days 13-20 of gestation in rats. Reprinted with permission from Kellogg et al., 1980.
Figure VI.8. Levels of norepinephrine (NE) and dopamine (DA) in the hypothalami of rats exposed to diazepam (DZ) at various doses on days 13-20 of gestation. Differences appeared only in NE levels and only in the hypothalamus. Reprinted with permission from Simmons et al., 1984.
Figure VI.9. Auditory startle response in rats exposed prenatally to various doses of diazepam on days 13-20 of gestation. In this test an interruption (gap) is introduced in the background noise 190 msec prior to onset of the startle signal. The abscissa shows the length of the gap. Significant effects of diazepam treatment were seen in offspring at 28 and 70 days of age. Note that the effects of prenatal diazepam on the startle response are in opposite directions on day 28 from those seen later on day 70. Reprinted with permission from Kellogg et al., 1983b.
exposure to reserpine. While it is difficult to determine whether or not the neurotransmitter changes are causing the behavioral effects, the temporal correlation between the behavioral and neurochemical indices are in agreement and are suggestive of a cause and effect relationship.

3. Behavior and neurophysiological measures of CNS injury

Fox has documented correlations between neurophysiological parameters and a variety of behavioral effects on offspring after early lead exposure (Fox, 1979; Fox et al., 1979a,b; Overmann et al., 1979). In these studies, lead acetate was placed in the drinking water of dams at concentrations of 0.02 and 0.2%. Fox has shown that lead treatment at the higher dose produces faster negative geotaxis turning times, delayed air-righting reflex development, a decrease in separation-induced body temperature loss, and altered patterns of maximal electroshock seizures (observed as blocked forelimb flexion during the test). Electrophysiologic measurements have shown that early lead exposure reduced the latency to appearance of several characteristic brain waveforms (the N1, P1, N2, and P2 components) of a VER (visual evoked potential) (Fox, 1979).

Although additional research is needed, these studies with lead acetate have shown that electrophysiologic measurements can be used as an indicator of behavioral and neurologic dysfunction. In discussing the detection of neurotoxicity, Fox concluded that the "... battery of developmental sensory-motor tests provide sensitive and quantifiable techniques for assessing developmental and long-term alternations in CNS functioning ... except for the VER ... following perinatal insult." Fox recognized the value of electrophysiologic analyses as an in-depth research technique, but saw them as ill-suited compared to behavioral tests as neurotoxicity screening techniques.

E. NATIONAL AND INTERNATIONAL CONTEXT OF NEUROTOXICITY TESTING

The Japanese and British have led the way in establishing behavioral teratology requirements for new drugs. In recent months the European Economic Community has adopted the British behavioral teratology language and has now finalized the guidelines (Council of the European Communities, 1983). The significance of behavioral testing methods are rapidly gaining acceptance in both scientific and regulatory communities throughout most of the industrialized world. In the United States, the Environmental Protection Agency (EPA) has also issued behavioral and neuropathologic premarket safety testing guidelines (U.S. Environmental Protection Agency, 1985, 1986). The EPA guidelines specify behavioral testing more
precisely than any guideline thus far issued, but the impact of these documents is mitigated by the fact that they apply only to chemicals regulated under the Toxic Substances Control Act and are only very rarely used because of the language of this law.

F. SUMMARY

Based on current data there is clear need for routine screening for neurotoxicity that goes beyond the very limited examinations in current use for preclinical studies required by FDA. Of the techniques available that have been subjected to adequate evaluation for reliability, sensitivity, and validity, only measures of behavioral dysfunction offer the best available means to screen CNS functions for damage. In addition, the evidence suggests that behavioral tests offer the most comprehensive means of evaluating CNS injury in a reasonable cost-effective way because other neurobiological assays are frequently too specific or require such a high level of expertise as to preclude their use for screening purposes. At present the CNS functional domains that should be monitored by standardized testing include reflex ontogeny (in developmental toxicology studies), locomotor activity, reactivity (e.g., startle response), learning and memory, and sensorimotor functions. Only those tests that have been established to be accurate for assessment of these functions should be considered acceptable, whereas new tests must bear the burden of peer review to establish their reliability, sensitivity, validity, and their nonredundancy with currently standardized behavioral screening tests. Many of the tests arising from the NCTR Behavioral Teratology Collaborative Study and the Cincinnati test battery meet these requirements and fit within the functional domains described above. These tests would be suitable for routine, large-scale testing of food additives, pharmaceuticals, pesticides, toxic substances, and other chemicals. The FDA could accomplish major gains in safety assessment by incorporating these tests into present toxicity protocols (Sobotka, 1986).
OPEN DISCUSSION

HUTCHINGS: Questions for Dr. Vorhees?

SPENCER: Did the NCTR study incorporate a negative control compound and if so, what were the results?

VORHEES: There is no easy answer to that question. The answer is yes and no. In the sense that amphetamine turned out negative, I would call it a negative compound. At the dose and exposure period used, all laboratories agreed that there was no effect; and that it was negative under the conditions of the Collaborative Study, even though the original intent was that it be a positive control.

SPENCER: In the experimental design, was it considered important to have a negative control?

VORHEES: There were of course negative control groups, one group for the vehicle used, and an untreated group. In that sense, every good study design has one or more negative controls.

SPENCER: Was there consideration given to include an agent which is generally accepted not to have a neurotoxic effect in order to discriminate between what might be significant (i.e., methylmercury) and what might not be significant?

VORHEES: Only two compounds could be evaluated because of a limitation of resources. Negative test agents have been evaluated in contexts other than in the NCTR Collaborative Study.

SPENCER: I would suggest that a negative control compound is an extremely important component of such studies.

In regard to the comments about neuropathology, many scientists correctly perceive that much of this work is relatively insensitive. Many neuropathology studies utilize techniques which were introduced in the 1890s. On the other hand, many of us have tried to incorporate more recent techniques introduced into cell biology perhaps 20 or 30 years ago. These are much more sensitive.
It was the conclusion of the morphology section of a recent WHO committee that an admixture of two techniques -- conventional paraffin histology useful for cytochemical studies, coupled with the contemporary methods of optimal fixation and plastic embedding, should be utilized for neuropathological assay. It is generally perceived, incorrectly, that the transmission electron microscope is needed to see fine cellular details. That is not the case; the light microscope is adequate for most studies. I know of no compound whose neurotoxicologic effect cannot be demonstrated with the bright field light microscope, which is an extraordinarily high resolution instrument if coupled with the appropriate preparative technique.

VORHEES: I agree. The intent of my presentation was to review the work that has been done, not what can be done. There are many other techniques that could provide better information.

O'DONOGRUE: My name is O'Donoghue, I'm with the Eastman Kodak Company. While I am not a statistician, my question is directed to your discussion about reliability. The two parameters that are commonly considered relevant to reliability are precision (i.e., how close the labs are to each other in their results) and accuracy (i.e., how close they are to the true answer). From the data presented for the high dose levels of the mercury, a potent neurotoxin, these tests are both precise in that all laboratories reported a similar finding and accurate in that they reported the correct answer. But at the low dose, the results are precise from all the laboratories in that the answers are close together, but not accurate in that some laboratories reported the correct answer and some laboratories reported the incorrect answer. It appears that this data showed that at high doses it is reliable, precise, and accurate; but at low doses, it is precise, but inaccurate and therefore, not reliable. It is important to have both negative test compounds and positive test compounds so that one can determine whether the result is correct or incorrect.

VORHEES: It is not appropriate to draw any conclusion about accuracy in this situation. The finding in the low dose group is not incorrect. It indicates that the dose that was selected is at the limits of the detection power of that particular test. That does not make it inaccurate.
O'DONOGHUE: But the data showed that the laboratories had different answers and that some called them positive and some called them negative. If this is true, which answer is correct?

VORHEES: With any toxicological measure, one can find sufficiently low doses where there will be ambiguity in the study findings between laboratories. This is one reason for the use of the dose-response curve.

O'DONOGHUE: If it is true that the finding for the high dose in the experiment was precise and accurate and at the low dose it was not, then the point that I would like to make is that the materials that we are concerned about most of the time are probably either negative or weakly positive and would fall into that later category, not in the upper category.

VORHEES: In this study there was no high dose in the sense of the conventional toxicology protocol. It is more appropriately termed the higher dose.

HUTCHINGS: Are there other questions?

KIMMEL: I would like to reiterate that the high dose in the Collaborative Study was actually not a high dose. It was a fairly low dose which was carefully defined in terms of minimal toxicities in both the mothers and the offspring. This is in contrast to the types of toxicity that are usually associated with the high dose. As Dr. Vorhees said, in the case of methylmercury, the low doses were set at the detection sensitivity level of the test. This is one of the issues that we feel is very important for people to calculate and to indicate in the results of their studies. The question being, what is the detection sensitivity of the test procedure in use or how much can you detect using the test design that is available for that test? Although not all the laboratories showed a difference in the ANOVA of our study, there were no post hoc tests on individual dose differences from controls. All of the laboratories showed the same tendency of effect with the low dose as with the high dose.

VORHEES: It should be called Dose 1 and Dose 2.

HUTCHINGS: There will be additional time for questions related to this paper in the general discussion. We will now move on to the next paper.
LITERATURE CITED


Wilson, J.R. 1986. Institute for Behavioral Genetics, Boulder, CO. Personal communication with C.V. Vorhees, University of Cincinnati College of Medicine, Cincinnati.
VII. PSYCHONEUROIMMUNOLOGY, NEW APPROACHES TO NEUROBEHAVIORAL TESTING

Nicholas P. Plotnikoff, Ph.D.
Oral Roberts University
Tulsa, Oklahoma

ABSTRACT

A review of the literature in the field of psychoneuroimmunology was reported by Robert Ader (Ader, 1981). Basically, three research approaches have been used to demonstrate a relationship between central sites within the brain that have modulatory influences on the immune system. These include: 1) both the electrical stimulation and surgical lesioning of nuclei in the hypothalamus; 2) traditional behavioral conditioning; and 3) behavioral stress studies. The mechanisms of action for the psychoneuroimmunologic response appear to be focused in three areas: 1) the autonomic nervous system, where direct anatomical connections from the brain to the immune system have been recently demonstrated; 2) the endocrine system where the individual hormones have direct and indirect effects on the immune system; and 3) enkephalins-endorphins in the brain, pituitary, and adrenals that also have direct and indirect effects on the immune system.

Examples of drug effects on the above complex systems will be centered on morphine and the enkephalins-endorphins with reference to drugs of abuse, food additives, and cosmetics.
A. INTRODUCTION

In the broadest context, behavioral medicine has always attempted to include the effects of behavioral changes on states of health (Spector, 1983). Spector (1980) notes that mood changes have always been associated with changes in health status. In more recent decades, an increasing number of studies have suggested that part of these changes in health may be related to changes in immune function, and numerous investigators have sought to understand the relationships between the central nervous system and the immune system. As a result, terms such as psychoneuroimmunology or neuroimmunomodulation have been used to relate alterations in central nervous system function with changes in immune function (Ader, 1981). Study in this field extends the basic principle of homeostatic self-governance as it has expanded basic research of the psychosomatic concepts of stress-related alterations in intercellular communication. Locke and Hornig-Rohan (1985) have listed over 1,400 sources of information in this rapidly expanding scientific discipline. Their compilation provides a useful collection of authored abstracts that link the nervous and immune system and the role of the brain and behavior in immunomodulation.

Specific aspects of the immune system have been studied, including both the humoral and cellular components of the immune system. Studies have focused on the effects on antibody formation as well as direct effects on T cells, B cells, cytotoxic cells, and macrophages. Connections via the autonomic nervous system, the endocrine system, and the lymphoid organs of the immune system have been studied by numerous investigators. Spector and Korneva (1981) have reviewed the literature describing the function of specific loci of hypothalamic (limbic system) areas of the brain which could be altered to evaluate corresponding changes in the immune system. Figure VII.1 provides a schematic summary of the interrelationships between the nervous system, the endocrine system, and the immune system.

B. PATHWAYS OF NEUROLOGIC AND IMMUNE SYSTEM INTERACTION

1. Autonomic nervous system

Direct autonomic nervous system connections between the brain, organs of the lymphoid system (i.e., thymus, spleen, bone marrow, lymph nodes), and gut-associated lymphoid tissue (GALT) have been established (Spector, 1983). Felten et al. (1985) have identified nerve endings in the parenchyma of the cortex and medulla of the thymus, spleen, bone marrow, and lymph nodes. In particular, adrenergic neurotransmitters released from nerve endings in the
Figure VII.1. Schematic summary of the interrelationships between the nervous, endocrine, immune, and circulatory systems as they interact in psychoneuroimmunomodulation (modified from Spector, 1983).
parenchyma have been associated with T-lymphocyte activation. Other factors include cholinergic fibers as well as peptides such as vasoactive intestinal peptide (VIP), neuropeptide Y, metenkephalin, cholecystokinin (CCK), and neurotensin (Bulloch and Moore, 1981; Williams and Felten, 1981).

2. **Neuroendocrine system**

   Earlier studies have demonstrated numerous interactions between the central nervous system and the endocrine system (Plotnikoff and Kastin, 1977). The principal pathways include the hypothalamus, pituitary gland, and the adrenal glands as well as the gonads (Comsa et al., 1982). Sachar (1978) observed that most of the principal psychoactive agents alter hormone levels and functions of the pituitary gland. Morphine-like agents are known to stimulate the release of growth hormone and prolactin, as well as decrease levels of adrenocorticotropic hormone (ACTH) and gonadotrophins (George, 1971). Changes in the level of pituitary hormones can have both direct and indirect effects on components of the immune system (Maclean and Reichlin, 1981). Many of these studies measured changes in the size of the thymus, spleen, and lymph nodes. For example, thymic involution has been demonstrated following large doses of ACTH, estrogens, and steroid hormones (Grossman, 1985). Adrenalectomy or gonadectomy are known to result in hypertrophy of lymphatic organs, whereas, treatment with growth hormone or thyroid hormone increases thymus gland size (Dougherty, 1952). In addition, all of the previously mentioned hormones also have direct effects on lymphocytes (Ahlqvist, 1981).

3. **Autonomic and neuroendocrine interaction**

   The central nervous system, the endocrine system, and the immune system are integrated and function interdependently. The central nervous system which regulates the outflow of the endocrine system as well as the autonomic nervous system can also affect the immune system (Hall and Goldstein, 1981). Figure VII.2 illustrates the relationship of the thalamic nuclei (i.e., thalamus, hypothalamus, and median eminence) as the interface between sensory and brain regulation in the release of endocrine hormones. Figure VII.3 illustrates brain and endocrine interrelationships and lists hormonal substances that circulate through the blood stream which can influence the immune system. These figures demonstrate both the complexity of information processing and the variety of individual elements within these systems (Spector, 1983).
Figure VII.2. Autonomic brain and sensory pathways involved in hypothalamic stimulation and hormone release (modified from Spector, 1983).
Figure VII.3. Brain and endocrine interrelationships and the transport of hormonal substances that can influence the immune system (modified from Spector, 1983).
C. ELEMENTS OF NEUROIMMUNOLOGIC FEEDBACK LOOPS

Earlier studies attempting to show "connections" between the brain and the immune system were criticized on the basis of inadequate controls (Ader, 1981). Frequently no controls were employed in these early investigations. Sham procedures had been omitted in studies involving either brain lesioning or stimulation of specific nuclei in the brain. However, recent studies with adequate controls have established the validity of a large series of electrophysiologic studies. These studies suggest that stimulation of specific areas in the brain could influence the activity level of the immune system (Cross et al., 1982; Spector, 1980; Stein et al., 1976, 1981).

Antibody levels to specific antigens that circulate within the blood stream can be altered by placing lesions in the hypothalamus. In particular, Jankovic and Isakovic (1973) have observed that stimulation of specific nuclei in the hypothalamus can alter both humoral (antibodies) and/or cellular (T cells) immunity. For example, lymphocyte proliferation in the presence of mitogens was significantly altered by lesioning or stimulation of specific areas of the hypothalamus (Jankovic and Isakovic, 1973). Cross et al. (1984) observed that the placement of bilateral electrolytic lesions in the preoptic-anterior hypothalamic area (AHT) of rats reduced natural killer cell activity. Furthermore, the number of circulating lymphocytes decreases after lesioning of the anterior hypothalamus and increases after hippocampal lesioning (Brooks et al., 1982; Cross et al., 1982; Jankovic and Spector, 1986; Stein et al., 1976).

Of special interest are the findings of Besedovsky and associates that the firing rates of single ventromedial hypothalamic units change when antigen is injected peripherally (Besedovsky and Sorkin, 1977; Besedovsky et al., 1979). These studies suggest that the cellular immune system releases lymphokines (e.g., interleukins, interferons) that also affect the brain. This finding provided evidence for a complete feedback loop (Besedovsky et al., 1983). The various tissue and cell components, cell products, and reactions of the immune system that are involved in the psychoneuroimmunomodulation feedback loop are summarized in Figure VII.4.

Studies by Blalock and Smith (1980) have found strong antigenic relatedness among human leukocyte interferon, ACTH, and endorphins, suggesting prohormones may also be involved in the feedback system. In addition, human leukocyte interferon binds to opiate receptors and produces endorphin-like effects providing further support for a regulatory circuit between the immune and nervous system (Blalock and Smith, 1981).
Figure VII.4. Components, cell products, and reactions of the immune system involved in the psychoneuroimmunomodulation feedback loop (modified from Spector, 1983).
D. ENVIRONMENTAL STRESS AND THE IMMUNE SYSTEM

1. Evidence from man (psychiatry)

Anecdotal reports have attempted to relate emotional status of the patient to the onset and course of several disease states (e.g., susceptibility to infection, ulcers, rheumatoid arthritis) (Alexander, 1950; Plaut and Friedman, 1981). Retrospective analyses of a patients' emotional history provides some support for this observation (Fox and Newberry, 1984; Solomon, 1981). However, more convincing are the studies that correlate states of depression with immune dysfunction. For example, Schleifer et al. (1983) observed that T-lymphocyte proliferation to mitogen stimulation was reduced in study participants who experienced a recent bereavement. In another study, Schleifer et al. (1985) observed that the lymphocyte proliferation in the presence of mitogens was lower in hospitalized depressed patients than in matched controls. Kronfol et al. (1983) also observed a marked decrease in lymphocyte mitogenic activity in depressed patients.

2. Animal studies

Classical Pavlovian conditioning studies have provided useful information regarding the relationship between the immune system and environmental stimuli. While earlier studies were highly questionable because of the small numbers of subjects and inappropriate controls, recent studies conducted with appropriate controls demonstrate that classical conditioning techniques could be employed to modulate immune function. The most definitive studies were conducted by Ader and Cohen (1981) who showed that immunosuppression by cyclophosphamide could be paired with a conditioning cue (saccharin). In their studies, Ader and Cohen (1981) found that immunosuppression could be induced by the saccharin conditioned cue alone after the cyclophosphamide treatment had been removed.

E. HORMONE INTERACTIONS -- BEHAVIORAL STUDIES WITH ENKEPHALINS-ENDORPHINS

Early behavioral studies with the hypothalamic releasing hormones indicated that these hormones exerted direct effects on the brain in addition to their well-established effects on the pituitary gland (Plotnikoff and Kastin, 1977). In a similar manner, Plotnikoff and Kastin (1977) found that methionine enkephalin could: 1) potentiate dopamine induced hyperactivity; 2) reduce fighting induced by footshock; and 3) reduce epileptic seizures induced by auditory stimulation. The three study models employed in rodents are sensitive measures of the central effects of hormones. In particular, the dopamine potentiation model
results in marked hyperactivity, fighting, and vocalizing at doses of methionine enkephalin as low as 0.1 mg/kg. The dopamine potentiation test is highly sensitive to known antidepressants.

These early studies led to the hypothesis that the enkephalins from both central and peripheral sources could alter mood states and behavior (Plotnikoff et al., 1976). In addition, a large number of other behavioral tests (e.g., spontaneous activity, consummatory behaviors, learning, swimming, grooming, avoidance conditioning) were also affected by the enkephalins-endorphins (Olson et al., 1982). These behavioral studies with the enkephalins-endorphins indicated that the opioid peptides exerted both neurotransmitter-like and hormone-like effects (Müller and Genazzani, 1984).

Numerous investigations have observed that behavioral stress in rodents could stimulate the release of corticotrophin releasing factor (CRF), ACTH-adrenocorticotrophin, and beta-endorphin from the brain, pituitary gland, and adrenal glands (Amir et al., 1980; Guillemin et al., 1977; Plotnikoff and Murgo, 1985; Riley, 1981; Selye, 1980). Bloom (1983) observed that the prohormone, proopiomelanocortin, gives rise to beta-endorphin and ACTH in the brain, pituitary, and adrenals, while proenkephalin A gives rise to methionine enkephalin in the brain and adrenals.

Subsequently, stimulation of the adrenals by ACTH and beta-endorphin results in the release of steroids, catecholamines, and enkephalins-endorphins (Hanbauer et al., 1982; Viveros et al., 1980). It was found that the enkephalins were stored in the chromaffin granules with the catecholamines and that both were released simultaneously when the study animal was subjected to stress. The steroid hormones released from the adrenals have been shown to cause involution of the thymus gland (Parrillo and Fauci, 1979; Riley, 1981). Recent studies by Plotnikoff and his associates have indicated that enkephalin treatment results in an enlargement of the thymus gland and can antagonize the thymic involutionary effects of steroids (Plotnikoff et al., 1984).

F. RECEPTOR SITES IN THE IMMUNE SYSTEM

Wybran et al. (1979) suggested that human T-cell lymphocytes had receptor sites for methionine enkephalin as well as morphine. The binding sites were measured by forming rosettes with sheep red blood cells. The number of rosettes formed increased in the presence of methionine-enkephalin and decreased with morphine. As a result Plotnikoff (1981, 1982) studied the possible effects of the enkephalins and morphine on the immune system. During the early 1980s, Plotnikoff and Miller (1983) found that the enkephalins stimulated lymphocyte proliferation in the presence of a mitogen and also extended survival time of mice inoculated with attenuated L1210 leukemia
cells. In human studies the enkephalins were found to increase the number of T cells forming rosettes with sheep red blood cells in normal volunteers and cancer patients (Faith et al., 1984a; Miller et al., 1983, 1984).

Murgo et al. (1985) have shown that the enhancement of active T-cell rosettes by methionine enkephalin is increased in the presence of zinc. This is particularly interesting since it has been previously shown that zinc affects the binding of opiate peptides in the brain (Stengaard-Pedersen, 1982).

Natural killer cells (NK) were found to be activated by the enkephalins in normal volunteers and also cancer patients (Faith et al., 1984b). In these studies, K562 tumor cells filled with $^{51}$Cr were used as the target cells. The degree of lysis produced by the activated NK cells was measured. Clinical studies have shown that methionine enkephalin increases both the number and function of T cells in cancer and AIDS patients (Plotnikoff et al., 1984, 1985a,b, 1986). Infusion of methionine enkephalin in doses of 10 to 25 μg/kg increased the numbers of T cells (OKT 3, 11, 4, and 8) as well as mitogen responsiveness. Most importantly, enkephalin increased the production of mono- kines and lymphokines (interleukin I, II, and gamma-interferon) (Brown and Van Epps, 1985; Plotnikoff et al., 1986; Youkilis et al., 1985).

In contrast to the enkephalins, morphine has been found to have opposite effects on the immune system (i.e., down regulation). Several groups have found in both animals and humans that morphine reduces mitogen induced lymphocyte blastogenesis, antibody response following antigen challenge, and the number of T cells forming SRBC rosettes (Faith et al., 1984a). These down-regulating immunological effects are particularly striking in heroin addicts, who are prone to numerous infections (Bocchini et al., 1983; Brown et al., 1974; McDonough et al., 1980). Immunosuppression of T-cell functions by morphine can actually accelerate mortality induced by L1210 in mice as well as infections induced by Klebsiella pneumoniae and Staphylococcus aureus (Koyuncuoglu and Gungor, 1986; Tubaro et al., 1983).

Phencyclidine (PCP) is another psychotropic drug which binds specifically to lymphocytes. Receptors for this binding may be akin to the sigma opiate receptor (Fudenberg et al., 1983). These findings have implications for the pathogenesis of the schizophrenic-like syndrome associated with PCP abuse and also for the use of PCP derivatives in general anesthesia where it may contribute to post-operative infections (Khansari et al., 1984). In addition, the benzodiazepines, which are widely prescribed for their anti-anxiety effects, are potent stimulators of human monocyte chemotaxis (Ruff et al., 1985).
G. CENTRALLY-ACTING COMPOUNDS THAT INFLUENCE THE IMMUNE SYSTEM

Many of the immunological effects associated with the enkephalins and morphine are attributed to direct effects on lymphocytes. Weber and Pert (1984) have identified receptor binding sites on the surface of the lymphocytes for morphine as well as the enkephalins. However, there is evidence that both the enkephalins and morphine have central components that may influence the immune system through the autonomic nervous system and the endocrine system (Müller and Genazzani, 1984). Buckingham and Cooper (1984) have shown that morphine-tolerant rats failed to secrete corticotrophic releasing factor (CRF), ACTH, or corticosterone in response to stress. In contrast, Shaar and Clemens (1980) have demonstrated that the enkephalins stimulate release of growth hormone and prolactin. In animals, the autonomic nervous system effects of the opiates and endorphins are seen as reductions in blood pressure and heart rate (Wei et al., 1980).

Other "drugs of abuse" (e.g., heroin, PCP, marijuana, barbiturates, and alcohol) have also been found to down regulate the immune system perhaps by effects in the brain. The suppression of the endocrine and/or autonomic systems via the hypothalamic nuclei may result in immune system down regulation (Fischer and Falke, 1984; Hole, 1984; Khansari et al., 1984; Watson et al., 1983). Alternately these drugs may have direct effects on the immune system as described above for PCP (Khansari et al., 1984). Other centrally-acting drugs include general anesthetics, antipsychotics, benzodiazepines, and anticonvulsants (Spector, 1980). Since psychoactive agents interact with various neurotransmitters (e.g., the catecholamines) in brain areas such as the hypothalamus, it would be expected that alterations in the endocrine and autonomic systems would impinge on the immune system (Johnson, 1975; Rugarli, 1983; Zimmerman et al., 1977). This has been demonstrated for drugs such as heroin, PCP, and alcohol where suppression of the immune system may have central origins (Descotes et al., 1985; Portoletes et al., 1982).

In a similar manner, it is possible that other substances, such as food additives and cosmetics, may have effects on the brain that may result in down-regulation of the autonomic nervous system, endocrine system, and, subsequently, the immune system (Asher, 1978; Koller, 1979; National Research Council, 1982; Norbury, 1982; Street and Sharma, 1975; Wierda et al., 1981).
H. SUMMARY

Psychoneuroimmunology is a developing science that will encourage continued study of the immunological effects of centrally-acting agents. Perhaps one approach that may be useful is to simply measure the size of the thymus gland and the spleen at the conclusion of traditional toxicity studies. Such measurements could be used to compare variations between treated and control animals and may offer some insight for immune system response to centrally-acting agents. Quite possibly this data may already be available, since many subacute and chronic toxicity studies routinely measure lymphoid organ size. However, more sensitive secondary assays may be required to measure immune dysfunctions (e.g., antibody formation, mitogen responsiveness, NK cell activity, and chemotaxis) and the proportions of lymphocyte subsets such as helper-suppressor T cells. An immunological measure may help to evaluate potential "behavioral toxicity" of test agents and in some way prevent the occurrence of disease related to immune system dysfunction, such as AIDS, infections, and cancer.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the constructive comments received from Drs. Anthony J. Murgo, West Virginia University, Morgantown, West Virginia; and Robert E. Faith University of Houston, Texas in the preparation of this manuscript.
OPEN DISCUSSION

HUTCHINGS: Questions for Dr. Plotnikoff?

CAVAGNARO: Joy Cavagnaro, Hazleton Laboratories America. In your in vitro systems, where you assess the different activities of cells (e.g., OKT-3), do you isolate the specific cells and incubate them with the enkephalin? The reason for my asking this question is that if you have a total system containing all cells and then you measure these cells at the end point, you may be getting a secondary effect by stimulating another lymphokine.

PLOTNIKOFF: Those studies were in vivo. We had taken blood samples before the administration of methionine kephalin. Then we took blood samples after administration at different times, to compare the results one to another before and after administration.

CAVAGNARO: Could the activation of T-cell function that you observed be a secondary effect following the administration of [Met]-enkephalin? That is, via stimulation of one of the interleukins and/or humoral thymic factors.

PLOTNIKOFF: Several groups have now demonstrated that methionine kaphalin specifically stimulates the production and release of interleukin-1 and -2 and gamma-interferon. In opposite contrast, morphine depresses all of them. The immunological effects in vivo, as well as in vitro at least in part, may be said to be related to the central effects. Obviously, there are peripheral effects also.
LITERATURE CITED


VIII. GENERAL DISCUSSION

HUTCHINGS: This is the general discussion portion of this meeting. It is an opportunity for anyone to come forward to make a comment or raise a question about the information presented during the symposium.

ISON: My name is Jim Ison from the University of Rochester. I want to comment about the data from our laboratory that was presented today (Kellogg et al., 1980), and the data of Young and Fechter (1983). I would also like to comment on the relationship between those data and validity, the relationship between validity and public policy, and the need for further research.

When we produced the data for the effects of prenatal exposure to valium on the sound-induced activation of the startle response, we were uncertain about what this phenomenon meant for humans; however, at this time we are approaching the point where we can make an educated guess. This is especially true for the work in Carol Kellogg's lab where she has come close to understanding the neurological basis for the startle response and the mechanism of arousal by noise.

Changes in the development of the startle response by diazepam was a clear behavioral effect that we now think is important as an index for the development of emotionality. Animals may be emotionally sluggish or lethargic and may develop the normal display of emotions later than usual. In this case, we have data that provide a warning sign and more work is needed in this area to clarify the meaning of these findings. There is a need to understand the basis for the startle response and its relation to the human condition.

In contrast, for other areas (e.g., vision, hearing, and cutaneous sensitivity) behavioral teratologists and behavioral toxicologists can measure behavioral effects in animals that have clear significance for human health.

On the issue of validity for behavioral testing, one asks such questions as, what can disturb sensory functioning in humans? In this case, we are reminded of viral conditions and viral exposure during the prenatal period. These deficits can be detected in infected animals as well as in humans and they can be shown in conjunction with anatomical changes by using the tests that have been developed. The same can be said about chemical effects on sensory processing in humans. For example, we have known for many years about the amino-glycosides which also have teratological effects. These drugs affect sensory processing in animals and we can show
the neuroanatomical differences. Another example could be the neurological expression of environmental effects. Humans are known to suffer from excessive exposure to noise. Auditory functioning can be greatly diminished by such exposure. This is also true in animals and it can be easily detected in both the neuroanatomical substrate and behavior. Finally, sensory function can be affected by degenerative disease processes related to genetic transmission. These can be investigated in animal models as well as in humans, and detected in both behavioral and neuroanatomical studies (Ison, 1984).

Each condition that can affect sensory processing in humans can also be shown to affect animals. One should be reminded that these effects can be detected by behavioral laboratory testing. There is no question about the validity of these data which can clearly demonstrate sensory dysfunction efficiently and inexpensively.

If we are going to take from these behavioral data a public policy message, it is that the behavioral test guidelines simply should be made available at this time. It can no longer be said that behavioral data are without meaning, or that we do not understand their validity and their reliability when we clearly understand both their validity and their reliability. These data have a clear connection to the human condition. As a public policy statement, regulatory guidelines should talk about specific sensory processes in animal testing.

In speaking about more problematic behavioral effects, such as emotion, social, affective, and cognitive behaviors, it is true that there is more difficulty in making firm comparisons between animal and human data. At this time, there is uncertainty regarding their anatomic or biochemical substrate, but there is no question that current research is getting closer to the answer. At least that is true for the startle response, with which I am most familiar. Both the anatomical work from Mike Davis at Yale, and the neurochemical work that has been done by a number of people including Larry Reiter, are providing insight for what is happening in the specialized domain of affective and cognitive function. There is little doubt about the ultimate correspondence between animal behavior and human behavior in these domains.

At this time, we know what is happening with sensory and motor processes, and from the standpoint of public policy, those guidelines should be written. I believe that with continued research support, we will soon know about other things.
HUTCHINGS: Thank you very much, Jim. It is important to keep in mind that this is an information-gathering session for the FDA. The question that was put before us was whether or not new behavioral tests of neurotoxicity should be incorporated into regulatory guidelines for the FDA. Since the Panel could address many issues regarding reliability and validity without focusing on the fundamental questions for conducting this symposium, I would like to ask Tom Sobotka from the Center for Food Safety and Applied Nutrition at FDA to comment on this question in order for us to reach some conclusions about these issues and so that we can fulfill our obligations to FDA. It might be helpful if you could provide us with a perspective from the regulatory point of view at FDA.

SOBOTKA: One of the main issues for consideration at this symposium is to address the question of the need for behavioral toxicity testing. There is a consensus and a general recognition of the fact that a number of different chemicals have the potential to produce some alteration in the nervous system. The question at issue is how our agency should respond to this information within the context of conventional toxicity testing. There are a large number of tests used in the traditional toxicologic evaluation of chemicals. The nature and extent of the data depends upon the chemical under evaluation -- food additives, color additives, environmental chemicals, or contaminants. Precluding the need for requiring additional testing, is there any information within that database which might indicate a potential neurotoxicologic problem?

The rationale for asking this question is the general awareness and concern by regulatory agencies and research scientists involved in chemical safety evaluation that one must not lose sight of the economic considerations. If there is a need for added testing insofar as certain toxicity is not adequately evaluated in conventional testing, then I think that we must request that appropriate information be generated. If sufficient information is provided from the available data, then there is no need to require that additional testing be included as part of an across-the-board screening of chemicals. If at any time the data indicate that a particular chemical has a potential neurotoxic effect, then we should proceed with the secondary levels of testing which have been referred to by a number of the speakers today.

To exemplify this issue, consider the conduct of a conventional reproduction study. In this situation, the animal is studied throughout mating, gestation, parturition, and litter rearing to a point at which the offspring can survive on their own and reproduce. At each stage of such a study there are numerous maternal
and offspring reflexes and behavioral patterns which must be functionally intact to ensure the successful completion of the process of reproduction and survival of the offspring. It is reasonable to assume that the information routinely obtained from reproduction studies might also be useful as an index of whether behavioral or neuronal processes are or are not affected by the experimental treatment. However, the question still remains to be answered. How much of this information is pertinent and how much of such information can we realistically collect from studies such as these? Is this information sufficient to derive conclusions about the potential for neurotoxicity? Would such information permit the detection of possible subtle behavioral toxicity such as learning disorders, or would it be limited only to those innate reflex behaviors which are so robust that they would require a considerable degree of toxicity to be affected? If the available information is not sufficient, then what additional information would be needed?

The first question that must be addressed is whether or not there is neurotoxicity information that can be gleaned from traditional toxicology studies. Depending on the answer to this, subsequent questions would involve the nature of the additional information, or test measures, that should be included. What kind of behavioral data should regulatory scientists routinely obtain in order to adequately evaluate the question of neurotoxicity?

HUTCHINGS: Thank you for helping us to focus on these questions.

ZENICK: Hal Zenick from EPA. One of the issues that I think Ron Wood brought up in an earlier question, was the distinction between the popular terms of hazard identification versus dose response relationships.

We know how to screen compounds and to detect if a compound will have a behavioral effect; however, I would like to comment on a couple of issues that I am not sure we are addressing, and for which the database may already exist. My question is not which test to use (e.g., a good apical test, because it taps a lot of abilities), but what is true of species differences and compensatory abilities? For example, if one uses an apical test, is it true that the rat can tolerate a certain amount of brain damage and has the ability to use compensatory mechanisms to mask the damage or is that ability in this species really no greater than in man? To what extent is the use of that apical test valid?

Another issue is one of individual susceptibility. This would fit in with some of the adult animal work that Hugh Tilson talked about, where we collect baseline information on this
animal before we expose it to the drug or to the compound. I am interested in the question of whether we ever make any use of the individual's basal behavioral robustness in terms of its activity level, the number of errors to learned criteria, and so forth. How susceptible is that animal as an individual when it is exposed to the chemical agent. When talking about dose response relationships, there is no question that we can determine a dose response curve, but these issues become critical when we ask how that no-effect level relates to man should a safety factor be used for individual susceptibility and for species differences. Some of the issues that FDA and other people feel are important to deal with are not just which test, but these other issues related to defining the dose at which man is at risk.

My question is: If the data is out there, are we really drawing upon it or if we are getting that data now in our experiments, are we making use of it?

HUTCHINGS: Thank you. You have raised some caveats and some questions that are of concern. At this time I would like to bring the discussion back to specifically answer those questions and issues that Tom Sobotka set forth.

ZENDZIAN: My name is Bob Zendzian from the Office of Pesticide Programs at EPA. It was mentioned earlier that Pesticide Programs has been resisting the application of behavioral tests in neurotoxicity and special neurotoxicity tests. I would like to play the devil's advocate and present the case as I see it.

The Office of Pesticide Programs at EPA asks for one specific neurotoxicity test which is to test organophosphates in the chicken for delayed neurotoxicity. We ask for this test because we know that that type of toxicity occurs, can be seen in man, and that the chicken is the valid experimental subject. In addition, we ask that petitioners do neurotoxicity testing on compounds and in species where they find neurotoxicity. This is an opening which gives the regulatory scientist an opportunity to ask for other specific tests if our standard tests detect neurotoxic damage. Everything that we have seen so far can be screened by the tests that are already in place. The tests that we are talking about here are more useful to find the no-effect level, or to find sensitivity. The EPA considers these to be secondary tests. At this time, our agency asks petitioners to conduct about five different tests for those compounds which have shown neurotoxicity in the standard toxicology tests. We have seen no need to ask for a specific test on a regular basis in the same way that we ask for the chicken test.
These other tests appear to be looking for intangible effects. To use the analogy of the fishnet, with the general toxicity tests we know the size and shape of the opening, the type of fish that can be caught, and what we will miss. As we choose the size of the holes in the fishnet, we also choose one test in preference to another.

Additional neurotoxicity tests should be done when we see neurotoxicity in the basic tests. These additional tests would allow us to look for sensitivity, to get a no-effect level, and to get some idea of the other things that the test chemical might affect.

HUTCHINGS: Thank you. The question to the panel is: Do we need all of these additional tests? This is a matter of concern to the FDA.

VORHEES: I want to ask one question about the assumption you made in your analogy about casting a net. I question the validity of the statement that you know what gets away with the current tests, because I do not think that you do know what gets through the net. You only know what you catch, you do not know what you miss. This is the problem. The question that we need to ask is whether or not we need to change the shape of the net by including neurotoxicity, so that we have a chance to catch the neurotoxins? Can you support the statement that you know what gets away?

ZENDZIAN: You have hit upon the fallacy that you are trying to overcome. It comes back to the point that so far we have not seen anything that has gotten away. In looking back, what has gotten away from detection in the basic tests for delayed neurotoxicity has only been found in man. This is why we use the chicken as the test model, because it is directly applicable to man.

STEINBERG: When talking about the current standard for screening pesticides, it is important to talk about how they will be used, whether that be long-term or short-term. For instance, many of the behavioral effects, including the condition of avoidance response changes that are seen in man and the rat, were missed.

One reason that these changes are not readily detected in the United States is because we do not run into lettuce-borne diseases as frequently as they do in those countries where these pesticides are used extensively. The neurotoxic effects of these compounds can be seen in places where this type of toxicity has been observed and can be readily demonstrated in a rat.
REITER: I would like to comment on the possibility that compounds make it through the "fishnet". In the National Academy of Sciences report on safe levels of chemicals in drinking water (National Research Council, 1977), ADIs were calculated for a number of carbamate pesticides. Those ADIs were based on no-effect levels of exposure in a functional observational battery. In particular, signs like salivation, lacrimation, etc., were considered for carbamates (and organophosphates) in the rodent. If one does even the most rudimentary behavioral testing (e.g., motor activity in figure-eight mazes) for several carbamates, the ED50s (defined as the dose producing a 50% decrement in that behavior) are far below those associated with these overt signs of toxicity. Although these differences in effects levels may be covered by imposing a safety factor, the no-effects levels are not always consistent with results obtained from objective behavioral measures. With respect to cortisone, it may be an example of a compound which slipped through the net.

ZENDZIAN: Behavioral analysis was not used to determine the ADI. A more sensitive method is to measure cholinesterase inhibition in the blood and other organs. The observational behavior test is used there.

MAILMAN: Perhaps it is not such a clear-cut issue on both sides. With some compounds, there is an opportunity to apply these tests creatively. For example, when evaluating a pesticide that presumably has a demonstrable effect on the insect nervous system, it is valid to put these compounds through an exhaustive behavioral screen. This is because subtle or possibly specific molecular actions may cause changes which will allow us to better define an allowable dose.

In this instance, knowledge of the mechanism of action can lead us to choose a specific type of testing. The issue is: What about those chemicals that are not pesticides, or those where we do not know the final mode of action? How exhaustively and at what dose level should we screen? The pesticide issue is clear-cut when compared to food constituents or things that may turn up in our water where there are more things to consider.

ZENDZIAN: These tests are referred to as secondary tests. When there is a known effect on the nervous system, then we look for the no-effect levels based on our knowledge about the effect of the compound. An example of this would be the procedures used to evaluate cholinesterase inhibitors. However, we have recently discovered a few unusual compounds for which there are neurological effects at doses that do not effect the cholinesterases, and which occur only after 18 months of chronic administration.
SPENCER: Before leaving the last question, it would be a tragic mistake to think that the nervous system was so limited in its repertoire of responses as to permit the detection of injury based only on the chicken assay. The chicken assay is used to detect central-peripheral distal axonopathy, which is only one of the many neurotoxicological responses that are recognized.

One compound which probably would not have been picked up with the chicken assay is Musk Tetralin. This compound was used for some 25 years in fragrances on the assumption that it was safe, based on short-term, acute toxicity studies. However, it was found to have the most profound effects, both on the spinal cord and on the cerebral cortex, following prolonged exposure via the skin. I suspect that we do need a rather special type of assay to detect the toxicity demonstrated by this compound. Musk Tetralin would be a good compound for the behavioral toxicologist to exploit in further studies. Concurrent with the profound neuropathological changes, the animals became aggressive.

Another compound which also falls within the domain of FDA is Musk Ambrette. This compound was discovered to be neurotoxic by the FDA some years ago. Although it has been around for some 25 years and there was an abstract placed in TAP (Toxicology and Applied Pharmacology), no one ever followed that up. The toxicity could have been picked up from a 90-day study and perhaps the chicken would have responded in a comparable fashion.

One could speculate that food additives such as glutamate and aspartame (which breaks down to aspartate), if studied at high doses in the correct species, at a suitable age, could develop seizures. The question is, what to do with this type of data on the basis of today's discussion. It is clearly an acute neurotoxic phenomenon, but what relationship does it bear to chronic toxicity? This question has come into focus recently because of a totally unrelated disease, latharism, which occurs in parts of the Third World. This disease is most likely induced by a very potent glutamate analog. The lathyrous glutamate analog produces a spastic paraparesis in monkeys. It would have been difficult if not impossible to predict that an excito-toxic compound that produced seizures following acute exposure would produce chronic neurotoxic effects in the form of spastic paraparesis. On the basis of this type of observation, do we need to reevaluate glutamate and aspartate to determine whether they have the potential following chronic toxicity to produce this type of change?

In returning to the specifics of Tom Sobotka's question, I would suggest that because of the wealth of endpoints in neurotoxicology and the complexity of the brain and the nervous system which Richard Mailman has so nicely
illustrated, that we should think about using tests which detect a large number of different neurotoxicologic responses. We cannot rely solely on the chicken axonopathy assay to screen all compounds with neurotoxic potential. Tissue culture methods are now available to assist us in this enormous task.

ZENDZIAN: At the beginning of this meeting we heard about the importance of training technicians as observers which is only one aspect of the problem. The other aspect is the importance of training neurotoxicologists to understand the uses for these tests. The chicken assay is used for one type of toxicity associated with the organophosphates. Unfortunately, I have seen this test misused many times. For example, there is no reason to use this test for carbon-ring compounds that do not have a phosphate group. Furthermore, it is incorrect to assume that a compound is not a neurotoxin on the basis of a repetitive finding in the chicken assay. While this is not a valid application of this test, I have seen both registrants and reviewers referring to it inappropriately. We must understand not only what the valid applications are for these tests before we use them, but also what information can be provided by these tests when we use them.

LUBIN: Harold Lubin with the American Medical Association. As a representative of an organization of physicians, I would like to ask the panel to consider a problem that physicians face. I believe that we are on the horns of a dilemma. Our populace is frequently told that "they are being poisoned." This frightens people and often they take one of several responses. On the one hand they might say "since I'm going to be poisoned anyway, I will be totally unconcerned and do everything that I wish to do, even though it may be unhealthy." On the other hand, some people might say "I'm very frightened so I won't use anything I don't absolutely have to use." These frightened people address their concerns to the physician, and the physician turns to many of the experts in this room, but physicians do not receive a cut-and-dried answer to these questions. I think this is because many times the answers do not exist.

We have been asked by FDA if we have enough information to begin to answer these questions. It seems to me that this is an instance where people could say (a) "Yes we do have sufficient information," (b) "No we do not," or (c) "We do in certain areas and we do not in others." This would help the physician to translate some of this information to patients. Is there anyone here who could address this issue? If nobody does, then the presumption that follows is that we do not, at least in our collective knowledge, have the requisite kinds of information available today.
MAILMAN: There are some words of sensible advice that can be given to everyone. You should always wear your seat belt. You should not smoke. And you should not take anything in excess.

There will always be new compounds that pose unanswered questions. Scientists may not always agree and a consensus of opinion may not be available even for compounds that have been rigorously tested; however, we would be well advised to rely on common sense. For example, despite what Linus Pauling says in regard to pyridoxine and the issue of megavitamin therapy (Pauling, 1968), the optimum level of a compound is not always related to the highest possible concentration. Common sense goes a long way to letting us deal with the important problems as opposed to having everything become a problem.

CHANG: Most of us look at the issue with tunnel vision. The compounds that we evaluate are not just neurotoxicants, but they are also general toxicants. When evaluating neurotoxicants from the neuroscience perspective, we must avoid being misled by the other toxic effects of the compound in question.

Dr. Plotnikoff described the important relationships between neurotoxicity, endocrine actors, and immune system interactions. It is important to understand that some toxic compounds affecting the extra neurologic organ systems will produce or exaggerate neurologic lesions. It is not always possible to immediately assign the changes taking place in neurologic function, morphology, or chemistry to be a direct toxicant effect of the compound on the nervous system.

When using the fishnet analogy, it is important to remember that it is not for us to say that one fish is more important than another. For example, some scientists may think that vision is important, while others may think that some other sensory function is important. No matter what type of fishnet is used, some fish will be missed. There are limitations to the methods used and to the fundamental assumptions of the discipline itself. It is more important to look at multi-disciplinary methods to screen neurotoxicants than to assume that one method will be the catchall. This is also true of neuropathology and neuromorphology in correlation with neurofunction and neurobehavior. If we do not know what we are doing, we will not see what is there and this is true even for neuropathology.

I agree with Dr. Spencer's earlier comment that new skills and new technology must be implemented into this science. We cannot rely on 20-year-old methods and call them adequate.
As Dr. Mailman has pointed out, the nervous system is extremely complex and research findings can reveal paradoxical correlations. There may be neurobehavioral changes that appear without new morphology, or new morphological lesions that occur contrary to what might have been expected to support the behavioral changes. The complexity of neurologic circuitry and the finding that each area of the brain could be interrelated or interconnected make correlations difficult. In my own experience, I have found that it is possible to produce lesions in one area of the brain from another area of the brain which is actually signaling target neurons that are two or three circuits down the road. The resultant outcome would be to produce neurobehavioral changes in that individual which might be totally unexpected. For example, one would not normally expect that changes in memory and/or learning could result from lesions in the brain stem.

We must look at neurotoxicity in a broad spectrum and not with tunnel vision. We can not depend on one or two assays to determine the adequacy or sensitivity of the test method. If we can look at neurotoxicant compounds as toxicants per se, we will find more answers than just looking at one small aspect of the toxicity.

REITER: I would like to address Dr. Lubin's question on two levels. At present, there are inventories of existing chemicals in commerce that vary in number from 40,000 to 66,000. With such a large number, it is not possible for neurotoxicologists to assume responsibility for knowing the extent to which all these chemicals are neurotoxic, especially when the extent to which many of these chemicals are toxic in general is not known. In the National Academy of Science's report on toxicity testing (National Research Council, 1984) it was necessary to divide the inventory of chemicals into seven categories to prepare a representative sample. They reported, for example, that approximately 13,000 industrial chemicals are produced in excess of 1 million pounds per year. Approximately 78% of these chemicals were totally devoid of any toxicology database. This is not unique to neurotoxicology; it is a problem to toxicology. We do not know which of these chemicals are of toxicological significance let alone which are neurotoxic. If we accept the estimate of Kent Anger and Barry Johnson (Anger and Johnson, 1985), then 25% of these chemicals can be expected to produce neurotoxicity suggesting that there are many neurotoxic chemicals produced in extremely large volumes for which we have little neurotoxicological information. There is obviously a need to develop an approach that will systematically evaluate how chemicals interact with the nervous system to produce toxicity. Since it is not possible to test each and every one of these chemicals, even with the most simple behavioral tests, an alternative solution must be found for dealing with these large numbers of chemicals.
In addition, the Office of Toxic Substances estimates that approximately 1300-1500 new chemicals are introduced into commerce each year. Again using Anger and Johnson's estimate that 25% of these chemicals will be neurotoxic, it becomes apparent that the number of neurotoxic chemicals in commercial use is increasing and, therefore, there is an increasing need to evaluate the neurotoxicity of new chemicals. In my opinion, the test methods described today can have the greatest impact if they are used to detect these potentially neurotoxic chemicals before they are introduced into commerce rather than relying on testing after they are in the marketplace and after potential human exposure.

Considering the number of potentially neurotoxic chemicals and the fact that regulatory agencies do not currently require systematic testing for neurotoxicity, the answer to Dr. Lubin's question has to be "no". My personal opinion is that we are not currently addressing neurotoxicity at the level that we should, and there is a pressing need to introduce some kind of systematic approach which will allow us to detect these chemicals as they are coming into commercial use whether they be drugs or industrial chemicals.

HANIG: My name is Joseph Hanig with FDA. I am not speaking as a representative of FDA, but would like to express my own views, amplify the important point Larry Reiter made, and make another point as well.

The process for approving drugs at FDA is completely different from the evaluation of chemicals at EPA. As I understand it, drugs do not enter into the marketplace until they are tested and evaluated for a period that is in excess of a number of years, then they enter into the environment. Years ago (pre-1962) the FDA had a deadline after which, if they failed to raise objections, then things automatically went into the marketplace. I think that this is the way things are done at this point at EPA, which is totally different from the current FDA process.

One should be cognizant of the fact that there are important economic and policy-type components to the regulatory process. One of the problems that regulators face is that they are trying to mix policy, science, and state-of-the-art concepts while the public is either asking or demanding that the regulatory agencies, by virtue of promulgating regulations, guidelines, standards, etc., shall define the state-of-the-art and also be responsible for handling or ensuring the public safety. In the absence of (a) a national initiative on neurotoxicity testing, (b) the establishment of a research institute, or (c) the ability to separate research and guideline formulation from the regulatory component so there is no conflict of interest, we are going to continue to have
problems wherein those people who attempt to define the state-of-the-art on the basis of their scientific judgments will be running headlong into those who deal with the economic realities. We should ask ourselves whether we can afford to allow thousands of chemicals to enter into the environment every year, as I understand is occurring, or whether human life is so dear that we must take whatever economic disadvantages are associated with either slowing down or expanding and improving the whole process?

HUTCHINGS: Thank you very much for that comment. I want to thank everyone for these comments and for the responses of the Panel.

CALLAWAY: This has been a useful meeting, not only to FDA and the other regulatory agencies, but also for those scientific investigators who will return home with more information, and ultimately, for the general public. I would add my thanks to the speakers for their prepared comments, for the Panel, and especially for those individuals who asked probing questions and made important contributions of both information and opinion, all of which will be useful to the Panel in preparing the final report.

Let me add a word of thanks from Don and myself to the LSRO staff, especially to Richard Leukroth, who is the project coordinator, and Dr. Kenneth Fisher, Director of LSRO, for their excellent staff support.
LITERATURE CITED


IX. ACKNOWLEDGMENTS

A. SUBMITTED MATERIALS

The Life Sciences Research Office (LSRO) received information and/or reference material in response to the request for information published in the July 9, 1985 Federal Register announcement [50 FR 28034]. Five (5) individuals and organizations submitted information for consideration by the Panel on Predicting Neurotoxicity and Behavioral Dysfunction from Preclinical Toxicologic Data. One individual requested an opportunity to make a voluntary presentation on the scientific issues related to this subject during the general discussion period of the symposium held on September 30, 1985.

A copy of the submitted information was made available to all members of the study panel prior to the symposium. The members of the LSRO ad hoc Panel discussed these materials during the workshop and gave consideration to them in the preparation of this report. Copies are also available for inspection at the Food and Drug Administration, Dockets Management Branch, 5600 Fishers Lane, Rockville, Maryland 20857 and at the Life Sciences Research Office, FASEB, 9650 Rockville Pike, Bethesda, Maryland 20814. Both the LSRO and the members of the ad hoc Panel wish to thank these individuals for their contribution.

J.K. Marquis
Boston University
School of Medicine
Boston, Massachusetts

M.R. Davis
Department of Health and Human Services
Office of Technology Assessment
Washington, D.C.

R.W. Wood
University of Rochester
School of Medicine and Dentistry
Rochester, New York

J.A. Edwardson
Medical Research Council
Newcastle, United Kingdom

R. Singer*
Occupational Health Consulting, Inc.
402 East 90th Street, Suite 6B
New York, New York

* The full text of Dr. Singer's October 30, 1985 oral presentation submitted before the ad hoc Panel is available for purchase from ACE Federal Reporters, Inc., 444 North Capitol Street, Washington, D.C. 20001.
B. DISCUSSION REMARKS

The LSRO and the members of the ad hoc Panel would also like to express their appreciation to those symposium participants that contributed in the open and general discussions during the September 30, 1985 symposium.

Dr. Carl Boast
CIBA-GEIGY Corp.
556 Morris Avenue
Summit, New Jersey 07901

Dr. Joseph Hanig
Food and Drug Administration
200 C Street, S.W.
Washington, D.C. 20204

Dr. Joy Cavagnaro
Hazelton Laboratories
America, Inc.
9200 Leesburg Turnpike
Vienna, Virginia 22180

Dr. Jane E. Harris
U.S. Environmental Protection Agency
Toxicology Branch/HED
Washington, D.C. 20460

Dr. Miriam Davis
Office of the Assistant Secretary of Health
200 Independence Ave., S.W.
Washington, D.C. 20037

Dr. David G. Hattan
Toxicology Information Management Systems
Center for Food Safety and Applied Nutrition
Food and Drug Administration
200 C Street, S.W.
Washington, D.C. 20204

Bruce Donzanti
U.S. Army Medical Research Institute of Chemical Defense
Aberdeen Proving Ground,
Maryland 21010

James R. Ison
Department of Psychology
University of Rochester
Rochester, New York 14627

Dr. James L. Emerson
Coca-Cola Company
P.O. Drawer 1734
Atlanta, Georgia 30301

Dr. Carol A. Kimmel
U.S. Environmental Protection Agency
401 M Street, S.W.
Washington, D.C. 20460

Dr. Vera Glocklin
Office of Drug Research and Review
Food and Drug Administration
5600 Fishers Lane
Rockville, Maryland 20857

Dr. Tina E. Levine
U.S. Environmental Protection Agency
401 M Street, S.W.
Washington, D.C. 20460
Dr. Jonathan J. Lipman  
Vanderbilt University  
Medical School  
B2214 Medical Center North  
Nashville, Tennessee 37232

Dr. Peter Spencer  
Albert Einstein College  
1300 Morris Park Avenue  
Bronx, New York 10461

Dr. A. Harold Lubin  
American Medical Association  
535 N. Dearborn Street  
Chicago, Illinois 60610

Gary L. Wenk  
Johns Hopkins University  
Department of Psychology  
Baltimore, Maryland 21218

Dr. John L. O'Donoghue  
Eastman Kodak Company  
Kodak Park  
Rochester, New York 14650

Dr. Ronald Wood  
Department of Radiation,  
Biology and Biophysics  
University of Rochester  
Medical School  
Rochester, New York 14642

Dr. Raymond Singer  
Mt. Sinai School of Medicine  
402 E. 90th Street  
New York, New York 10128

Dr. John S. Young  
Johns Hopkins University  
615 N. Wolfe Street  
Room 7006  
Baltimore, Maryland 21205

Dr. Thomas J. Sobotka  
Center for Food Safety  
and Applied Nutrition  
Food and Drug Administration  
200 C Street, S.W.  
Washington, D.C. 20204

Robert P. Zendzian  
U.S. Environmental Protection  
Agency  
Office of Pesticide Programs  
401 M Street, S.W.  
Washington, D.C. 20460

Dr. Harold Zenick  
U.S. Environmental Protection  
Agency  
401 M Street, S.W.  
Washington, D.C. 20460