Neurobehavioral Aspects of Developmental Toxicity Testing

Beate Ulbrich¹ and Anthony K. Palmer²

¹Bundesinstitut für Arzneimittel und Medizinprodukte, Berlin, Germany; ²Huntingdon Research Centre, Huntingdon, Cambridgeshire, Great Britain

Tests for detection of neurobehavioral changes in the offspring have been a regulatory requirement in developmental toxicity testing of drugs for almost 20 years. Keeping their purpose of hazard identification and risk assessment for humans in mind, investigators and agency reviewers have become deeply ingrained with some stereotyped behaviors with respect to such relevant issues as choice of animal species and data evaluation. Other problematic areas of study design and conduct, selection of litter representatives for testing, what methods to combine in a testing battery, and statistical treatment of results and their interpretation, will need more research and discussion in the future. — Environ Health Perspect 104(Suppl 2):407–412 (1996)

Key words: prenatal, postnatal, developmental toxicity, behavioral teratology, regulatory studies, guidelines, detection of effects, drugs

Introduction

The thalidomide tragedy in the early 1960s brought about a worldwide realization that drugs, pesticides, and other chemical substances have the potential to induce damage in the unborn child. This was followed by the introduction of new guidelines for preclinical testing requirements. Naturally, interest focused mainly on structural abnormalities, and testing strategies were devised that were expected to detect and characterize prenatal insults leading to gross morphological changes in the embryo and fetus. Although experience has shown that terata occur only rarely compared with other end points of developmental toxicity, such as effects on growth and viability, the attitude that malformations are all important still persists with many investigators and regulatory agency reviewers.

As early as 1963, a fourth area of concern, behavioral teratology, was introduced in a review by Werboff and Gottlieb (1) on the postnatal effects of prenatal X-irradiation and exposure to psychoactive drugs. However, regulatory action was not taken until 1975 when Great Britain and Japan incorporated requirements for developmental neurotoxicity testing into their respective guidelines for testing of medicinal products for reproductive toxicity. At this time, the relevance to humans of the potential of chemicals to induce damage to the central nervous system following prenatal exposure and exposure during childhood had become widely accepted based on the data that were available for organic mercury, lead, and alcohol. Although no validated methods were available, it was felt that early detection of a substance's potential for developmental neurotoxicity in animal experiments could prevent widespread exposure of pregnant women and thereby minimize or eliminate the risk for the growing and developing child. It was assumed that, for an unknown compound, where no clues about the possible localization of a potential lesion existed, functional tests might give greater sensitivity than histopathological and biochemical methods. The underlying biological mechanisms could then be elucidated by secondary studies from the functional changes observed in first-pass testing. Based on this rationale, testing of drugs for end points of developmental neurotoxicity commenced in the mid-1970s, but it was not until the early 1980s that behavioral testing batteries became established as routine tests in the pharmaceutical industry.

Since that time a large amount of data on tests and test combinations for medicinal products has accumulated in the archives of regulatory agencies and pharmaceutical companies, data that should be reexamined critically with the aim of identifying methods that may be recommended.

Developmental Neurotoxicity within the Framework of Regulatory Studies

It was clear from the beginning that, for therapeutic agents, neurobehavioral toxicity testing would have to be incorporated into existing study designs for the detection of any (adverse) effect on development. These would have to be adapted to allow the collection of information on functional changes, in addition to the data on viability, growth, and gross structural abnormalities of conceptus and offspring (2). Regulatory studies for developmental toxicity do not detect these different end points equally well. They can be considered fairly sensitive for effects on viability and general growth parameters (body weight); however, when the emphasis is placed on rare events like malformations, a study size of usually 10 to 20 pregnant animals per group will always be insufficient in picking up any but the strongest effects. Also, the different variables constituting an embryofetotoxic effect do not usually occur with an even distribution within and between litters of a dose group. Some litters may be free of any relevant findings, others may contain only one or a few affected fetuses or pups, or, alternatively, the whole litter may be abnormal. For nonfunctional end points these distributions can be determined without great difficulty. But what about effects on the functional integrity of the central nervous system? How can we be sure that the litter representatives chosen for testing (in most studies by random methods) do indeed carry an alteration? If individual distributions of functional changes are in any way similar to those existing for other end points of aberrant development, with only a selection of animals from each litter being tested, in routine studies we should expect to miss quite a few substances that affect CNS function.

Behavior may well be the most variable parameter of all the responses an organism can make to a developmental toxicant.
This is logical because uncompromised animals will exhibit a wide range of complex, adaptive behaviors without having to compensate for substance-induced deficits, and relatively small changes in the environment, e.g., handling of animals, maternal influences, food restriction, and prior test experience, have been shown to produce alterations in normal behavior.

**Characteristics of Developmental Neurotoxicity**

Studies to detect developmental toxicity differ in several important aspects from tests in which toxicity is elicited in adults. The main difference is the impossibility of deriving untreated (control) as well as treated values from the same animals. Comparisons to determine whether there is a treatment effect always will be made between groups of animals that have different experiences related to exposure and, even in inbred animals, different genetic compositions.

The developing nervous system may be more sensitive to toxic effects than that of the adult. Neurotoxic effects of a chemical may occur at lower doses than in adults, different functions may be affected, and substances that show no neurotoxic potential in adults may induce effects during development. Therefore, we may suspect a potential for developmental neurotoxicity if a compound induces neurotoxic changes in adults, but we have to be aware of the fact that compounds that do not affect the adult nervous system may very well do so in the developing organism. As testing is performed to detect the unexpected, it will be necessary to study end points for developmental neurotoxicity with all substances for which exposure of the embryo, the fetus, the newborn, the child, and the juvenile cannot be excluded definitely. Unlike malformations that can only be induced during a narrow time window in organogenesis, functional changes can be expected to occur during this differentiation phase and, additionally, for as long a period as the organ system needs to attain full functional competence. For detection of such effects, animals will have to be exposed during embryo-fetal development and postnatally through puberty. Observation, however, will have to continue for a longer time period, ideally to old age, to make sure that delayed manifestations are not overlooked. None of the study designs currently recommended by guidelines includes effects that may become apparent only in aging animals, e.g., premature onset of senescence or, with respect to CNS function, senility.

**Animal Species**

Rabbits, rats, and mice are the animal species primarily used in routine developmental toxicity testing; however, the potential of inducing neurobehavioral toxicity in the offspring is evaluated almost exclusively in rats (Table 1). This is due to the fact that regulatory agencies have accepted data from rats in cases in which this species proved to be an unsuitable animal model for the substance under study when they should have encouraged the use of another animal species. This practice reveals astonishing insights into how great an importance is attached to possible effects with postnatal manifestations in humans (including neurobehavioral findings) during the process of hazard identification and risk assessment. At present, we are making the world safer for rats. But how secure can we feel about the detection of hazards for the developing nervous system when this animal model is not even reasonably close to humans? Even if we do not yet know how changes in animal behavior may translate to the situation in humans, the least we can do is to use the most appropriate animal model available to us, i.e., that closest to humans with respect to metabolism, pharmacokinetics, pharmacodynamics, and physiology. If such a species cannot be found or used, we should consider conducting studies in more than one animal species. This has been standard procedure in the testing for structural abnormalities and still is, despite intellectual acknowledgement that one relevant species is better than two or more less suitable ones. If we want to increase the predictability of animal results for humans, it will be necessary to develop methods for species other than rats and to apply them in those cases where the rat is not a relevant model.

**Methods**

Certainly, none of the commonly used laboratory animal models can match the complexity of human behavior. For detection studies, the animal model, the testing situation, and the available methods provide the limiting factors and restrict investigators to analyzing basic neurological functions and simple behaviors. Even given these restrictions, there are more specific functions than we could hope to incorporate and test in a single, comprehensive study design. It may be considered advantageous that the first guidelines required testing of specific functions of the central nervous system but, for lack of experience, did not specify which tests were to be used. This has resulted in a diversification of methods and in a great variety of testing batteries that are in regular use today. It should be possible to identify sensitive and reliable tests with predictive value for the human situation. It is unlikely, however, that a single ideal combination of testing procedures could be defined—one that would cover all aspects of developmental neurotoxicity and that could be conducted at reasonable costs.

Criteria for the selection of tests for a testing battery have been described (3). For detection of any (adverse) effect, preference is given to apical tests that require the integrated function of several subsystems. These may offer the best chance to discover whether the substance poses a hazard to development and function of the CNS based on the assumption that a change in any of the subsystems can lead to an alteration in behavioral output. On the other hand, with an increasing number of subsystems involved, the animal will have greater possibilities of compensating for deficits in one subsystem. The choice of methods should be aimed at having available a set of apical tests that are neither too complex nor too specific to incorporate into routine developmental toxicity studies and to supplement this battery with close observation of the animals. If these give indications for changes in behavioral end points, other more sophisticated tests can be used to clarify and characterize the results obtained by the base set. Testing batteries normally combine measurements of growth and physical development (2–6) with tests for the development of sensory functions, reflexes, and body control, and protocols for detecting changes in locomotor activity.
learning/memory, and social/reproductive behavior. As it will not be possible here to describe methods in detail, the reader is referred to several comprehensive reviews on testing procedures and their respective merits (3–8).

Reliability/Reproducibility
Regarding reliability of testing procedures, agency experience shows that most of the tests incorporated into testing batteries and retained by the investigators over the years can be considered standardized and validated with respect to intralaboratory reproducibility. Investigators do not tend to continue using tests that will give vastly different results from study to study, and it can be seen from the submitted reports that comparable values are found for control groups over time. Interlaboratory reproducibility has not been evaluated for all the tests used, but the results from the comparison of some of these methods in the study of the National Center for Toxicological Research (NCTR) have been encouraging (9,10).

Sensitivity
Detection sensitivity of behavioral measurements has been evaluated for the methods used in the NCTR Collaborative Behavioral Teratology Study, and it can be stated that variability of the measured parameters will allow detection of effects if they are large enough (approximately 10–20% change from control). However, it would appear that neither this testing battery, nor any other in current use, could be relied on to detect a developmental neurotoxicant among a series of unknown compounds. One of the presumed positive control substances for the Collaborative Behavioral Teratology Study, d-amphetamine, later gave negative findings consistently within and across the participating laboratories (9). So either the assumption of d-amphetamine being a strong developmental neurotoxicant was wrong or the test battery was not suited to detect the deficit the substance did induce (11).

For a broader comparison of methods not only for detection of effects but also for characterisation, a European collaborative study group was initiated. During this study each participating group applied the methods used in their laboratory to the task of detecting neurobehavioral changes induced by a known positive. The outcome of this investigation shows that it is not necessary to work with a standardized set of methods to detect adverse effects on the behavior of offspring (12,13).

Comprehensiveness
Selection of a comprehensive testing battery is a crucial point, especially as it will not become apparent until much later whether the aim has been achieved. Although no consensus for recommending specific tests is in sight, there seems to be some general agreement on the functions that should be tested, namely, sensory systems, reflexes, neuromotor development, locomotion/activity, reactivity/habituation, learning/memory, and social/reproductive behavior. To integrate behavioral data into the context of other manifestations of developmental toxicity, data on physical development of the offspring have to be available. These commonly include data on body weights and postnatal weight gain, viability, physical landmark development and maturation. It has also been recommended to maintain records of organ weights, especially brain, functional observation battery results, neuropathologic examinations (14), and brain biochemistry (15,16).

Predictability
Little can be said about whether the tests in current use predict that similar (or different) effects on CNS development and function would be elicited in humans. They are able to identify known human developmental neurotoxicants, but it can be argued that this is due to selection bias and to the fact that we already know what to look for with these substances. Predictability could be evaluated by using data on new therapeutic agents, but lack of human data effectively prevents this.

Computerized Procedures versus Human Observers
For a novel, unknown compound, detection of an effect will depend to a large extent on the observational skill and the knowledge of the investigators, who can do what a standardized test is unable to accomplish; that is, they can pick up unexpected effects by observation and verify them by specifically designed procedures. Most tests yield not only variables that can be measured exactly but also give rise to findings for which measurement is difficult or impossible and that will have to be observed and described.

A simple water maze, for example, which is part of many routine testing batteries, will be used to collect data on learning ability and memory. The parameters recorded routinely are whether the animal is successful within the time limit, the number of errors made, and the time needed to escape from the maze. Experience shows that most (all?) animals will learn the route that takes them to the exit easily once they have managed to discover (or have been shown) where it is situated. Probably this is not a very sensitive test for the detection of subtle differences in learning/memory functions, as the performance of rats is quite variable even in control groups, and the demands on the central nervous system of this simple task do not seem to be high enough to bring out clear effects on learning ability when brain damage is slight. In addition, the way the test is applied and evaluated, often only as a measure of learning, does not make use of its full potential. The first trial, in which the naive animal has no clue about the location of the exit, more often than not is treated as a training run, and, therefore, is not considered for further analysis of (learning) behavior. In a study report, the reviewer will be told how many animals failed to reach the exit in time, but the reasons why they failed to do so are never described. If this were done, we could gain insight into problem-solving abilities and strategies that might be more sensitive to chemical insults than simple learning tasks; this also may be more relevant for extrapolation to humans and for risk assessment.

Here human observers have definite advantages over automated systems. They are able to recognize behavioral changes in the subjects that have not been anticipated and are therefore not covered by the recording procedure of the program. On the other hand, humans are at a severe disadvantage when they are asked to carry out robotic functions, such as observing large numbers of animals in a specific test for hours and recording behavioral parameters. Human operators become bored and tired and their attention wanders unless it is triggered by something unusual. To design tests that can be employed safely in the detection of neurobehavioral toxicity, it is necessary to understand these limitations and to use both human observers and automated tests for the purposes they can serve best—humans to spot any uncommon and unpredicted response and computers for counting and recording tasks that can be anticipated and programmed.

What Have We Learned from Over 10 Years of Testing Therapeutic Agents?
In the overview that follows, we have followed the interpretation of the investigators who conducted the studies in categorizing...
findings as positive or negative. It must be kept in mind, however, that the most commonly used methods of statistical analysis in these studies apply measures of central tendency and that these are inappropriate to analyze values with a skewed distribution that are generated by many behavioral test procedures. Strategies to improve validity of developmental neurotoxicity testing should include improvement of data exploration and analysis.

Virtually all preclinical testing for developmental neurotoxicity with new medicinal substances has been carried out in rats (Table 1). Even the Japanese guidelines that required postnatal testing of offspring in embryotoxicity studies, in which two species traditionally have been used, have done so only for rodents but not for other species. Considering requirements for species selection in current guidelines for industrial chemicals and pesticides, the situation is similar in these areas and not likely to change in the near future.

Relevance of reproductive studies in rats, when rodents are inappropriate models for humans, is beginning to be addressed by harmonized guidelines. These aim at detection of reproductive toxicity for medicinal products that require the use of a relevant animal model. These considerations also apply for developmental neurotoxicity.

Effects on behavioral parameters are not uncommon in reproductive toxicity studies, regardless of whether the period of exposure occurs in early or late pregnancy or during lactation. They occur in an order of magnitude similar to skeletal variants or effects on postnatal viability and development. Not surprisingly, behavioral alterations in offspring are associated with maternal toxicity, with decreased pup weight, and with effects on postnatal physical development—usually delays (Tables 2–4).

If drugs are analyzed according to their indication group, it becomes apparent that positive findings on behavior are encountered in very different drug classes (Table 5) and not only with those compounds that are known to be centrally acting. In 24 of all the substances tested, behavioral changes were found either to be the only adverse effects that could be detected at any dose, or they occurred at the LOAEL together with other signs of developmental toxicity. Seven of these 24 compounds were antibiotic drugs. Since the effects were not expected, this shows the necessity of conducting developmental neurotoxicity tests for all substances to which the developing human will be exposed.

Almost all behavioral testing batteries contain one or more tests to measure activity. From experiences with the testing of new drugs, these seem to be very sensitive in picking up effects at low doses, maybe overly so, but for a detection study this would not be considered a disadvantage. Other tests and parameters that showed significant changes at low doses are active and passive avoidance learning and center latency in the open field test (Table 6).

Often effects are detected only in one sex (Table 7). Whether this is due to a true sex-specific action of the compound cannot be decided, as studies for secondary characterization are usually performed only if malformations are encountered in the routine studies, not for a suspected

| Table 2. Fertility and general reproduction studies. |
|-----------------|-----------------|-----------------|
|                 | Normal          | Affected        |
|-----------------|-----------------|-----------------|
| Total studies   | 141             | 47              |
| No maternal toxicity | 62             | 16              |
| Maternal toxicity | 73             | 26              |
| Maternal death   | 6               | 5               |
| Postimplantation loss normal | 35             | 7               |
| Postimplantation loss increased | 22           | 9               |
| Pup mortality normal | 86           | 25              |
| Pup mortality increased | 53           | 22              |
| Fetal weight normal | 72             | 23              |
| Fetal weight decreased | 56           | 23              |
| Pup weight normal | 72              | 18              |
| Pup weight decreased | 64           | 28              |
| Pup weight increased | 1            | 1               |
| Malformations normal | 93             | 32              |
| Malformations increased | 7            | 2               |
| Skeletal variant negative | 75           | 22              |
| Skeletal variant positive | 36            | 15              |
| Pregnancy duration normal | 97           | 36              |
| Pregnancy duration shortened | 3            | 0               |
| Pregnancy duration increased | 34          | 11              |
| Development normal | 107            | 25              |
| Development affected | 26             | 21              |

| Table 3. Embryotoxicity studies (Japanese guidelines). |
|-----------------|-----------------|-----------------|
|                 | Normal          | Affected        |
|-----------------|-----------------|-----------------|
| Total studies   | 71              | 26              |
| No maternal toxicity | 21             | 4               |
| Maternal toxicity | 46             | 18              |
| Maternal death   | 4               | 4               |
| Postimplantation loss normal | 32           | 16              |
| Postimplantation loss increased | 12          | 8               |
| Pup mortality normal | 49           | 19              |
| Pup mortality increased | 21           | 6               |
| Fetal weight normal | 33             | 8               |
| Fetal weight decreased | 34           | 16              |
| Pup weight normal | 24              | 7               |
| Pup weight decreased | 18           | 10              |
| Pup weight increased | 2            | 0               |
| Malformations normal | 58            | 20              |
| Malformations increased | 9            | 3               |
| Skeletal variant negative | 31           | 14              |
| Skeletal variant positive | 34            | 11              |
| Pregnancy duration normal | 26          | 12              |
| Pregnancy duration shortened | 2            | 0               |
| Pregnancy duration increased | 12          | 7               |
| Development normal | 52             | 17              |
| Development affected | 17             | 8               |

| Table 4. Peri-postnatal studies. |
|-----------------|-----------------|-----------------|
|                 | Normal          | Affected        |
|-----------------|-----------------|-----------------|
| Total studies   | 133             | 64              |
| No maternal toxicity | 50             | 19              |
| Maternal toxicity | 73             | 36              |
| Maternal death   | 10              | 9               |
| Postimplantation loss normal | 44           | 12              |
| Postimplantation loss increased | 30          | 14              |
| Pup mortality normal | 64             | 26              |
| Pup mortality increased | 69           | 38              |
| Pup weight normal | 51              | 19              |
| Pup weight decreased | 73            | 45              |
| Pup weight increased | 2            | 0               |
| Pregnancy duration normal | 102          | 47              |
| Pregnancy duration shortened | 3            | 1               |
| Pregnancy duration increased | 26          | 14              |
| Development normal | 89             | 38              |
| Development affected | 37            | 22              |

| Table 5. Studies in rats with behavioral changes by drug indication classes.a |
|-----------------|-----------------|-----------------|
| Drug class      | Fertility | Embryotoxicity | Peri-postnatal |
|-----------------|-----------------|-----------------|
| Analgesic/        | 3               | 2               | 3              |
| antiphlogistic    | (9)             | (6)             | (15)           |
| Antiallergic      | 5               | 2               | 3              |
| Corticoid         | (17)            | (8)             | (11)           |
| Antimicrobial     | 10              | 6               | 13             |
| Antiviral         | (21)            | (21)            | (33)           |
| Antimycotic       | 2               | 0               | 2              |
| Cardiovascular    | (15)            | (18)            | (41)           |
| Antiocoagulant    | 1               | 0               | 0              |
| Anticonvulsant    | (4)             | (1)             | (4)            |
| Psychoactive      | (3)             | (3)             | (3)            |
| Anticonvulsant    | 1               | 3               | 4              |
| Anticoagulant     | (15)            | (5)             | (15)           |
| Muscle relaxant/  | (1)             | 1               | 3              |
| spasmolytic       | (2)             | (2)             | (4)            |
| Expectorant/      | 1               | 2               | 1              |
| broncholytic      | (8)             | (5)             | (3)            |
| Hyperlipemic      | 0               | 0               | 1              |
| Gastrointestinal  | (3)             | (3)             | (5)            |
| Hormone/hormone   | 2               | 0               | 1              |
| Cytostatic        | (8)             | (5)             | (6)            |
| (2)             | (3)             | (3)             |

---

a Data represent the number of positive studies, with the total number of studies shown in parentheses.
Table 6. Behavioral tests giving positive results in detection studies for therapeutic agents.

| Test system or parameter | All studies | Only at mat tox | Below mat tox | As LOAE dev tox |
|--------------------------|-------------|----------------|--------------|----------------|
| Activity (all)*          | 34          | 19             | 15           | 9              |
| Hole board               | 3           | 2              | 1            |                |
| Open field               |             |                |              |                |
| Activity                 | 14          | 10             | 4            | 3              |
| Latency                  | 7           | 1              | 6            | 5              |
| Rearing                  | 10          | 4              | 6            | 3              |
| Grooming                 | 2           | 1              | 1            | 2              |
| Defecation               | 5           | 2              | 3            | 2              |
| Avoidance learning       | 7           | 1              | 6            | 2              |
| Water mazes              | 13          | 9              | 4            | 2              |
| Rotarod                  | 14          | 6              | 8            | 4              |
| Swimming development     | 4           | 4              |              |                |
| Cliff avoidance          | 1           |                | 1            |                |
| Negative geotaxis        | 3           |                |              |                |
| Olfactory discrimination | 1           |                |              |                |
| Homing                   | 2           | 2              |              |                |
| Righting                 | 16          | 9              | 7            | 1              |
| Grasping                 | 4           | 1              | 3            | 1              |
| Placing                  | 1           |                | 1            | 1              |

Abbreviations: mat tox, maternally toxic dose; LOAE, lowest observed adverse effect; dev tox, developmental toxicity. *Includes open field and hole board locomotor activity.

effect on behavior. Unfortunately, with the possible exception of rearing behavior, the available database is still not large enough to identify specific functions or parameters that are influenced preferentially in either males or females.

Results

The relevance of the results within the animal model has to be decided before extrapolation to humans is attempted. Magnitude of the effect, reversibility, and possible relations to other effects, developmental or maternal, will have to be considered. Changes that do not persist as the animal gets older will probably be judged to have a different impact than permanent effects, although with behavioral end points the ability of the animal to compensate for deficits has to be taken into account.

Risk Assessment and Risk Management

Are behavioral results of animals exposed to a substance during development predictive of safety or hazard for developing humans? Animal models are definitely affected by the potent neurotoxicants that have been found to induce brain damage and dysfunction in the human conceptus and children, but these substances are too few in number to allow generalization. Moreover, substances presumed to be safe in humans have not been evaluated to the same extent, and no good evidence can be gained from the testing of new compounds because data on effects in humans will not be available for many years to come.

The aim of preclinical testing is primary prevention. If neurobehavioral changes are encountered at a relevant dose in an animal model exposed to a drug during development, the regulatory agency will not be in a great hurry to find out what happens when this drug is given to pregnant women. The drug would be treated like a substance that induces structural abnormalities in animals. If no other problems prohibit granting of a license, the drug could be placed on the market. However, use during pregnancy, during lactation, and in children would be contra-indicated unless the drug has lifesaving properties or another clear benefit that would justify the perceived risk. Women of child-bearing potential would be advised to take contraceptive measures during and after treatment (for drugs with a longer half-life). For these reasons, we will not learn—unless by accident—whether the toxic potential, teratogenic or functional, is relevant for humans. Inadvertent exposure does occur, but given that behavioral responses are more variable and differences from normal may be more subtle rendering them less conspicuous than morphological effects, it is difficult to imagine how isolated cases of behavioral abnormalities could be noted and reported.

REFERENCES

1. Werboff J, Gottlieb JS. Drugs in pregnancy: behavioral teratology. Obstet Gynecol Surv 18:420-423 (1963).
2. Lochry EA. Concurrent use of behavioral/functional testing in existing reproductive and developmental toxicity screens: practical considerations. J Am Coll Toxicol 6:433-439 (1987).
3. Buelke-Sam J, Kimmel CA. Development and standardization of screening methods for behavioral teratology. Teratology 20:17-30 (1979).
4. Zbinden G. Experimental methods in behavioral teratology. Arch Toxicol 48:69-88 (1981).
5. Alder S, Zbinden G. Neurobehavioral tests in single- and repeated-dose toxicity studies in small rodents. Arch Toxicol 54:1-23 (1983).
6. Tilson HA. Behavioral indices of neurotoxicity: What can be measured? Neurotoxicol Teratol 9:427-443 (1987).
7. Spear LP. Neurobehavioral assessment during the early postnatal period. Neurotoxicol Teratol 12:489-495 (1990).
8. Sobrian SK, Pappas BA. Advantages and disadvantages of lon-

Environmental Health Perspectives • Vol 104, Supplement 2 • April 1996 411
9. Buelke-Sam J, Kimmel CA, Adams J, Nelson CJ, Vorhees CV, Wright DC, Omer VS, Korol BA, Butcher RE, Geyer MA, Holson JF, Kutscher CL, Wayner MJ. Collaborative Behavioral Teratology Study: results. Neurobehav Toxicol Teratol 7:591–624 (1985).
10. Vorhees CV. Reliability, sensitivity and validity of behavioral indices in neurotoxicity. Neurotoxicol Teratol 9:445–464 (1987).
11. Weissman A. What it takes to validate behavioral toxicology tests: a belated commentary on the Collaborative Behavioral Teratology Study. Neurotoxicol Teratol 12:497–501 (1990).
12. Elsner J, Suter KE, Ulbrich B, Schreiner G. Testing strategies in behavioral teratology: IV. Review and general conclusions. Neurobehav Toxicol Teratol 8:385–390 (1986).
13. Elsner J, Hodel B, Suter KE, Oelke D, Ulbrich B, Schreiner G, Cuomo V, Cagiano R, Rosengren LE, Karlsson JE, Haglid KG. Detection limits of different approaches in behavioral teratology and correlation of effects with neurochemical parameters. Neurobehav Toxicol Teratol 10:155–167 (1988).
14. Buelke-Sam J, Mactutus C. Workshop on the qualitative and quantitative comparability of human and animal developmental neurotoxicity. Work Group II Report: testing methods in developmental neurotoxicity for use in human risk assessment. Neurotoxicol Teratol 12:269–274 (1990).
15. Saillenfait AM, Vannier B. Methodological proposals in behavioral teratogenicity testing: assessment of propoxyphene, chlorpromazine, and vitamin A as positive controls. Teratology 37:185–199 (1988).
16. 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).