The Intersection of Risk Assessment and Neurobehavioral Toxicity

Bernard Weiss and Jürg Elsner

--- Environ Health Perspect 104(Suppl 2):173–177 (1996)

After nearly three decades of research in many parts of the world, neurobehavioral toxicity is now acknowledged as a significant outcome of chemical exposure. In contrast to the view prevailing even in the recent past, many observers now concede that its health and economic costs may exceed even those of cancer, the prototype for risk assessment, by substantial amounts. This new perspective has been accompanied by a surge of efforts designed to promote effective test methods, to explore the responsible mechanisms, to design applicable risk assessment procedures, and to determine the consequent policy implications (1,2).

The process of recognition did not proceed as smoothly as expected, given the resonant scientific foundations provided by the behavioral neurosciences. One of these, behavioral pharmacology, the discipline that emerged in the 1950s in response to the introduction of chemotherapy for psychological disorders, provided a readily adaptable technology for exploring adverse effects. Workplace exposure criteria, such as threshold limit values (TLVs), had long relied on behavioral criteria such as work efficiency and alertness to danger to infer hazard. Perhaps the problem lay in how easily misunderstandings can arise about the definition and measurement of behavior.

Although the discipline has generated an abundant literature and established a robust scientific footing, translating such efforts into policy decisions remains perplexing, mainly because of the difficulties posed by how to express them in risk terms. The conventional prototype for risk assessment is cancer, but numerous dissimilarities between neurobehavioral toxicity and carcinogenesis render it a rather imperfect model. Because behavior is often cited as the integrated product of a highly complex system, with numerous modes of expression, it should come as no surprise that it may be altered in equally diverse ways by xenobiotic influences and that the significance of any but the most blatant behavioral change eludes simplistic measures and interpretation.

After all, behavior is a dynamic and plastic phenomenon. It would be deceptive to compare it to functions that are much more rigid and deterministic such as those of the cardiovascular system. Scientists unaccustomed to phenomena as malleable as behavior sometimes find it difficult to grasp both its essential lawfulness and the degree to which, concurrently, it may undergo critical modifications without displaying any overt abnormalities. Some consider behavioral changes to be analogous to alterations in software which, by proper reprogramming, may be overcome without major difficulties. Others may claim that behavioral deficiencies attributed, for example, to elevated exposure to metals, are more likely the product of deficiencies in social conditions. Such claims tend to erode when confronted jointly by data from properly conducted animal research and from epidemiological studies that deliberately and carefully weigh and balance the influence of potentially confounding social variables. Several of the joint chapters and individual papers review these issues.

A broad, permeating issue derives from one of the original aims of SGOMSEC: to make its contributions pertinent to countries lacking advanced industrial economies and resources. Chemicals and chemical production facilities tend to be transferred to such countries without an accompanying transfer of the technology of toxicology and environmental health science. This discrepancy results in unsafe control practices, excessive exposure levels, and, ultimately, mass chemical disasters. SGOMSEC 11 strove to confront this issue by describing a range of methods from the relatively simple to the rather complex and by illustrating the different contexts in which different methods are appropriate. But even in advanced industrial societies, policy analysts, regulators, and others with decision-making responsibilities are confronted with irksome questions about neurobehavioral toxicity. In that arena, the challenges range from how to determine whether the potential for neurotoxicity exists to how to translate such potential into policy.

SGOMSEC 11 was also designed to learn from the history of neurobehavioral toxicology. It sometimes proved difficult to convince toxicologists from other specialties and policy makers that even substances already dispersed in the human environment require careful evaluation of their neurobehavioral toxicity, despite no cogent evidence of adverse effects at environmental levels. Once a substance is widely distributed in the communal, or even the industrial environment, barriers to its removal are riveted in place. Especially if the arguments for its control are based, not on immediate threats to life but on a less tangible behavioral metric, inertia exerts a potent force. The arguments for premarket testing for neurobehavioral toxicity flow from such experiences.

The Choice of a Focus on Behavior

The adjective neurobehavioral is commonly applied because the nervous system determines the contours of its ultimate product, behavior. Any measure of nervous system status or function incurs immense complexities. Behavior’s credentials as a valid toxicity index are often questioned because its determinants converge from many paths. The consequences of a specific neurochemical aberration such as a shift in receptor density, for example, may be expressed behaviorally in almost limitless ways depending on the specific end points and indices chosen for measurement and the constitutional capacities and behavioral history of the individual organism. Consider the numerous behaviors linked to the neurotransmitter dopamine: a variety of cognitive functions, mediation of
reinforcement processes, tremor and other indices of motor function, sexual performance and motivation, and even species-specific behaviors. Naturally, the most appealing situation is one in which neurochemical findings could be correlated with behavioral data, but most behaviors are joined to more than one neurotransmitter system and embrace more than a single brain structure. Such multiple connections explain why neurochemistry, morphology, and even electrophysiology would normally be introduced only at the later stages of assessment.

Because it arises from multiple sources, behavior might be viewed as a confusing index of toxicity. That potential for confusion, however, is also an argument in its favor. If it is subject to such a wide array of influences, the argument goes, it can then serve as an apical tool for testing general toxicity. If such evidence emerges, more specific behavioral or other measures can be applied to narrow the contributing variables or mechanisms. The opposing argument claims that, because behavior reflects the integration of a highly redundant system in which compensatory mechanisms may obscure a deficit in any particular functional domain, it is not a sensitive measure of adverse effects in all circumstances.

Both arguments, despite their apparently conflicting stances, invoke equivalent conclusions: toxic potential should be assessed by choosing behavioral endpoints that offer the greatest breadth and precision of information. It should be recognized that the appeal of simplicity and economy may prove deceptive and even costly if they merely manipulate the intrinsic ambiguities of risk assessment. SGOMSEC 11 aimed to deal explicitly with such supramethodological issues while offering critical reviews of the prevailing approaches.

The final design of SGOMSEC 11 divided the issues into four sections: neurobehavioral toxicity in humans, neurobehavioral toxicity in animals, model agents, and risk assessment. Anyone familiar with the discipline appreciates that these rubrics do not describe fixed boundaries, but convenient classifications. In fact, the extensive overlap between these categories proved to be an advantage because members of one group could be enlisted, in preparing the joint report, to assist another group when their special qualifications were required.

The outline below provides a list of topics for which individual papers were commissioned. Each of the participants was asked to feature three points: How did we get to the current status of the topic? How can we relate it to risk assessment? What methodological advances should we seek to make a firmer connection with policy?

Identification of Neurobehavioral Toxicity in Human Populations

This section was designed to explicate the ways in which information about hazard and risk might be procured from human populations. In some past instances, this information came from clinical observations, usually on the basis of extreme exposure levels. The current mode of defining risks depends mostly on the use of psychological test instruments, but questions remain about their relevance and suitability.

Clinical Data as a Basis for Hazard Identification

Many of the neurobehavioral toxicants now viewed as hazardous to humans originally earned recognition through the observations of clinicians. These toxicants came to their attention because of signs and symptoms overtly expressed by patients. What are the lessons to be learned from this history? What tools should clinicians be prepared to deploy in such instances? Is hazard identification the only role fulfilled by clinical observations? Is there a series of steps, undertaken in a clinical context, that might lead to a firmer basis for identifying and estimating risk once such observations are validated? How can clinical observations be translated efficiently into epidemiological studies? Can a useful guide be designed for doing so? Is a tiered strategy, that is, one that builds systematically from one set of observations to another more complex set the most appropriate one to adopt, or does such staging of questions tend to delay the risk assessment process? Are there useful examples of such a progression?

Designing and Validating Test Batteries

Beginning in about the early 1970s, psychological test batteries began to be applied to the definition and assessment of adverse consequences stemming from exposure to central nervous system-active agents such as volatile organic solvents. By now, a plethora of test collections has penetrated the literature. Although these batteries possess many elements in common, they also diverge in philosophy and design. What are the strengths and weaknesses of the present array of batteries? How might they be improved while maintaining their advantages of ease of use and broad acceptance? Would they still be suitable for critical applications in less advanced countries? What about their suitability for longitudinal assessments? How well do they evaluate sensory and motor function?

The most widely adopted batteries are anchored in diagnosis. Their roots lie in neuropsychology and the assessment of brain damage and psychopathology. Should other approaches be considered? Test batteries are generally constructed to use brief samples of behavior to screen for adverse effects in populations such as workers. Is the breadth of test items in the typical battery a problem? What are the advantages and disadvantages of adopting a more intense focus? This approach might be used for pilot and astronaut selection or to represent translations from complex performance in animals. Do such approaches hold any lessons for the evaluation of neurobehavioral toxicity?

Translation of Symptoms into Test Variables

A problem now looming for neuropsychology and neurobehavioral toxicology is the collection of quasi-clinical, often vaguely defined syndromes labeled as Multiple Chemical Sensitivity, Sick Building Syndrome, and Chronic Fatigue Syndrome. All are reflections of patient complaints lacking consistent objective verification such as that provided, say, by clinical chemistry profiles. As a result, many clinicians and biomedical scientists tend to view such complaints skeptically, or find themselves unable to propose any course of action. Does part of the problem arise from the emphasis by neuropsychology on diagnosis rather than on functional variables or on labeling of deficits rather than on determinations of how effectively the individual functions in his or her environment? How can such data be collected or synthesized or estimated? Are there especially suitable experimental designs for such questions, such as single-subject designs? What alternatives to current assessment procedures hold promise? Would they be suitable for longitudinal evaluations such as those that might become necessary for monitoring the aftermath of a poisoning episode?

Developmental Neurotoxicity

The period of early brain development is a precarious stage because insults inflicted during this time seem to ramify in many directions, often first becoming perceptible only after reaching a particular epoch of the life cycle. As a consequence, a full evaluation
of the neurotoxic impact of prenatal and neonatal exposure virtually demands longitudinal investigations for definitive answers. What are the essential elements of such designs? Which features are absolutely indispensable? What are the most serious confounders? Can these kinds of studies somehow be streamlined? To what extent can cross-sectional studies serve as surrogates or, at least, pointers? Can the procedures of animal testing, which do not require the same degree of compensation for social and cultural variables, be adapted for testing children?

Identification and Confirmation of Neurobehavioral Toxicity in Animals

Evaluations in laboratory animals fulfill two purposes. First, for new chemicals, these evaluations should make it possible to determine whether an agent presents a significant hazard. They also allow exploration of the potential dimensions of the hazard. Finally, they may make it possible to distill quantitative risk estimates for humans, in parallel with the way in which biosy data are used in cancer risk assessment. Tumors, however, are presumed to reflect processes that will occur in human hosts. Neurobehavioral deficits in animals are less directly translatable into human functions. What should be the role of animal research and in what ways can it serve the ultimate purpose of risk assessment?

Many critics attack the validity of extrapolating behavioral data from animals to humans. Indeed, behavior seems to be highly species-specific and exquisitely adapted to the organism’s and its species survival needs. Although such critics grant the universality of the genetic code, they are less willing to grant the universality of the neural mechanisms governing the operation of nervous systems in different species. In this framework, humans are viewed as beyond extrapolation, with human behavior accorded the status of some emergent phenomenon disconnected from the brain structures they share with other species.

No one denies that the structural differences between rodent and human brains and the differences in behavioral repertoires vitiate any facile and superficial extrapolations. But the underlying functional mechanisms of the brain, and their expression in behavior, are shared by these organisms. Rat behavior can be used as a model of human behavior if a model is defined as a system possessing essentially the same functional properties as the one it simulates, except in a simplified version. Deficits in human behavior ascribed to neurotoxicants tend to manifest themselves in fundamental functional properties shared with other species. Labels such as attention, emotional responsivity, sensory processing, motor coordination, learning disabilities, and others are not specifically human properties of behavior. Human language is distinctive, of course, but its acquisition displays a pattern common to many other behaviors that follow a developmental sequence in which environmental and constitutional variables merge continuously. The primary source of confidence in the power of extrapolation though is a body of findings that supports the congruence of human and animal responses to neurotoxicants.

Natural Populations as Sentinels

Safety evaluation of environmental chemicals has been broadened to include ecological risk assessment. The U.S. EPA’s Science Advisory Board report, Reducing Risk (3), is one instance of this growing appreciation, but the impact of chemical pollution on natural populations rose to a subject of widespread concern after Rachel Carson’s seminal book (4). We now acknowledge that a major element in this impact derives from disruptions in behavior; one example is a reported diminution of nest attentiveness by birds in the Great Lakes. What are the indicators that up to now have proven useful in natural populations? In which directions should improvement in these methods be pointed? What is the extent of concordance between such observations and human health effects or with laboratory animal studies? How can ecological observations be converted into the kinds of quantified variables characteristic of laboratory experiments without losing essential information?

Laboratory Approaches: Scope and Selection of End Points

For new chemicals, laboratory assays provide the first filtering stage for potential toxicity. Currently, a standardized set of observations, such as a functional observation battery (FOB), is used to probe for neurobehavioral effects. Certain regulatory bodies have also required measures of motor activity, perhaps accompanied by neuropathology at this stage. These criteria are acknowledged as broadly suggestive rather than as definitive, especially at the point when dose–response modeling enters the risk assessment process.

For many purposes, the clinical examination, as in humans, will represent the first initiative, and often the first clues that a neurotoxic agent has appeared on the scene. Can a standardized protocol be designed that will prove feasible, in settings lacking other resources, and sensitive as well? How should such a protocol be modified for examinations in the field, as for wild animal populations?

If a more comprehensive evaluation is sought, what should be its constituents? What considerations should guide the selection of experimental parameters? What research should be conducted to help refine such a process? What constraints are imposed by the extrapolation issue? How vital is it to assure that observations in animals reflect analogous functions in humans? Is it more important to select end points that reflect the functional capacities of the particular species?

What economies of approach are feasible when resources are limited? Does the strategy of tiering, in which assessments branch to increasingly specific and complex assessments, make sense in such situations? How might low cost and sensitivity be combined? What should be the priorities in such a process?

Developmental Neurotoxicants

Exposure to chemicals during early development often inflicts toxic consequences rather different from the consequences inflicted on mature nervous systems. In addition to the modes of damage, however, differences arise in how the damage may be expressed. For example, it may emerge only after a prolonged latency, perhaps as late as senescence. Or, it may appear in different guises at different phases of the life cycle. U.S. EPA and other regulatory bodies have prescribed standardized protocols for assessing developmental neurotoxicity. Do these protocols offer support for a comprehensive, quantitative risk assessment? If not, how should they be modified? Are they efficiently designed and are some elements of these protocols possibly redundant? For example, does the absence of functional impairment at a particular exposure level preclude morphological aberrations at that level? Or must all potential sources of information be examined?

Model Agents

The agents discussed in this section offer cogent history lessons. Organic solvents and chlorinated hydrocarbons were widely used for many years without much concern
over their possibly adverse effects. By the time these properties had been identified in a painfully slow process, the agents had already pervaded the environment or had become so essential that their removal, even if technically possible, became impractical. Methylmercury and lead had been recognized as neurotoxicants long before their current prominence, but an appreciation of their more abstruse expression at low exposure levels required an abundance of resources and investigator dedication in the face of sometimes monumental skepticism.

Current neurobehavioral toxicology largely owes its standing to these agents because they exemplify the power of behavioral end points. We asked the participants to review what we have learned from investigations of agents now established as prototypes. For example, would a retrospective analysis of the literature built around such model agents provide guidance for how to approach new agents? What would have been the most appropriate testing schemes and toxic end points and which assessment strategy would have yielded maximum information at the least cost?

Those enumerated below all owe their original identification as neurobehavioral toxicants to observations in humans, typically at high doses. What might have been the outcome had these agents first been examined as new chemicals? Which endpoints would have proven to be sensitive? To what degree, for each agent, have we observed a convergence between progress in human and animal research?

Lead

Lead was recognized as a hazard even in antiquity but was frequently ignored. Only with the accumulating, incremental evidence provided by methodological refinements did we progress to the present situation. The current Centers for Disease Control (CDC) guidelines denote blood levels above 10 µg/dl as a potential index of excessive exposure—a sharp fall from the standards prevailing only a short time ago. Animal and human data show periods both of convergence and divergence but, on the whole, took parallel paths. Attaining convergence, the current situation, required improvements in both sets of methodologies, but the animal data proved critical because of the criticisms aimed at the epidemiological studies. In essence, investigators learned how to ask the appropriate questions. It was not a process that would have succeeded without the inevitable but instructive blunders.

Methylmercury

Not long ago, methylmercury was viewed only as a hazardous chemical confined to narrow purposes and distribution. A chain of mass chemical disasters gradually altered this view, but the extrapolation from mass disasters to broad implications for public health came slowly. On the basis of knowledge acquired from these disasters, 26 states in the United States have posted fish advisories. Animal research contributed significantly to our understanding of the underlying mechanisms of toxicity, but the risk issues are still being played out, primarily with the human disaster data. How has animal research illuminated the human risk perspective? What has it taught about the approach to untested chemicals? What lessons should be drawn about the longitudinal monitoring of human populations? Do the animal data allow reasonable dose extrapolation?

Organochlorine Pesticides and Related Compounds

Compounds ranging from dichlorodiphenyltrichloroethane (DDT) to 2,4-dichlorophenoxyacetic acid (2,4-D) to the polychlorinated biphenyls (PCBs) to 2,3,7,8-tetrachlorobiphenyl-p-dioxin (TCDD) have been implicated in neurotoxicity. Especially for the last two classes of chemicals, recognition of their potential neurotoxic properties emerged only gradually, perhaps because it was submerged by concerns about carcinogenicity. What is the current perspective about the health risks of these compounds, and what lessons does its evolution provide for how other classes of chemicals should be examined? Such substances are also now implicated as environmental estrogens with a new spectrum of neurobehavioral issues to address, some of which may even be lurking in data we already possess.

Solvents

Volatile organic solvents became an early focus of human neurobehavioral toxicology. Their neurotoxic properties have always been recognized, even in setting exposure standards in the workplace. Wider recognition of these properties, especially in the absence of gross dysfunction, is attributable to the application of psychological testing methods. Because methodological advances moved in parallel with improvements in study design, the solvents literature has provided guidance for similar questions. The evolution of this research area to its current state should offer lessons on how to cope with related issues such as those stemming from chemical sensitivity syndromes. As with lead, animal models came on the scene only after solvent neurotoxicity had been well established. The same degree of parallelism seen with lead has yet to be achieved and awaits the application of equally sensitive behavioral criteria.

Quantification, Modeling, and Definition of Risk

The ultimate goal of neurobehavioral toxicology, apart from its inherent contributions to basic science, is formulating risk. Although, by tradition, toxicity data are transformed into values such as NOAELs, this is simply a regulatory convenience rather than a risk assessment. The conversion of neurobehavioral data into quantitative risk assessments presents numerous challenges. Cancer risk assessment, the prototype, is based on premises that cannot be applied to neurobehavioral toxicity. Among these are the assumption of a unitary biological process, cumulative dose as a valid exposure parameter, and the irrelevance of acute animal toxicity data for the prediction of carcinogenic potential.

Translation of Neurobehavioral Data into Risk Figures

Another legacy of the cancer risk model is its dependence on quantal data. Such measures are easier to handle for risks expressed in probabilistic terms, but most neurobehavioral measures are continuous rather than discrete. One result of this disparity is that risk for systemic outcomes is typically framed in terms such as NOAELs. Furthermore, many effects are graded over time, so that they present features best expressed, perhaps, as 3-dimensional surfaces. What are possible models for expressing risks based on such graded outcome measures? Do they hold implications for experimental design such as choices between number of dose levels and number of observations at each dose? How should they reflect repeated measures on the same subjects? Are there examples from the currently available literature?

Choosing End Points

Unlike the model of cancer, neurobehavioral toxicology is compelled to rely on several different types of measures as guidance for risk estimates. For example, the U.S. EPA has requested data from a FOB, motor activity, schedule-controlled operant behavior, and neuropathology to help it formulate the health risks of exposure to volatile organic solvents. Even pathology,
which in the past constituted the primary criterion of toxicity, is inadequate by itself. Furthermore, even a single criterion, such as schedule-controlled operant behavior, itself comprises multiple measures. Which measures derived from such techniques might be suitable for guiding the risk assessment process?

A unique assortment of questions is posed by developmental neurotoxicology because the process of development itself offers inherent enigmas. Species extrapolation in this context, despite fundamental commonalities among species, poses an additional layer of uncertainty upon those already confronting risk evaluations based on species comparisons. Is the prevailing strategy adequate for even gross prediction or do its deficiencies herald further errors or even disasters?

**Neurobehavioral Epidemiology**

How do neurobehavioral end points coincide with the requirements of epidemiology? Rather than cases, for example, the data may consist of dose-effect relationships in which the effect may be expressed as alterations in a spectrum of deficits, or, because of individual patterns of susceptibility, individuals may differ in their relative responsiveness to different end points. What would be an appropriate epidemiological framework for assessing neurobehavioral toxicity?

**Setting Exposure Standards: A Decision Process**

Most observers recognize that, barring rejection of an agent at the earliest stage of risk assessment, a broad but necessarily superficial appraisal of potential neurobehavioral toxicity may be insufficient for quantitative risk assessment or even for identifying critical end points that are not easily appraised with simpler techniques. Under what conditions should a superficial appraisal be relied upon to formulate risk? Assume that further investigations beyond the simplest may have to be conducted. Can a cogent design for a sequential strategy be formulated? What are satisfactory starting and stopping points? One model of a quasi-tiered approach is the assessment of developmental neurotoxicity, a model imposed simply by the inability to reach definitive conclusions about the impact of exposure at one particular age from results determined at another age. What should be the major decision points in evaluations not aimed at developmental questions or in evaluations of developmental toxicity? Is it more efficient to begin with the later decision points than to proceed, say, from simple to complex in several stages? That is, would the later decision points embody, as well, the earlier ones? Are there decision rules that can be constructed to guide such a process? Can decision nodes be established at which certain paths can be taken for more definitive conclusions?

Tiered testing schemes generally proceed from simple to complex criteria. This direction generally implies corresponding dimensions such as from cheap to expensive, from crude to sensitive, from high-dose to low-dose effects, from acute to chronic effects, from adult exposure to developmental toxicity, from hazard identification to quantitative risk assessment. Such progressions reveal where the problem lies in a tiered testing approach: If merely the absence of toxicity in tier 1 procedures is legally required for approval of substances that may invade the environment and expose humans and animals, new substances will be tested by relatively simple and insensitive tests following acute high-dose administration in adult animals. Would such a strategy be adequate to offer protection against the recurrence of situations such as those described under Model Systems? Will more scientific battles have to be fought in 10 years to prompt an assessment of the neurobehavioral toxicity of substances introduced today?

**Summary**

Neurobehavioral toxicology is now established as a core discipline of the environmental health sciences. Despite its recognized scientific prowess, stemming from its deep roots in psychology and neuroscience and its acknowledged successes, it faces additional demands and challenges. The latter, in fact, are a product of its achievements because success at one level leads to new and higher expectations. Now the discipline is counted upon to provide more definitive and extensive risk assessments than in the past. These new demands are the basis for the appraisals presented in the SGOMSEC 11 workshop. They extend beyond what would be offered in a primer of methodology. Instead, these appraisals are framed as issues into which what are usually construed as methodologies have been embedded.

**REFERENCES**

1. National Research Council, Committee on Neurotoxicology and Models for Assessing Risk. Environmental Neurotoxicology. Washington: National Academy Press, 1992.
2. Office of Technology Assessment, U.S. Congress. Neurotoxicity: Identifying and Controlling Poisons of the Nervous System. New York: Van Nostrand Reinhold, 1990.
3. U.S. EPA. Reducing Risk: Setting Priorities and Strategies for Environmental Protection. Rpt SAB-EC-90-021. Washington: U.S. Environmental Protection Agency, 1990.
4. Carson, R. Silent Spring. Boston: Houghton Mifflin, 1962.