Setting Exposure Standards: A Decision Process

Hugh A. Tilson, Robert C. MacPhail, and Kevin M. Crofton

Neurotoxicology Division, National Health and Environmental Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, North Carolina

Increased emphasis on routine screening of chemicals for potential neurotoxicity has resulted in the development of testing guidelines and standardized procedures. A multiphased, tiered-testing strategy has been proposed by numerous expert panels to evaluate large numbers of chemicals. In a regulatory context, however, a formal tiered-testing approach is not used, mostly because of the constraints of differing regulatory authorities and the potential cost of such a testing strategy. Instead, current regulatory decision making utilizes all available animal and human data to identify a critical adverse effect which is then used for setting standards. Although the current decision-making process does not use a formal tiered-testing approach, it appears to identify chemicals with neurotoxic effects. An analysis of U.S. Environmental Protection Agency integrated risk information system (IRIS) indicates that about 20% of the chemicals having standards or health advisories are based on neurotoxicity. — Environ Health Perspect 104(Suppl 2):401–405 (1996)

Key words: neurotoxicity, screening, tiered testing, risk assessment, IRIS

Introduction

There are between 50,000 and 100,000 chemicals used commercially and 1,000 to 1,600 new chemicals are submitted for premanufacture notification each year in the United States alone (*1*). Several regulatory agencies in the United States are charged with enforcing legislation designed to protect the public and environment from the hazards of chemical exposure (*2*). Premarket testing, for example, is required for pesticides by the U.S. Environmental Protection Agency (U.S. EPA) under the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA) and for drugs and food additives by the Food and Drug Administration (FDA). The Consumer Product Safety Commission (CPSC) and U.S. EPA can require testing of consumer products and industrial chemicals if there is a demonstrated justification for testing. The U.S. EPA can also require evaluation of industrial chemicals under the Toxic Substances Control Act (TSCA).

It is now recognized that many chemicals can have an adverse effect on the function and/or structure of the nervous system. Anger and Johnson (*3*), for example, identified more than 750 industrial chemicals as having neurotoxic effects following acute or repeated exposure. From the list of 750 chemicals, Anger (*4*) found 65 chemicals for which it has been estimated that 1 million or more people are potentially exposed. Of 588 chemicals listed by the American Conference of Government and Industrial Hygienists, 167 have threshold limit values based, in part, on neurologic or behavioral end points (*5*). Neurotoxic disorders, hearing loss, and psychological dysfunction are among the 10 leading occupational problems in the workplace (*6*). A recent publication from the Office of Technology Assessment estimated that 3 to 28% of all chemicals may be neurotoxic (*7*).

To identify agents producing neurotoxic effects, considerable efforts have been made to develop valid, sensitive measures of neurotoxicity. Many of the tests currently used in neurotoxicological studies are functional indicators of neurotoxicity, especially behavioral end points, that assess chemical-induced changes in sensory, motor and cognitive functions. Since behavior appears to be the net result of the integrated output of sensory, motor and cognitive processes of the nervous system, chemical-induced changes in behavior may be a relatively sensitive indicator of dysfunction in the nervous system (*8*).

Reliance on behavioral measures in neurotoxicological testing is reflected in the testing guidelines published by regulatory agencies such as the U.S. EPA. For several years, the U.S. EPA has had guidelines for a functional observational battery (FOB), motor activity, and schedule-controlled behavior. A neurotoxicity screening battery (*9*) consisting of motor activity, FOB, and neuropathology is now required for the registration and reregistration of pesticides. The U.S. EPA has also published testing guidelines for developmental neurotoxicity that include motor activity, learning/memory, and acoustic startle reactivity. The U.S. EPA also relies on behavioral end points in screening chemicals that produce organophosphate-induced delayed neuropathology. In the assessment of human therapeutic agents, the FDA relies on a battery of animal toxicity data that includes clinical observations and neuroanatomical assessments. Neurotoxicity, including adverse effects on behavior, is of concern to other regulatory agencies, including the National Institute for Occupational Safety and Health, Occupational Safety and Health Administration, and the CPSC.

Countries other than the United States also require behavioral tests in animal toxicity studies. As early as 1975, Great Britain and Japan began requiring behavioral testing in animals to evaluate the developmental neurotoxicity of new drugs (*10*). In adopted testing protocols similar to those required by Great Britain and Japan. The World Health Organization has also published proposed

This paper was prepared as background for the Workshop on Risk Assessment Methodology for Neurobehavioral Toxicity convened by the Scientific Group on Methodologies for the Safety Evaluation of Chemicals (SGOMSEC) held 12–17 June 1994 in Rochester, New York. Manuscript received 1 February 1995; manuscript accepted 17 December 1995.

Address correspondence to Dr. Hugh A. Tilson, Neurotoxicology Division (MD-74B), Health Effects Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711. Telephone: (919) 541-2671. Fax: (919) 541-4849. E-mail: tilson@herl45.herl.epa.gov

Abbreviations used: U.S. EPA, U.S. Environmental Protection Agency; FIFRA, Federal Insecticide, Fungicide, and Rodenticide Act; FOB, functional observational battery; NRC, National Research Council; LOAEL, lowest observed adverse effect level; NOAEL, no observed adverse effect level; UF, uncertainty factor; RfD, reference dose; IRIS, integrated risk information system; HAs, health advisories; RfCs, inhalation reference concentrations; OTA, Office of Technology Assessment; WHO/IPCS: World Health Organization/International Programme for Chemical Safety.

testing guidelines for drugs and other agents. Finally, the Organization for Economic Cooperation and Development is now considering a number of neurotoxicity testing protocols, many of which involve behavioral assessments.

Strategies for Assessment of Chemicals for Neurotoxicity

Since 1975, a number of expert panels have recommended a tiered-testing strategy for the evaluation of chemicals for neurotoxicity, and all have recommended the use of behavioral techniques (*11,12*) (Table 1). Tiered testing involves two or more phases of evaluation in which each stage incorporates decision points as to whether available information is sufficient for determining the neurotoxicity of a chemical.

In 1978, Evans and Weiss (*13*) described a multiphased strategy of assessment based on a scheme published by the National Academy of Sciences in 1975 (*8*) (Figure 1). In cases where little toxicological information exists, there is a need to first determine whether a chemical is capable of producing neurotoxicity, i.e., hazard identification. Frequently, an LD_{50} (or equivalent) and a chemical's structure are all that are available in the early stages of hazard identification. In such cases, assessments at the first tier may be required or a decision must be made to develop a toxicological profile before subsequent product development can proceed.

First-tier tests are typified by the capability to assess large numbers of animals, they usually require little or no training of test animals before exposure, and they are relatively simple to perform. Such tests, however, may be labor intensive, subjective, and at best semiquantitative. Examples of first-tier tests include neurological screens such as the FOB and cage-side observations. Motor activity is another example of a first-tier test that is, in contrast to others, objective and quantitative (*14*). Behavioral tests are sometimes also used with other routine measures such as neuropathology or neurochemical assays. If a chemical is found to be neurotoxic in the first tier, a decision would have to be made to test at the next tier. Such a decision might be based on the beneficial properties of the chemical relative to the neurotoxic effects observed in the screening battery. Positive findings in the first tier, then, could either suspend further development of a chemical or facilitate the design of experiments at the next tier.

Studies at the second tier are designed to characterize the nature of a chemical's neurotoxicity. A decision to test at the second tier might be based on results from the first tier, on published data, or on new toxicological data indicating that a chemical already in the environment or workplace produces neurotoxicity. Second-tier tests are generally thought to be more sensitive than those at the first tier and can require special automated equipment and extensive training of the animals. Examples of second-tier tests include procedures to measure chemical-induced alterations in learning and memory or signal detection techniques to measure sensory dysfunction.

Evans and Weiss (*13*) also indicated a third tier of testing involving assessment of human susceptibility to environmentally relevant chemicals using procedures studied in animals. Fine motor disturbances produced by exposure to methylmercury, for example, could be studied in humans using essentially the same methodology used in laboratory animals.

In a recent report published by the National Research Council (*15*), tiered testing was also discussed as a strategy for the assessment of chemicals. The tiered testing approach is similar to that described by Evans and Weiss (*13*) for the first two tiers. The NRC (*15*) report, however, indicated that mechanistic studies focusing on how a chemical produced a neurotoxic effect would be conducted at the third tier (Table 2). In this case, a detailed examination would be undertaken of a chemical's effect at several levels of nervous system organization (i.e., behavioral, cellular, molecular). Information derived from these studies might be used in the development of biologically based dose–response models. The approach taken by Choi (*16*) linking the neurotoxicity produced by an excitatory amino acid-mediated increase in intracellular calcium is an example of a mechanistic study.

In a recent overview of screening approaches, Tilson and Moser (*17*) found similar tier-testing approaches by a number of laboratories engaged in animal testing. All of the laboratories surveyed used some type of FOB or clinical neurological assessment at the first tier. Specialized measures of motor function, such as strength, splay and activity, and routine neuropathology were included in the first tier by some laboratories. Second-tier evaluations were used to characterize neurotoxicity detected in the first tier and included more specialized behavioral, neuropathological, electrophysiological, or neurochemical assays.

Table 1. Expert panels or committees recommending neurotoxicity testing.

Study	Year	Recommendation
NRC	1975	Tiered testing using behavioral tests
NRC	1977	Conditioned and unconditioned tests for hazard identification
NRC	1984	Behavior and neuropathology as Tier 1 tests
WHO/IPCS	1986	Two tiers of behavioral tests
U.S. EPA/FIFRA	1987	Motor activity, FOB, and neuropathology as Tier 1 test for pesticides

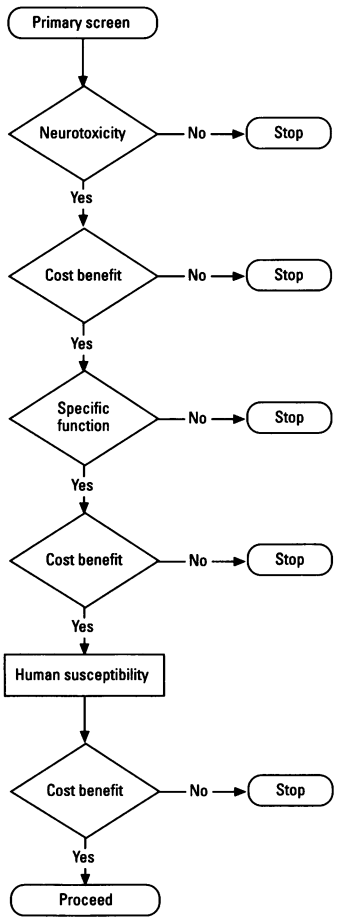

Figure 1. Three-tiered testing scheme recommended by Evans and Weiss (13) including a first tier for hazard identification, a second tier for characterization of specific functions, and a third tier for human susceptibility.

Table 2. Tier-testing approach.

Tier 1	Hazard identification
Tier 2	Characterization
Tier 3	Mechanism of action

Data from the National Research Council (*15*).

Several measures at different levels of neural organization were frequently taken to validate or improve interpretation of the toxicological findings. Since the focus of the review by Tilson and Moser (*17*) was on hazard identification, a third tier of testing to assess human neurotoxicity or conduct mechanistic studies was not discussed.

Tiered Testing in a Decision-making Context

Although a systematic tiered-testing strategy as described by Evans and Weiss (*13*) and the NRC (*15*) seems logical and relatively straightforward in principle, such an approach is not frequently used in a regulatory decision-making context. For example, although a U.S. EPA testing guideline for developmental neurotoxicology has been in place for a number of years, it has been invoked only rarely (*18*). Instead, U.S. EPA has historically chosen to use the developmental neurotoxicity guideline in test rules and negotiated consent agreements where only a single level of testing was outlined (*19*). A positive finding in such studies in a regulatory context would very likely not result in further testing.

One example of a tier-testing approach that has been adopted by the U.S. EPA is in the registration regulations for fuels and fuel additives. Recently, the U.S. EPA published a notice of proposed rule making for fuels and fuel additives registration regulations (*20*). The purpose of the proposed regulation was to establish a requirement for the registration of motor-vehicle fuels and fuel additives as authorized by sections of the amended Clean Air Act. Under the proposed regulations, makers of fuels and fuel additives would be required to conduct tests and submit information regarding effects fuel emissions might have on public health and welfare.

The new proposed registration requirements for fuels and fuel additives have been organized within a three-tiered health effects structure. At the first tier, makers of fuels and fuel additives would be required to conduct a literature search on the health and welfare effects of fuel and fuel additive emissions, characterize the emissions and provide exposure information. Short-term biological testing to screen for specific health effects in animals of whole emissions of fuels or fuel additive mixtures would be performed at the second tier. Neurotoxicological end points include neuropathology and assessment of brain glial fibrillary acid protein, a putative marker of neural injury. At the third tier, U.S. EPA would determine on a case-by-case basis if additional testing were needed based on data submitted from the first two tiers and any other data available. Information developed under the first two tiers would be used to provide the requisite information for registration, subject to subsequent satisfaction of any tier-three requirements determined by the U.S. EPA.

Although a formal tier-testing strategy is not routinely employed in regulatory decision making, it is clear regulatory decisions are based on data obtained from procedures normally defined as being first- or second-tier tests. For example, for some chemicals regulated under TSCA, the only available information is the structure and some estimation of the levels and numbers of people that might be exposed (*21*). A decision to require testing might be based on structure–activity relationships, production levels, and proposed use. Tier-one tests such as an FOB, motor activity, and neuropathology might be used to provide the data in such cases. On the other hand, if there are data that suggest neurotoxicity or if the chemical is already on the market and neurotoxicity has been reported in occupational or environmental settings, then testing might be required; depending on the circumstances, first- and/or second-tier tests might be used to characterize the nature of the neurotoxicity.

The use of data from first- or second-tier testing in a decision-making context is dependent on the process by which regulatory risk assessments are performed. At the U.S. EPA, quantitative risk assessment begins with the definition of a critical adverse effect based on evaluation of all the available human and animal data (Figure 2). Determination of an adverse critical effect is based on data obtained from animal experiments using first- and/or second-tier tests, mechanistic studies, or from human epidemiological, clinical, or case studies, from which a critical study is identified. The critical effect from that study along with supporting data are then used to determine a lowest observed adverse effect level (LOAEL) or no observed adverse effect level (NOAEL) to calculate a reference dose (RfD) for regulatory decision making. The RfD is an estimate of a daily exposure to the human population that is likely to be without appreciable risk over a lifetime (*22*) and has historically been derived by dividing the NOAEL or LOAEL by one or more factors of up to 10 to account for one or more uncertainties in the experimental data. These uncertainty factors (UF) include intraspecies (human) sensitivity, animal-to-human extrapolation, less-than-lifetime exposures, and use of a LOAEL rather than a NOAEL. An additional uncertainty factor for an incomplete dataset has also been used. A modifying factor ranging from 1 to 10 may also be used in the denominator to reflect any uncertainties in the critical study used to establish the critical effect (*22,23*).

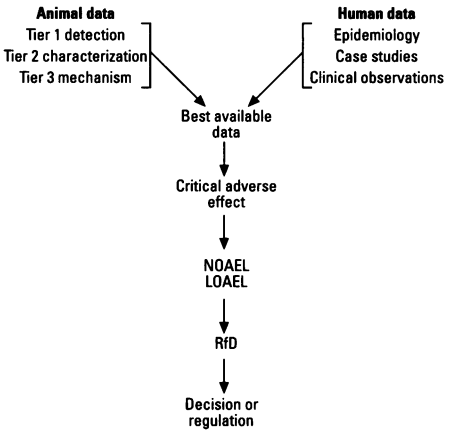

Figure 2. Data from all tiers of animal testing and all human data from epidemiology, case studies, or clinical observations and experimental studies are evaluated and a critical adverse effect is determined to calculate standards for regulatory action.

It is unlikely that the data used to determine the critical adverse effect would be derived from a formal systematic tier-testing strategy. From an industry or pesticide registrant's point of view, the outcome of further testing along with the added time and expense may create disincentives for tiered testing. Consider the hypothetical example in Figure 3. The results of first-tier screening studies suggest a chemical has neurotoxic potential. Second-tier testing is then undertaken to characterize the chemical's neurotoxic effects. If the results of second-tier testing do not indicate neurotoxicity, it may be doubtful whether a prudent risk assessor would summarily dismiss the first-tier results. On the other hand, if the results of second-tier testing indicate neurotoxicity, the NOAEL or LOAEL may be either equal to or lower than that established on the basis of first-tier testing. If it is equal then the RfD would be equal, while if it is lower the RfD would be correspondingly lower. Additional resources would then have been spent either not to change or to lower the regulatory standard.

Since a formal tiered-testing strategy is not currently employed by regulatory agencies such as the U.S. EPA, it is relevant to

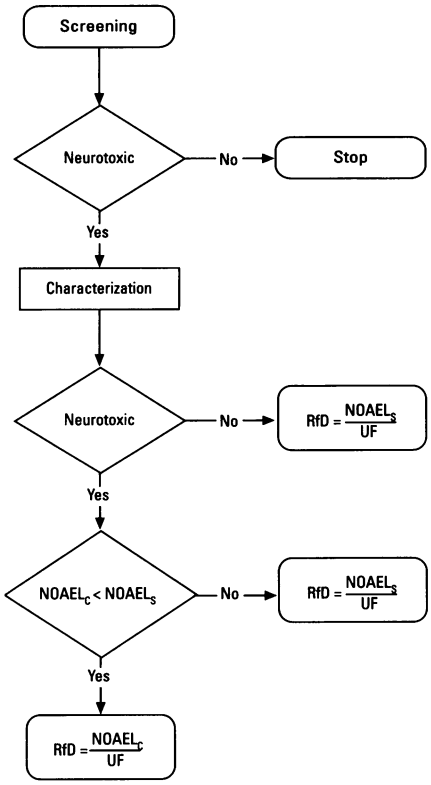

Figure 3. Hypothetical scenario for standard setting based on a tiered testing approach. The conservative nature of standard setting, based upon fundamental uncertainties in predicting human risk, may lead to selection of the lowest NOAEL regardless of the level of testing.

ask about the capability of the current system to detect and set standards for chemicals having neurotoxic effects. To determine this, the integrated risk information system (IRIS) of the U.S. EPA was accessed, and the use of neurotoxicological findings in establishing standards and health advisories (HAs) was determined. IRIS is a computerized database designed by the U.S. EPA's Office of Health and Environmental Assessment for use by staff at U.S. EPA program offices and research centers to inform U.S. EPA personnel concerning the regulatory status of specific chemicals.

Table 3. Neurotoxicological findings in IRIS database[a].

Standard/advisory	Total number of chemicals	Chemicals with neurotoxic end points[b]	% Neurotoxic
RfDs	336	65	19.4
RfCs	32	12	37.5
1-Day child HA	13	6	46.1
10-Day child HA	35	7	20.0
Longer term child HA	37	7	18.9
Longer term Adult HA	37	7	18.9
DWEL/life-time HA	57	11	19.3

[a]As of January 1994. [b]Chemicals for which critical effect was based wholly or in part on neurotoxic end points.

The IRIS database provides detailed information concerning the adverse effects used to generate standards and HAs. Table 3 summarizes the number of chemicals contained in the database for each category of standards or HAs. Of the 336 chemicals for which there were RfDs, 65 or 19.4% listed neurotoxicity as the scientific data used wholly or in part for regulation. Of the 32 chemicals having inhalation reference concentrations (RfCs), which are equivalent to RfDs, 12 or 37.5% were based on neurotoxicity. Neurotoxicity was also listed for a number of 1-Day Child HAs (46.1%), 10-Day Child HAs (20%), Longer-Term Child HAs (18.9%), Longer-Term Adult HAs (18.9%), and Lifetime HAs (19.3%). The percentage of chemicals identified as having neurotoxic effects in IRIS compares favorably with the estimates reported by the OTA (*7*) that 3 to 28% of all chemicals in the environment may be neurotoxic. The chemicals identified in the IRIS database included a number of solvents (e.g., styrene, toluene, xylenes, acrylamide, *n*-hexane, propylene glycol monomethyl ether), pesticides (e.g., carbofuran, methyl parathion, pydrin, thiram, methamidophos), and metals (e.g., manganese, methylmercury).

Therefore, in spite of the lack of a tiered-testing strategy, neurotoxic effects are used by the U.S. EPA as the basis for setting standards and health advisories.

Summary and Conclusions

A tier-testing strategy for the assessment of neurotoxicity has been proposed by numerous expert panels and, most recently, by the NRC (*15*). Although such a testing strategy is conceptually attractive, it is not practical in most regulatory contexts. The current decision-making process at U.S. EPA utilizes all available animal data collected in hazard identification, characterization, or mechanism-of-action studies, as well as human epidemiological, case, and clinical studies, to arrive at a critical adverse effect. The risk assessment process then utilizes the critical adverse effect to support regulatory decision making. Considering the proportion of cases in which neurotoxicity is used as the critical adverse effect, the current risk assessment system identifies about 20% of the chemicals listed as neurotoxic, which agrees well with the range of values suggested by the OTA (*7*). As more is learned about the mechanism of action of neurotoxic chemicals, future testing may evolve into a modified tier-testing strategy where short-term, *in vitro* tests are used at the first tier and information from these studies is used to determine if testing at the second tier should occur. Second-tier tests would include *in vivo* tests selected on the basis of the results of the first tier.

REFERENCES

1. Anger WK. Worksite behavioral research: results, sensitive methods, test batteries and the transition from laboratory data to human health. Neurotoxicology 11:629–720 (1990).
2. Tilson HA. Screening for neurotoxicity: principles and practices. J Am Coll Toxicol 8:13–17 (1989).
3. Anger WK, Johnson BL. Chemicals affecting behavior. In: Neurotoxicity of Industrial and Commercial Chemicals (O'Donoghue J, ed). Boca Raton, FL:CRC Press, 1985;51–148.
4. Anger WK. Workplace exposures. In: Neurobehavioral Toxicology (Annau Z, ed). Baltimore:Johns Hopkins University Press, 1986;51–148.
5. Anger WK. Neurobehavioral testing of chemicals: impact of recommended standards. Neurobehav Toxicol Teratol 6:147–153 (1984).
6. Centers for Disease Control (CDC). Leading work-related diseases and injuries—United States. Morb Mortal Wkly Rep 32:24–26 (1983).
7. Office of Technology Assessment. Neurotoxicity: Identifying and Controlling Poisons of the Nervous System. Rpt OTA-BA-436, Washington:U.S. Government Printing Office, 1990.
8. National Academy of Sciences. Principles for Evaluating Chemicals in the Environment. Washington:National Academy

of Sciences, 1975.
9. U.S. EPA. OPPTS Test Guidelines: Neurotoxicity. PB91-154617. National Technical Information Service, 1991.
10. Kimmel CA. Current status of behavioral teratology: science and regulation. Crit Rev Toxicol 19:1–10 (1988).
11. Tilson HA. Neurotoxicology in the 1990s. Neurotoxicol Teratol 12:293–300 (1990).
12. U.S. EPA. Draft Report: Principles of Neurotoxicity Risk Assessment. Fed Reg 58:41556–41599 (1993).
13. Evans HL, Weiss B. Behavioral toxicology. In: Contemporary Research in Behavioral Pharmacology (Blackman DE, Sanger DJ, eds). New York:Plenum Press, 1978;449–487.
14. MacPhail RC, Peele DB, Crofton KM. Motor activity and screening for neurotoxicity. J Am Coll Toxicol 8:117–125 (1989).
15. National Research Council. Environmental Neurotoxicology. Washington:National Academy Press, 1992.
16. Choi DW. Glutamate neurotoxicity and diseases of the nervous system. Neuron 1:623–634 (1988).
17. Tilson HA, Moser VC. Comparison of screening approaches. Neurotoxicology 13:1–14 (1992).
18. Rees DC, Francis EZ, Kimmel CA. Scientific and regulatory issues relevant to assessing risk for developmental neurotoxicity: an overview. Neurotoxicol Teratol 12:175–181 (1990).
19. Levine TE, Butcher RE. Workshop on the qualitative and quantitative comparability of human and animal developmental neurotoxicity. Work Group IV report: triggers for developmental neurotoxicity testing. Neurotox Teratol 12:281–284 (1990).
20. U.S. EPA. Notice of proposed rulemaking for fuels and fuel additives: registration regulations. Fed Reg 57:13168 (1992).
21. Francis EZ, Farland WH. Application of the preliminary developmental toxicity screen for chemical hazard identification under the Toxic Substances Control Act. Teratog Carcinog Mutagen 7:107–117 (1987).
22. Barnes DG, Dourson DG. Reference dose (RfD): description and use in human health risk assessments. Regul Toxicol Pharmacol 8:471–486 (1988).
23. Sette WF, MacPhail RC. Qualitative and quantitative issues in assessment of neurotoxic effects. In: Neurotoxicology (Tilson H, Mitchell C, eds). New York:Raven Press, 1992;345–361.