

## WORKSHOP OVERVIEW

# The IPCS Collaborative Study on Neurobehavioral Screening Methods<sup>1,2</sup>

VIRGINIA C. MOSER,\*<sup>3</sup> GEORGE C. BECKING,† ROBERT C. MACPHAIL,\* AND BEVERLY M. KULIG‡

\*Neurotoxicology Division, National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina 27711; †International Programme on Chemical Safety, World Health Organization, Research Triangle Park, North Carolina; and ‡Department of Neurotoxicology and Reproduction Toxicology, TNO Nutrition and Food Research, Zeist, The Netherlands

Received November 18, 1996; accepted November 20, 1996

The IPCS Collaborative Study on Neurobehavioral Screening Methods. MOSER, V. C., BECKING, G. C., MACPHAIL, R. C., AND KULIG, B. M. (1996). *Fundam. Appl. Toxicol.* 35, 143–151.

The International Programme on Chemical Safety sponsored a collaborative study to evaluate the utility of neurobehavioral test methods for identifying neurotoxic chemicals. The protocol consisted of a functional observational battery and automated assessment of motor activity. The study involved four laboratories in the United States and four in Europe, each of which evaluated the dose- and time-related effects of seven prototypic chemicals following both single and 4-week repeated exposures. The protocol was designed to assess the general utility and reliability of neurobehavioral screening procedures in a diversity of testing situations. The results of chemical testing indicated that all participating laboratories generally could detect and characterize the effects of known neurotoxicants, despite some differences on specific endpoints. These data provide important information regarding the reliability and sensitivity of neurobehavioral screening methods over a range of laboratory conditions. The purpose of this workshop was to describe the background and study design of the collaborative effort, present the data (including comparison of results across laboratories), and discuss issues regarding the conduct and interpretation of these behavioral tests, as well as future directions for neurotoxicity screening. © 1997 Society of Toxicology.

### INTRODUCTION—V. C. MOSER, U.S. EPA

Screening for the neurotoxic potential of chemicals is now recognized as an integral component of toxicity testing. A functional observational battery (FOB), along with an auto-

mated assessment of motor activity, is widely used for neurobehavioral testing (Tilson and Moser, 1992). The FOB is a type of neurological exam that includes a series of observations and measurements taken (1) in the subject's home cage, (2) while the subject moves about in an open field, and (3) during manipulative tests. The FOB end points are intended to evaluate the subject's neuromuscular integrity, responses to sensory stimuli, and signs of autonomic nervous system activity (Gad, 1982; Irwin, 1968; Moser *et al.*, 1988), whereas motor activity is viewed as a more apical measure of neurobiological integrity (Reiter and MacPhail, 1979). The protocol used in this collaborative study consisted of approximately 30 measures, which were grouped into domains according to the neurobiological functions evaluated by each test (see McDaniel and Moser, 1993).

These tests are currently recommended by the U.S. Environmental Protection Agency (U.S. EPA) to screen chemicals for neurotoxic potential, and the results of this collaborative study are timely in light of ongoing efforts to harmonize testing guidelines worldwide. A full report of this project will soon be published in *NeuroToxicology*. This workshop described the genesis and conduct of the collaborative study, the results, recommendations for modifying the behavioral screening battery, and future developments anticipated in neurotoxicity testing. The results provide important information regarding the reliability and sensitivity of neurobehavioral screening methods over a range of laboratory conditions and the utility of these methods for identifying and characterizing neurotoxic hazards. Thus, this collaborative study represents an important step toward acceptance of neurobehavioral screening methods on an international level.

### ROLE OF THE INTERNATIONAL PROGRAMME ON CHEMICAL SAFETY IN THE DEVELOPMENT OF AN INTEGRATED APPROACH FOR ASSESSMENT OF THE NEUROTOXICITY OF CHEMICALS—G. C. BECKING, IPCS/WHO

The presentation was designed to provide an overview of the International Programme on Chemical Safety (IPCS) and

<sup>1</sup> Presented at the 35th Annual Meeting of the Society of Toxicology, Anaheim, CA, March 1996.

<sup>2</sup> This paper has been reviewed by the National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation of use.

<sup>3</sup> To whom correspondence should be addressed. Fax: 919-541-4849

information on the scientific activities within IPCS related to neurotoxicology. Since its inception in 1980, IPCS has placed a major emphasis on the neurotoxicity of man-made and natural chemicals. The program is managed by the World Health Organization (WHO) in Geneva on behalf of the other partners, the United Nations Environmental Programme (UNEP) and the International Labour Organization (ILO). The Scientific Secretariat provides technical advice and assistance in coordinating the scientific activities carried out by IPCS participating institutions, three of which participated and provided major financial and intellectual support for the IPCS Collaborative Study on Neurobehavioral Screening Methods.

Two major interrelated objectives of the IPCS are (1) evaluation of the health and environmental risks from exposure to chemicals and (2) to improve the scientific foundation for human health risk assessment. Obviously, success in the second objective will assist in the preparation of scientifically sound evaluations of chemical risks for use by member nations worldwide.

Since 1982, IPCS has carried out several projects in neurotoxicology in an effort to develop an internationally accepted integrated approach for assessing the neurotoxic potential of chemicals. The collaborative study is part of this objective. Other activities include (1) publication in 1986 of an Environmental Health Criteria (EHC-Methodology) entitled "Principles and Methods for the Assessment of Neurotoxicity Associated with the Exposure to Chemicals" (WHO, 1986); (2) preparation of an EHC-Methodology entitled "Use of *in Vitro* Tests in Neurotoxicology"; (3) collaboration with the Organization for Economic Cooperation and Development (OECD) in developing test guidelines for neurotoxicity and the preparation of a guidance document on testing strategies for neurotoxicity; and (4) international efforts to harmonize neurotoxicity risk assessment methodologies.

IPCS-sponsored activities in neurotoxicology will be expanded as member countries make recommendations, but only if support is forthcoming. For example, this IPCS collaborative study was a strong recommendation of the scientists that prepared EHC 60 (1986). No work would have been initiated, however, without the support of the U.S. EPA and its National Health and Environmental Effects Research Laboratory. Also, support by the U.S. National Institute of Environmental Health Sciences/National Toxicology Program (NIEHS/NTP), the Commission of the European Union, University of Bari, and the National Council of Research, Italy, was substantial, as were the commitments made by each of the laboratories participating in the collaborative study.

#### THE IPCS COLLABORATIVE STUDY: BACKGROUND, PROTOCOL DESIGN, AND TESTING PROCEDURES— R. C. MACPHAIL, U.S. EPA

The past two decades have witnessed a substantial increase in research into the neurotoxicity of environmental

chemicals. Against a backdrop of numerous episodes of neurotoxic poisoning in humans and other animals (see NRC, 1991, 1992; OTA, 1990), several expert committees recommended increased attention to screening compounds for neurotoxic potential (see Sette, 1989). These recommendations have consistently included functional (in-life) assessments of chemical-exposed laboratory animals along with assessment of pathology within regions of the nervous system. While cage-side functional assessments have been carried out in toxicity testing for several decades, basic uncertainties over details of the assessments have led to formalized approaches for hazard identification in neurotoxicology. The U.S. EPA, for example, has published recommendations for neurotoxicity screening that include an FOB, motor activity, and neuropathology (U.S. EPA, 1991). These screening tests are consistent with those recommended by several national and international expert committees.

The likelihood of widespread application of neurotoxicity screening tests in science and regulation raises basic questions regarding the reliability of the tests and the generality of test results. Collaborative interlaboratory comparisons are indispensable for this purpose. In fact, an expert committee convened by WHO reviewed the state of the science and concluded that "test development in preclinical neurotoxicology has evolved to the point where interlaboratory validation of procedures using prototypic neurotoxic agents could be attempted" (WHO, 1986). Ideally, such collaborative studies should be carried out on an international scale and in a variety of research environments. A collaborative study was therefore sponsored by IPCS on the use of a neurotoxicity screening battery to evaluate environmental chemicals.

A steering committee was established by IPCS to oversee the collaborative study. The steering committee was comprised of scientists from different countries who were responsible for recruitment of participating laboratories, protocol design, and study coordination. Members included the authors of this report, as well as Drs. V. Cuomo (University of Bari, Italy), E. Frantík (Institute of Hygiene and Epidemiology, Czech Republic), H. Tilson (U.S. EPA), and G. Winneke (Institute of Environmental Hygiene, Germany). Through advertisements and announcements at a number of scientific congresses, eight laboratories (listed in Table 1) ultimately agreed to participate. The laboratories included academic, government, and contract laboratories.

The protocol specified evaluations of the effects of both acute and repeated (4-week) exposures to toxicants. Adult male laboratory rats were to be tested before, during, and after exposure. Test methods consisted of an FOB and an automated assessment of motor activity. Chemicals included those that were either expected to produce different patterns of neurotoxic effect or judged likely to be without neurotoxic potential.

A training video and manual were prepared by the steering committee. In addition, 2-day training sessions were held in

**TABLE 1**  
**Laboratories Participating in the IPCS Collaborative Study**  
**on Neurobehavioral Screening Methods**

Europe
Institute of Pharmacology, School of Medicine, University of Bari, Bari, Italy
Ministry of Health, National Food Agency, Soborg, Denmark
National Institute of Public Health, Prague, Czech Republic
TNO Nutrition and Food Research, Rijswijk, The Netherlands (now located in Zeist, The Netherlands)
United States
Bushy Run Research Center, Export, PA (now closed)
G. D. Searle and Co., Skokie, IL
ManTech Environmental Technology, Inc., Research Triangle Park, NC (under contract to U.S. Environmental Protection Agency)
National Institute for Occupational Safety and Health, Cincinnati, OH (interagency agreement with National Toxicology Program)

Research Triangle Park, North Carolina, for one representative from each of the participating laboratories. The training sessions included review of the tests and their scoring criteria, plus evaluations of control and toxicant-treated laboratory rats. Details of the testing protocol were also reviewed.

Each laboratory then carried out pilot studies to demonstrate their proficiency with the test methods (Moser and MacPhail, 1992). The following treatments were used: (1) a single dose of parathion, (2) a single dose of *p,p'*-DDT (DDT), and (3) short-term (1 to 2 weeks) repeated dosing with acrylamide. The steering committee identified several cardinal signs that each laboratory should obtain following these treatments (e.g., tremor produced by DDT, cholinergic signs with parathion, and hindlimb weakness following acrylamide). Proficiency also included establishing dose-effect functions to demonstrate the ability to detect triadimefon-induced increases in motor activity and chlorpromazine-induced decreases in activity. The proficiency phase was considered an essential step prior to the formal data-collection phase of the study to insure a comparable level of experience in testing and evaluation.

Formal studies consisted of evaluating the dose-response and time-course characteristics of seven chemicals: acrylamide, *N,N'*-methylene-bisacrylamide (bis-acrylamide), DDT, lead acetate, parathion, toluene, and triethyl tin (TET). Some of these chemicals were expected to produce different patterns of effect when given acutely and repeatedly. Chemicals with different putative mechanisms of action were also included in order to evaluate the ability of the methodology to differentiate compounds acting at different sites within the nervous system. For example, autonomic effects should be a prominent feature of parathion exposure, while tremor and hyperreactivity should be observed following DDT. Both TET and acrylamide should produce hindlimb weakness, but through different mechanisms. Bis-acrylamide is a

nonneurotoxic analog of acrylamide, while lead is thought to be relatively nonneurotoxic when given to adult laboratory rats. Each experiment included a vehicle-control group plus four dose groups. All chemicals were supplied to each laboratory from a common source and were pure or of known composition. Chemicals were identified only by a unique code assigned by the supplier, and testing was to be conducted with the observer unaware of the treatment. A designate at each laboratory received health and safety information for the chemicals in case of a spill or other accident.

While the tests, chemicals, and dosing durations were standardized in the study protocol, several key variables that could affect the outcome of the experiments were not standardized. For example, no effort was made to standardize the strain or stock of rat, husbandry conditions, or the device used for motor activity assessments. The steering committee decided it would be unrealistic to dictate all details of the experiments for all laboratories. Test results should therefore be viewed as systematic replications of each experiment under laboratory conditions likely to be encountered in research in support of regulatory decision making.

The lack of standardization of several variables necessitated an algorithm to select dosages for each toxicant. Each laboratory received the starting dose for a given chemical. Pilot experiments were then carried out with one rat each receiving the starting dose, three doses higher, and three doses lower, with a fixed dose spacing (increments and decrements). A "top dose" (TD) was defined as the highest dose that did not produce lethality within seven days after administration. In no case, however, did the maximum exceed a limit dose of 2 g/kg. Time of peak effect (TOPE) was next determined with four rats receiving the TD and four receiving vehicle. Functional assessments were conducted with an abbreviated battery at half-hour intervals for the first 2 hr, hourly intervals for the next 4 hr, and again 24 hr after dosing. In addition, if more than one treated rat died within 7 days, the TD and TOPE determinations were repeated at the next lower dose. In this way, laboratories were equated with regard to the acute potency and the time course of action for each toxicant. Doses for the formal testing phase of the collaborative study were then expressed as proportions of the TD determined for each toxicant in each laboratory: 100, 50, 25, and 12.5% TD for acute dosing and 50, 25, 12.5, and 6.25% TD for repeated dosing. With the exposure parameters thus derived, formal testing of the seven chemicals was initiated.

Several collaborative studies have been conducted in toxicology that have had varying degrees of stringency in the testing protocol (e.g., Christenson *et al.*, 1994; Claxton *et al.*, 1992; van den Heuvel *et al.*, 1987). The expectation, of course, is that more stringent protocols will lead to more reliable results. It is rarely the case, however, that all relevant variables can be identified and controlled in an experiment. As a result, the laboratory typically has a significant impact

on the results of collaborative studies—in much the same way that differences in results can occur in an individual laboratory conducting direct replications of the same experiment (e.g., Dews and Berkson, 1954; Frantík *et al.*, 1994). It is therefore unrealistic to expect a complete conformity of results across all laboratories in any collaborative study. As a consequence, a principle focus of the present collaborative study was to evaluate the amount of variability that is likely to occur with the test methods.

#### RESULTS OF CHEMICAL TESTING AND CONTROL DATA IN THE IPCS COLLABORATIVE STUDY— V. C. MOSER, U.S. EPA

This collaborative study was conducted to evaluate the utility of a neurobehavioral screening battery for identifying neurotoxic substances. To be acceptable for screening, the FOB and motor activity measurements should be able to identify potential neurotoxic agents, detect dose- and time-dependent effects following an acute or repeated dosing, and distinguish these effects from those of nonneurotoxic compounds. The protocol, as described above, was developed to address these issues.

Two exposure regimens were used for each chemical in the study. In the single-dose study, rats were dosed with either vehicle or one of four dosages (see above) of the test chemical. They were tested before dosing, on the day of dosing at the TOPE, and at 1 and 7 days after dosing. For the repeated-dose study, rats were dosed 5 days/week for 4 weeks with either vehicle or one of four dosages. Testing occurred before dosing began, at the end of the second and fourth weeks of dosing, and then 2 weeks after dosing ended. On each test day, rats were tested with the FOB followed shortly thereafter by motor activity.

Since all eight laboratories were to conduct acute and repeated-dosing studies with each of the seven chemicals, the total number of studies in this collaborative effort was 112. However, a few laboratories deviated from the protocol, and a few of these deviations were deemed serious enough to exclude the data from the formal comparisons across laboratories. Still, there were 102 studies to examine, and data were available from no fewer than seven laboratories for any one chemical.

The data collected with the screening battery include categorical responses, which record the presence or absence of a sign; ordinal data, which include ranking the severity of a sign; and continuous data, which display a range of values usually approximating a normal distribution. All measures were collected from the same rats at multiple time points. The continuous data were analyzed with traditional analysis of variance techniques, while descriptive and ordered-response (rank) data were analyzed using a categorical data modeling procedure (see Creason, 1989).

Because of the large amount of data, an additional step

was taken to integrate and reduce the data. The data were grouped into functional domains (autonomic, activity, convulsive, excitability, neuromuscular, and sensorimotor). To enable comparison of endpoints within and between domains, a severity scoring procedure was used to convert individual data to scores ranging from 1 to 4. The conversion was based on the concurrent control data, such that a 1 represented data that occurred most frequently in the control rats, and a 4 indicated data that rarely occurred in controls (McDaniel and Moser, 1993; Moser *et al.*, 1995).

The protocol described how each laboratory would determine their TD for each chemical. The TD was critical since the laboratories used fractions of the TD for the dose-response assessments. In general, there was good agreement in the resultant TDs. All but one or two laboratories agreed on a top dose for acrylamide (113 mg/kg), bis-acrylamide (147 mg/kg), lead acetate (200 mg/kg), and toluene (2 g/kg; limit dose). There was, however, more of a range of TDs for the other chemicals, with no more than three laboratories choosing the same dose: DDT (58–196 mg/kg), parathion (3–10 mg/kg), and TET (4–9 mg/kg).

Under acute conditions, only 3 of 51 studies recorded >50% lethality, and this occurred only in the highest dose group (100% TD). For the repeated-dosing studies, however, there was considerable lethality at the highest dose (50% TD) and sometimes at the next highest dose (25% TD). In 32 of 51 studies at least one dose group was lost to lethality. This compromised the data from the repeated-dose studies, especially those of acrylamide, lead acetate, and TET. Thus, while this dose-selection method generally worked well for single-dose studies, additional range-finding studies are clearly needed for repeated-dosing studies. The spacing of doses worked well for most compounds, except for those with a steep dose response (e.g., parathion). We concluded that as much information as possible should be utilized from range-finding studies in order to select the dose range for testing.

To determine the appropriate testing time on the day of dosing, each laboratory ran preliminary TOPE studies. There was generally good agreement in the choice of the TOPE. Acrylamide, bis-acrylamide, and DDT showed the longest peak-effect times (up to 6 hr), while the others were shorter (1–4 hr). This method of choosing a TOPE worked fairly well for acutely neurotoxic compounds. It was more difficult to determine a TOPE, however, for compounds showing either nonspecific effects or a biphasic pattern of changes, and as a result a wider range of times was chosen. The choice of test times was critical for the ability to detect acute effects.

Analysis of the control data included examining baseline data in all rats (since all rats were tested before dosing), as well as the vehicle-control group in each study. Within each laboratory, the variability within each study depended on the endpoint, with some measures more variable than others.

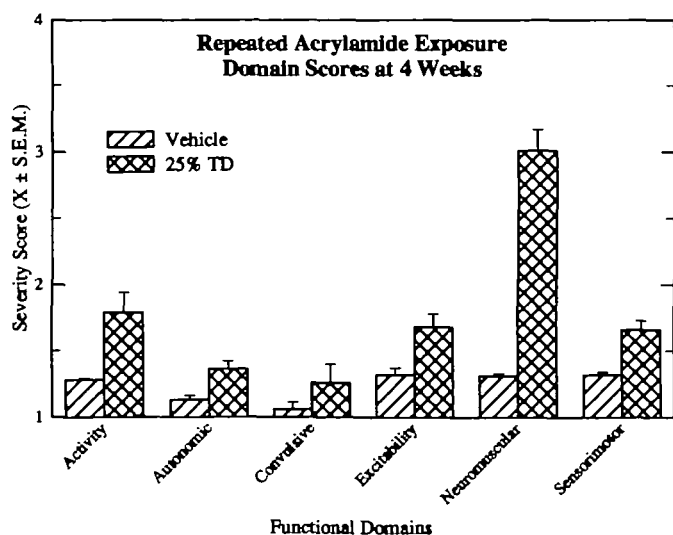


FIG. 1. Domain scores representing the neurotoxicity of acrylamide after 4 weeks of dosing (5 days/week). The scores for each functional domain are averaged across all laboratories, and data are presented as means  $\pm$  SEM. For clarity, only the vehicle and the 25% TD groups are shown.

Variability of group mean data across studies, however, was generally  $<20\%$ . Baseline values across laboratories were quite similar for some endpoints (e.g., rearing), but varied greatly for others (e.g., motor activity). Variability of group mean control data was greater across laboratories than within a laboratory. For some end points, control variability influenced the laboratory's ability to detect significant changes. For example, one laboratory recorded high variability in motor activity; as a result, treatment-related changes were never detected.

The data from the chemical studies from each laboratory were interpreted in light of what was expected, based on the published literature and the experience of the steering committee. The functional domain analysis was used to focus on the major effects of each chemical. To accomplish this, the mean domain scores from all laboratories were averaged and analyzed. The magnitude of the effect, or effect size, was then compared across domains. Once the domains with the largest effects were determined, those domains were examined most closely to see which endpoints were affected and in which laboratories.

Acrylamide is a prototypic compound producing a central-peripheral distal axonopathy (Kulig, 1994; Tilson, 1981). In this collaborative study, repeated exposure to acrylamide produced considerable effects on motor function. To illustrate this, Fig. 1 presents the mean domain scores of the vehicle and the 25% TD groups (which was the highest dose group surviving 4 weeks of dosing). All laboratories documented pronounced neuromuscular domain changes, which are evident in Fig. 1 by comparing the effect size of the different domains. These effects were due mostly to a progressive pattern of increased landing foot splay, de-

creased grip strength with a preferential effect on the hindlimbs, and pronounced ataxia and other gait abnormalities; all hallmarks of acrylamide neurotoxicity (Kulig, 1994).

Bis-acrylamide, which does not produce an acrylamide-like axonopathy (Edwards, 1975), did not show the same pattern of effects with repeated dosing. While the neuromuscular domain was most affected, this was due to decreased landing foot splay (in contrast to increased splay with acrylamide), decreased grip strength which was not specific for the hindlimbs, and only slightly altered gait. Furthermore, weight gain was depressed considerably more than with acrylamide, and hypothermia was recorded throughout dosing: these general toxic effects were much more prominent than the behavioral changes.

Acute administration of acrylamide was expected to produce incoordination, lowered activity, and tremors (Kulig, 1994; Tilson, 1981). Indeed, the predominant effect of acute acrylamide was in the activity domain, and most laboratories reported decreased motor activity. Half of the laboratories obtained increased reactivity to being handled or removed from the home cage; this accounted for significant effects in the excitability domain. Neuromuscular domain effects consisted of altered gait, and in the convulsive domain, most of the laboratories reported mild tremors. Somewhat similar effects were obtained with acute administration of bis-acrylamide, in that activity domain changes (decreased motor activity) were the most pronounced effect and there were reports of altered gait. In contrast to acrylamide, bis-acrylamide did not produce tremors or changes in reactivity.

A single dose of DDT is known to produce tremors, myoclonus, increased reactivity to auditory stimuli, and hyperthermia (Henderson and Woolley, 1970; Woolley, 1982). In the acute study, the convulsive domain was affected in almost all laboratories, and increased body temperature was recorded in all. In addition, half of the laboratories detected an increased reaction to the click (auditory) stimulus. These effects were obvious in all laboratories except one, which used a time point of only 1 hr. Since DDT effects are known to be protracted (Hudson *et al.*, 1985), it may be assumed that the TOPE used in this laboratory was the cause of their failure to detect the classic signs of DDT toxicity.

Most laboratories also detected the acute syndrome produced by parathion, which appears as autonomic signs (salivation and miosis), motor effects (decreased motor activity, altered gait, and deficits in righting response), and mild tremors (Murphy, 1980; Taylor, 1985). Most parathion effects were seen only at the highest dose, and in most laboratories the two lowest doses produced no effects whatsoever. Thus, the dose spacing was probably too wide for this particular compound. As was the case with DDT, one laboratory did not report the classic autonomic signs of toxicity; this laboratory explained that the autonomic effects occurred at a different time than the motor effects, which the protocol specified to be one of the criteria for TOPE selection.

Tolerance is known to develop to both DDT and parathion (Hayes, 1982; Hoskins and Ho, 1992). As expected, all laboratories reported no effects in the repeated-dosing studies with these two chemicals.

TET produces myelinopathy which is manifest as generalized weakness, depression, and motor incoordination (Reiter and Ruppert, 1984; Watanabe, 1980). Neuromuscular domain effects were a prominent feature of TET's profile, due to altered gait, lowered grip strength, and a pronounced effect on righting reaction. There were also prominent effects on the activity (lowered motor activity) and excitability (decreased reactivity) domains. Repeated dosing with TET produced lethality in at least the two highest dose groups in all laboratories, compromising the remaining dose-response data. Nonetheless, a similar profile was obtained with the same neuromuscular domain effects.

Lead acetate was expected to produce nonspecific changes, if any (Pryor *et al.*, 1983). Most laboratories reported lowered motor activity and decreased arousal in the single-dose studies, but hypothermia and weight loss were the more pronounced effects. With repeated dosing, lead acetate became even more toxic, and all laboratories lost most or all of the two or three highest dose groups. As in the acute studies, predominant behavioral effects were lowered arousal and motor activity. These behavioral changes, however, were overshadowed by the extreme toxicity evident in the surviving rats which lost weight and were hypothermic during the 4 weeks of dosing.

The seventh chemical was toluene, a solvent that produces generalized CNS depression (Benignus, 1981). Only one laboratory detected these effects of toluene, while the others showed good agreement in not finding any effects. Most laboratories did report a decreased tail-pinch response, but only the one laboratory recorded gait and righting changes, decreased grip strength, and biphasic effects on activity levels. It is important to note that toluene was administered orally to nonfasted rats, and the highest dose was limited to 2 g/kg. Solvents in general require higher doses when given orally to produce effects, and it is possible that the doses were not high enough: alternatively, the rats in the one laboratory may have been genetically more sensitive. Repeated dosing with toluene produced no effect in any laboratory.

Thus, the data from all laboratories reflected the detection and characterization of the effects of most of the known neurotoxicants. The exception was toluene, for which only one laboratory detected motor incoordination and CNS depression. All laboratories reported activity and reactivity signs of toxicity following a single dose of acrylamide, along with progressive neuromuscular toxicity during repeated dosing. In addition, data from all laboratories indicated TET to be neurotoxic, producing neuromotor dysfunction following single and repeated dosing. The acute DDT neurotoxic syndrome was observed in all laboratories, except the one which used the shortest TOPE. Likewise, acute parathion

produced cholinergic signs in all laboratories, except in the laboratory which felt they had used an inappropriate TOPE. There were some differences in effects across laboratories on specific end points, but these did not impact greatly the overall profile of each chemical.

All laboratories obtained statistically significant changes in particular measures which varied from chemical to chemical, i.e., the chemicals produced different profiles of effect. These profiles differentiated the known neurotoxicants from the nonneurotoxicants; e.g., the effects of repeated exposure to bis-acrylamide were distinct from those of acrylamide. Lead acetate generally lowered activity and reactivity levels, but the pronounced general toxicity overshadowed the neurological effects. The chemical profiles also revealed differences between the patterns of effect following acute and repeated exposures, e.g., tolerance to the effects of DDT and parathion and progressive neurotoxicity with acrylamide and TET. Thus, this study provides practical data regarding the use of neurobehavioral screening methods over a range of laboratory conditions as well as the ability of the tests to detect potential neurotoxicity.

**NEUROTOXICITY SCREENING: METHODOLOGICAL  
ISSUES AND FUTURE DEVELOPMENTS—B. M. KULIG,  
TNO NUTRITION AND FOOD RESEARCH INSTITUTE,  
THE NETHERLANDS**

Despite the documented importance of neurotoxic exposures to human health in individual countries (Schaumburg and Spencer, 1980), it has been only in the past 10 years that the problem has received international recognition. Although morphological data have been historically used for determining neurotoxicity, there has been increasing interest in the use of behavioral endpoints for evaluating the effects of toxic exposures on the nervous system. This IPCS collaborative study was one of the first efforts to address this issue. Given the unfamiliarity of most toxicologists with behavioral methods and the potential of these methods to significantly contribute to screening strategies, it is not surprising that this collaborative study focused upon these methods for study.

Cage-side observations of neurological and behavioral changes have been part of toxicologic screening studies for many years, but it is only recently that more systematic observational methods have been incorporated into a clinical screening battery for the purpose of identifying neurotoxicants (Moser, 1989; O'Donoghue, 1989). The use of behavioral methods has been controversial and many questions have been raised as to their validity and reliability for neurotoxicity screening. The IPCS collaborative study provided the opportunity to gain practical knowledge in carrying out behavioral neurotoxicity screening studies, and to help clarify some of the issues surrounding the use of these methods. The purpose of this presentation is to discuss some practical considerations in the application of behavioral techniques

which can affect the reliability and validity of these measures, as well as future directions in the area of neurotoxicity assessment.

Observation is the backbone of every scientific discipline, and thus there is nothing particularly novel in the use of observation in science and medicine. The observation of behavior, however, does require some unique knowledge and experience in order to be meaningfully applied to neurotoxicity assessment. The first, and seemingly simplistic, axiom for using observational methods to identify neurotoxic effects is *you have to look at the animal and its behavior*. Although self-evident, this is not as easy as it may sound. To an untrained and casual eye, a laboratory rat's behavior may appear to be rather dull and uninteresting. On the other hand, if one closely observes the behavior of that rat over a period of time, the variety of its activities can be somewhat overwhelming. How to understand such a plethora of behavioral events, or whether meaningful scientific data can even be extracted from such observations, has, in fact, been at the heart of much of the debate surrounding the use of behavioral methods in neurotoxicology.

The second axiom in the use of behavioral methods is *behavioral observations are subject to observer bias*. What influences our perceptions at any given moment may be difficult to determine, but it is obvious that humans tend to categorize and interpret what they see. A technique for reducing experimenter bias is to employ blinding procedures which prevent the observer from knowing *a priori* which animal received a treatment. It has been argued that blinding is more a matter of form than substance in toxicology experiments since the observer will usually be able to tell that a given rat has received some experimental treatment. Although this may be true for animals receiving high doses of the test compound, it is not necessarily the case at lower dose levels. Since most toxicology studies are conducted with controls and at least three dose groups and have as one of their aims the determination of a no-effect level, the most unbiased evaluation of effects at the lower doses is of obvious importance.

In addition to observer bias, there are also other potential confounders which must be taken into account when using behavioral measures. These include the effects of repeated measurements, order effects, time-of-day effects, apparatus effects, and the effects of extraneous environmental stimuli, including those produced by the observer. Luckily, experimental design techniques (e.g., counterbalancing) are available to minimize observer bias and other potential sources of confounding. Table 2 lists some of these techniques.

Another important technique in reducing artifacts in behavioral experiments is the use of controls. Both experimental control and statistical control are necessary for conducting valid behavioral experiments.

One of the major differences between the standardized observational battery used in the IPCS collaborative study

**TABLE 2**  
**Experimental Design Features for Minimizing Artifacts in Behavioral Experiments**

Operational definitions
Blinding of observer to treatment
Control Procedures
Experimental control to insure constancy of experimental conditions
Control tests for all groups to check equivalency of baseline responding before treatment
Control groups to control for the effects of experimental manipulations and the effects of repeated measurements
Balancing to control for
The effects of the testing device
The order of testing
The time of day
Statistical analysis

and the cage-side observations commonly used in general toxicity studies is the specification of the conditions under which an observation is made and the operational definition of behavior and the severity of an effect. Although observational methods remain subjective in nature, the use of operational definitions allows other investigators to determine what was and was not looked for and/or observed. Of course, despite the inclusion of many different endpoints, it is always possible that a neurological or behavioral abnormality will occur which is not described in the standardized battery. For this reason, the category "Other" will always be a part of any set of observations, no matter how extensive. This problem cannot be addressed by strategies in experimental design, but rather by training, experience, and a thorough familiarity with the behavioral repertoire of the test animal.

A major challenge in the collaborative study involved the different types of measurement scales for the various endpoints (nominal, ordinal, and interval scales) which require different statistical approaches. Nonparametric tests are available for the evaluation of differences among groups for variables measured on nominal and ordinal scales, whereas variables measured on interval scales are typically analyzed using parametric statistics. Both parametric and nonparametric tests are available in many of the commercially available software packages, together with corresponding *post hoc* comparisons of group differences.

One of the difficulties with including a large number of measures in a test battery is the chance of finding a significant difference when in fact none exists. The different types of data collected in the collaborative study were therefore converted and combined to provide overall scores for different functional domains which could be analyzed parametrically. Whether such an approach is the most appropriate for treating neurotoxicity-screening data is still an open question.

Overall, one of the conclusions of the collaborative study is that behavior can be a useful tool for assessing neurotoxicity. Examination of the coefficients of variation in control

TABLE 3  
Future Directions in Neurotoxicology

Refinements in application, e.g., time of peak effect determination and dose selection
Refinements in statistical analysis
Multidisciplinary approach
Mechanistic studies
Inclusion of biochemical markers
<i>In vitro</i> models for neurotoxicity screening
Specific tests for sensory functions and for learning, memory, and/or performance
Extrapolation models
PBPK modeling in conjunction with effects testing
Quantitative risk assessment models

animals obtained for the different measures provides evidence for the consistency and reliability of behavioral measures, and the different profiles of effect obtained for different chemicals demonstrates their specificity. The tests cannot, however, be applied haphazardly. Their application requires careful attention to details of experimental design, observer training, and experience in order to maximize their effectiveness as screening instruments.

A list of possible future developments in neurotoxicity screening is presented in Table 3. In some cases, these developments reflect refinements in existing measures, while in others, they involve the inclusion of new types of information to achieve a more comprehensive picture of the risks associated with a particular chemical exposure.

One example of a refinement with existing measures is the method for determining the TOPE. In this collaborative study, determination of the TOPE was based on the time of maximum changes in arousal and/or gait. One problem with this approach is that some important neurological signs may be missed if studies are conducted based only on these two scores. For example, not all laboratories were able to detect significant changes in the autonomic domain with acute parathion treatment. This was probably not due to a lack of observational acumen, but rather to the fact that the time course for different effects was not unitary. For example, using chlorpyrifos (another organophosphate cholinesterase inhibitor) we have found that at a time when gait changes were maximal, no clonic movements could be detected and autonomic signs had either returned to baseline or were only marginally affected (Kulig, unpublished data). Given these findings, it is apparent that the determination of the TOPE could be improved by including several simple observations of autonomic and convulsive effects.

In addition to refinements of existing techniques and, as discussed above, their statistical analysis, future developments in neurobehavioral testing will likely include quantitative tests of sensory function as well as tests of memory, learning, and/or the performance of learned behavior. With respect to sensory function, the sensitivity of the simple tests

of reactivity to visual, somatosensory, and auditory stimuli was not directly addressed in this collaborative study. Thus, no conclusions can be drawn as to the sensitivity of these techniques compared to other measures of sensory function. Although none of the laboratories was able to detect changes in reactivity to the click response in animals treated with toluene, this may have been due to a lack of effects with the dosing regimen rather than a lack of sensitivity of the test method. However, only a direct comparison of different methods will answer this question.

With respect to screening for possible effects on cognitive processes, no tests of memory, learning, or performance were included in the collaborative study. This is particularly unfortunate since changes in cognitive abilities have been reported in many human studies of neurotoxic exposures. Although many tests have been developed to examine changes in cognitive processes of laboratory animals, which tests are the most cost-effective for screening purposes is still under discussion.

Although observational batteries are one of the most practical approaches to neurotoxicity screening, these observations, no matter how well they are conducted, can only suggest a possible locus or mechanism of action of a chemical. Although the collaborative study focused on the use of behavioral screening methods, few neurotoxicologists would deny the importance of a multidisciplinary approach. Neuroscience research on the cellular and molecular mechanisms underlying the behavioral effects produced by chemicals has made great strides in the last decade. From this research, neurochemical, electrophysiological, and histopathological techniques are evolving which have increased our understanding of how and where chemicals work in the nervous system, while *in vitro* test systems are being increasingly advocated as an adjunct to neurotoxicity assessment. Although some techniques may prove impractical for screening purposes, others may prove useful. The quantitative determination of glial fibrillary acidic protein (GFAP), for example, has been advocated as a possible neurochemical approach for neurotoxicity screening (O'Callaghan, 1988). We have found GFAP determinations to be particularly useful for establishing the specificity of a particular compound's effects. For example, one of the difficulties in the repeated-dose lead acetate experiments was mortality in the high-dose groups. Such findings make it difficult to determine whether the effects are indicative of neurotoxicity or more generalized toxicity. We repeated the lead study using three dose groups and a control group, with the highest dose equal to the lowest dose used in the collaborative study. Interestingly, we found marked regional changes in GFAP levels even at the lowest dose level, thus leaving little doubt that the brain is a target for lead (Kulig, unpublished data).

Finally, one area which can be expected to gain increasing attention is the application of extrapolation models in neurotoxicity risk assessment. We need to know not only what the



compound does to the body but what the body does to the compound. Significant progress is being made in the development of physiologically based pharmacokinetic (PBPK) modeling which can be used for both high- to low-dose and cross-species extrapolation. As neurotoxicology matures and its multidisciplinary essence receives greater recognition, the emphasis will shift from questions regarding what methods should be used for detecting neurotoxic effects in laboratory animals to what the effects mean for human health.

## REFERENCES

- Benignus, V. A. (1981). Neurobehavioral effects of toluene: A review. *Neurobehav. Toxicol. Teratol.* **3**, 407–415.
- Christenson, W. R., Van Gaethem, D. L., Schroeder, R. S., Wahle, B. S., Dass, P. D., Sangha, G. K., and Thyssen, J. H. (1994). Interlaboratory cholinesterase determinations and the effect on results of statistical evaluation of cholinesterase inhibition. *Toxicol. Lett.* **71**, 139–150.
- Claxton, L., Lewtas, J., Becking, G., and Shelby, M. (1992). Collaborative study on complex mixtures. *Mutat. Res.* **276**, 1–144.
- Creason, J. (1989). Data evaluation and statistical analysis of functional observational battery data using a linear models approach. *J. Am. Coll. Toxicol.* **8**, 157–170.
- Dews, P. B., and Berkson, J. J. (1954). On the error of bio-assay with quantal response. In *Statistics and Mathematics in Biology*. (O. Kempthorne, T. A. Bancroft, J. W. Gowen, and J. L. Lush, Eds.), pp. 361–370. Iowa State College Press, Ames, IA.
- Edwards, P. M. (1975). Neurotoxicity of acrylamide and its analogues and effects of these analogues and other agents on acrylamide neuropathy. *Br. J. Ind. Med.* **32**, 31–38.
- Frantik, E., Hornychova, M., and Horvath, M. (1994). Relative acute neurotoxicity of solvents: Isoeffective air concentrations of 48 compounds evaluated in rats and mice. *Environ. Res.* **66**, 173–185.
- Gad, S. C. (1982). A neuromuscular screen for use in industrial toxicology. *J. Toxicol. Environ. Health* **9**, 691–704.
- Hayes, W. J. (1982). *Pesticides Studied in Man*, pp. 172–205. Williams and Wilkins, Baltimore, MD.
- Henderson, G. L., and Woolley, D. E. (1970). Mechanisms of neurotoxic action of 1,1,1-trichloro-2,2-bis(*p*-chlorophenyl)ethane (DDT) in immature and mature rats. *J. Pharmacol. Exp. Ther.* **175**, 60–68.
- Hoskins, B., and Ho, I. K. (1992). Tolerance to organophosphorus cholinesterase inhibitors. In *Organophosphates: Chemistry, Fate, and Effects*. (J. E. Chambers and P. E. Levi, Eds.), pp. 285–299. Academic Press, New York.
- Hudson, P. M., Chen, P. H., Tilson, H. A., and Hong, J. S. (1985). Effects of *p,p'*-DDT on the rat brain concentrations of biogenic amine and amino acid neurotransmitters and their association with *p,p'*-DDT-induced tremor and hyperthermia. *J. Neurochem.* **45**, 1349–1355.
- Irwin, S. (1968). Comprehensive observational assessment. Ia. A systematic quantitative procedure for assessing the behavioral and physiologic state of the mouse. *Psychopharmacologia* **13**, 222–257.
- Kulig, B. M. (1994). Acrylamide. In *Handbook of Clinical Neurology*, Vol. 20 (64), *Intoxications of the Nervous System, Part I* (F. A. deWolff, Ed.), pp. 63–80. Elsevier Science BV, Amsterdam.
- McDaniel, K. L., and Moser, V. C. (1993). Utility of a neurobehavioral screening battery for differentiating the effects of two pyrethroids, permethrin and cypermethrin. *Neurotoxicol. Teratol.* **15**, 71–83.
- Moser, V. C. (1989). Screening approaches to neurotoxicity: A functional observational battery. *J. Am. Coll. Toxicol.* **8**, 85–94.
- Moser, V. C., Cheek, B. M., and MacPhail, R. C. (1995). A multidisciplinary approach to toxicological screening. III. Neurobehavioral toxicity. *J. Toxicol. Environ. Health* **45**, 173–210.
- Moser, V. C., and MacPhail, R. C. (1992). International validation of a neurobehavioral screening battery: The IPCS/WHO collaborative study. *Toxicol. Lett.* **64/65**, 217–223.
- Moser, V. C., McCormick, J. P., Creason, J. P., and MacPhail, R. C. (1988). Comparison of chlordimeform and carbaryl using a functional observational battery. *Fundam. Appl. Toxicol.* **11**, 189–206.
- Murphy, S. D. (1980). Pesticides. In *Casarett and Doull's Toxicology: The Basic Science of Poisons* (J. Doull, C. Klaassen, and M. Amdur, Eds.), pp. 357–408. MacMillan, New York.
- National Research Council (1991). *Animals as Sentinels of Environmental Health Hazards*. National Academy Press, Washington, DC.
- National Research Council (1992). *Environmental Neurotoxicology*. National Academy Press, Washington, DC.
- O'Callaghan, J. P. (1988). Neurotypic and gliotypic proteins as biochemical markers of neurotoxicity. *Neurotoxicol. Teratol.* **10**, 445–452.
- O'Donoghue, J. L. (1989). Screening for neurotoxicity using a neurologically based examination and neuropathology. *J. Am. Coll. Toxicol.* **8**, 97–115.
- Office of Technology Assessment (1990). *Neurotoxicity: Identifying and Controlling Poisons of the Nervous System*. U.S. Government Printing Office, Washington, DC.
- Pryor, G. T., Uyeno, E. T., Tilson, H. A., and Mitchell, C. L. (1983). Assessment of chemicals using a battery of neurobehavioral tests: A comparative study. *Neurobehav. Toxicol. Teratol.* **5**, 91–117.
- Reiter, L. W., and Ruppert, P. H. (1984). Behavioral toxicity of trialkyltin compounds: A review. *Neurotoxicology* **5**, 177–186.
- Reiter, L. W., and MacPhail, R. C. (1979). Motor activity: A survey of methods with potential use in toxicity testing. *Neurobehav. Toxicol.* **1** (Suppl. 1), 53–66.
- Schaumburg, H. H., and Spencer, P. S. (1980). Selected outbreaks of neurotoxic disease. In *Experimental and Clinical Neurotoxicology* (P. S. Spencer and H. H. Schaumburg, Eds.), pp. 883–889. Williams and Wilkins, Baltimore, MD.
- Sette, W. F. (1989). Adoption of new guidelines and data requirements for more extensive neurotoxicity testing under FIFRA. *Toxicol. Indust. Health* **5**, 181–194.
- Taylor, P. (1985). Anticholinesterase agents. In *Goodman and Gilman's: The Pharmacological Basis of Therapeutics* (A. G. Gilman, L. S. Goodman, T. W. Rall, and F. Murad, Eds.), 7th ed., pp. 110–129. MacMillan, New York.
- Tilson, H. A. (1981). The neurotoxicity of acrylamide: An overview. *Neurobehav. Toxicol. Teratol.* **3**, 445–461.
- Tilson, H. A., and Moser, V. C. (1992). Comparison of screening approaches. *Neurotoxicology* **13**, 1–14.
- U.S. Environmental Protection Agency (1991). *OPPTS Test Guidelines: Neurotoxicity (PB91-154617)*. National Technical Information Service, Springfield, VA.
- van den Heuvel, M. J., Dayan, A. D., and Shillaker, R. O. (1990). Evaluation of the BTS approach to the testing of substances and preparations for their acute toxicity. *Food Chem. Toxicol.* **28**, 469–482.
- Watanabe, I. (1980). Organotin (Triethyltin). In *Experimental and Clinical Neurotoxicology* (P. S. Spencer and H. H. Schaumburg, Eds.), pp. 545–557. Williams and Wilkins, Baltimore, MD.
- Woolley, D. E. (1982). Neurotoxicity of DDT and possible mechanisms of action. In *Mechanisms of Actions of Neurotoxic Substances* (K. N. Prasad and A. Vernadakis, Eds.), pp. 95–138. Raven Press, New York.
- World Health Organization (1986). *Principles and Methods for the Assessment of Neurotoxicity Associated with Exposure to Chemicals*. Environmental Health Criteria 60, Geneva.