

---

# Economical Test Methods for Developmental Neurobehavioral Toxicity

Giorgio Bignami

Laboratorio di Fisiopatologia di Organo e di Sistema, Istituto Superiore di Sanità, Viale Regina Elena, Roma, Italy

The assessment of behavioral changes produced by prenatal or early postnatal exposure to potentially noxious agents requires both the designing of ad hoc tests and the adaptation of tests for adult animals to the characteristics of successive developmental stages. The experience in designing tests is still more limited than in the adaptation of tests, but several tests have already proven their usefulness; some examples are the suckling test, the homing test, and evaluations of dam-pup and pup-pup interactions. Functional observational batteries can exploit the development at specified postnatal ages of several reflexes and responses that are absent at birth in altricial rodent species with a short pregnancy such as the rat and the mouse. In neonates, the assessment of early treatment effects can rely not only on deviations from normal responding but also on changes in the time of appearance of otherwise normal response patterns. The same applies to other end points such as responses to pain and various types of spontaneous motor/exploratory activities, including reactivity to a variety of drug challenges that can provide information on the regulatory systems whose development may be affected by early treatments. In particular, the analysis of ontogenetic dissociations (i.e., differential early treatment effects depending jointly on developmental stage at the time of exposure, age of testing, and response end point) can be of considerable value in the study of treatments' mechanisms of action. Overall, it appears that behavioral teratological assessments can be effectively used both proactively, i.e., in risk assessment prior to any human exposure, and reactively. In the latter case, these assessments could have special value in the face of agents suspected to produce borderline changes in developing humans, whose innocuousness or noxiousness can be difficult to establish in the absence of hard evidence of teratogenicity. — *Environ Health Perspect* 104(Suppl 2):285–298 (1996)

Key words: behavioral teratology, functional observational batteries, pain reactivity, motor/exploratory activities, methylmercury, ethanol, benzodiazepines, ozone (O<sub>3</sub>), rat, mouse

---

## Introduction

Considerable progress has been made over the past two decades in the validity assessment of a variety of response end points in experimental behavioral teratology, i.e., in studies that are aimed at verifying whether prenatal or postnatal

exposure to a physical or chemical agent produces significant changes in neurobehavioral development [for reviews of methods and results, see (1–9); for proceedings of ad hoc meetings, see (10–19)].

---

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.

The research presented in this paper was performed at the Istituto Superiore di Sanità (Laboratory of Organ and System Pathophysiology) and was supported by the Sub-Project on Behavioral Biology and Pathophysiology (Project on Non-infectious Pathology) of the Istituto Superiore di Sanità (1991–1995), by grant 91.00494.CT04 from the Italian National Research Council (CNR), and by the CNR-ENEL Project "Interaction of energy systems with human health and environment"—Rome, Italy (1991–1994). The author gratefully acknowledges the help given by several colleagues in the Behavioral Pathophysiology Section and the Comparative Psychology Section of the laboratory, whose extensive experience in developmental and behavioral teratological research was exploited in the present analysis of test methods.

Address correspondence to Dr. Giorgio Bignami, Laboratorio di Fisiopatologia di Organo e di Sistema, Istituto Superiore di Sanità, Viale Regina Elena, 299, I-00161 Roma, Italy. Telephone: 39-6-444.0061. Fax: 39-6-495.7821.

Abbreviations used: po, *per os*; ip, intraperitoneal; MM, methylmercury; MAM, methylazoxymethanol; CNS, central nervous system; PCs, personal computers; NGF, nerve growth factor; CPP, conditioned place preference; ANOVA, analysis of variance.

A first distinction that needs to be made is that behavioral teratology assessments must often be performed with different background information, under different conditions, and sometimes with different goals, depending on the agent considered. In the case of pharmaceutical drugs, for example, the information on biological effects and mechanisms of action, body burden, and other important properties is generally much more extensive than in the case of industrial chemicals. Moreover, in the case of medicines the routine routes of exposure (po or ip) can generally suffice in most investigations, and negative results are only moderately useful; in the case of industrial agents, inhalation or percutaneous exposure often need to be considered, and negative and positive results are equally useful (20).

Since the early work in behavioral teratology (21), economical test methods have made a substantial contribution to risk assessment; in addition, they have often provided information that can be readily translated into useful working hypotheses concerning the significance and the nature of the effects observed, which is essential for deciding whether higher-tier studies are necessary. Moreover, these methods have played a major role in collaborative studies aimed either at checking the replicability of the results by experiments performed in different laboratories under identical conditions (22) or at widening the range of assessments after a specified treatment by subdividing the test burden between different laboratories (23).

Behavioral teratology covers a very broad area that includes at one end the study of behavioral changes produced by treatments whose teratogenic effects are clearly documented by neuropathological assessments, as is the case with quite different agents such as methylmercury (MM), methylazoxymethanol (MAM), and ethanol. Even in these extreme cases, animal behavior studies can play an essential role in the detection and definition of changes produced by exposures to doses or concentrations whose effects are below the threshold of neuropathological methods. In humans, this applies to children exposed either *in utero* to low doses of ethanol (24,25) or prenatally and postnatally to low doses of lead (26). In both instances, human and animal data have been shown to be in good agreement with each other (27,28).

At the opposite end, behavioral teratology studies are essential to assess possible risks created by exposures to agents that do

not have a substantial teratogenic potential as assessed by the use of somatic end points. The continuing discussion on the teratogenic potential of drugs such as some of the antiepileptics and anxiolytics shows that the distinction between teratogenic and nonteratogenic agents cannot be a sharp one; however, there are considerable differences (both qualitative and quantitative) between the CNS damage produced by prenatal exposure to these drugs and that produced by agents such as MM, MAM, and ethanol (at high doses). In addition, information on the short- and long-term biochemical effects produced by agents of both types is not readily amenable to an interpretation in the absence of behavioral assessments or when negative behavioral data or minimal changes are found after treatments that produce considerable CNS changes [see (29) for ethanol; (30) for parathion; (31) for methyl demeton (Meta-Systox)].

Essentially, behavioral investigations must be sufficiently extensive and specific to assess the meaning of observed correlations or, at least, to help identify the mechanisms that induce a behavioral expression of a given CNS change in some situations, but not in others. For example, the dose-dependent increase in preweaning locomotor activity observed in rats after prenatal anticonvulsant (phenytoin) treatment when using a square open field failed to occur in the same laboratory when using a circular field (32). Neonatal antidepressant (clomipramine) treatment resulted in open-field hyperactivity at the adult stage, whereas activity in a closed chamber was not affected; this suggested that hyperreactivity to mild stress was the cause of the former change (33). In the case of haloperidol treatment on postnatal days 4 to 21, which is known to produce considerable effects on CNS dopaminergic mechanisms, mice did not show activity changes in a photocell activity cage, in a circular alley, and in an operant chamber. However, activity was significantly increased in a presumably more stressful situation (the open field), but only in 4 out of 12 combinations of animal strain, age, and time of testing (34). These examples of complex interaction profiles should not be construed as an obstacle to the use of simple and economical assessment procedures. The experimenter, however, should be aware of the frequent occurrence of such interactions so as to understand the risk of false negatives and false positives and to be able to decide whether more complex and expensive higher tier assessments are a real need or, vice versa, a useless luxury.

The emphasis on the need for adequate behavioral investigation does not mean that any deviation from the behavioral profile of control animals is to be interpreted as a sign of pathology (see section "Pain Reactivity and Analgesic Drug Effects" for some indications on response variation produced by a variety of early influences). Specifically, several types of isolated behavioral changes can be essentially neutral with respect to adequate functioning of homeostatic processes; this, for example, is likely to be the case with changes in activity levels when the scores of all treated subjects remain within the range of the control population, and other types of changes are consistently absent. At the present state of the art, no rule of thumb is available to separate the normal from the pathological in borderline cases; therefore, any reproducible behavioral change of less than negligible size should be taken as a warning that the agent causing it may require a more thorough assessment.

The survey of economical test methods will be preceded by some considerations concerning inference strategies (particularly those that exploit response and age dissociations of treatment effects) and research goals in relation to available resources.

### Inference Strategies, Research Goals, and Resources

In behavioral teratology studies, the expectancy is that different effects may be produced depending on when exposure occurs during development. Equally important is the fact that even studies using simple designs and economical test methods can often make an effective use of several types of contrasts in the effects' profiles, without a substantial increase of the test burden. A first type of contrast is that between responses which are modified, or not modified after a particular treatment; such contrasts often provide useful preliminary information on the treatments' mechanisms of action. A second type of contrast is that between presence or absence of a given response change depending on the age at the time of testing either during the treatment (if postnatal or combined prenatal and postnatal) or after treatment discontinuation. Contrasts of this kind allow the experimenter to compare the trend over time of treatment effects with the developmental pace of specified response systems; this provides information on the relative vulnerability of various regulatory mechanisms with different maturational rates, including recovery or compensation processes.

Additional useful contrasts can emerge from the analysis of pharmacological reactivity as assessed by appropriate drug probes or challenges. These contrasts can provide fairly specific information on neural regulatory mechanisms that are affected or spared, particularly when response and age variables turn out to play a role in determining whether reactivity to a particular agent is affected by an early treatment.

Some of the effects observed in developing mice after prenatal benzodiazepine treatment [oxazepam, 15 mg/kg po twice daily to the dam on pregnancy days 12–17; (35)] can be exploited to show the informative value of the contrasts so far mentioned. In fact, the pups showed a substantial reduction of open-field activity and response to amphetamine at 2 weeks of age, which was probably not due to the slight and transient retardation of somatic and neurobehavioral development produced by the treatment. At 3 weeks of age, activity and amphetamine reactivity were indistinguishable from those of the controls; in addition, the maturation of both habituation and hyperactivity response to an antimuscarinic (scopolamine) took place normally. These contrasts suggested a selective and transient effect on the development of monoaminergic regulatory mechanisms, which was further supported by the delayed appearance of morphine hyperactivity in the absence of changes in pain sensitivity or morphine analgesia [for additional biochemical and behavioral data that favor this interpretation, see (36)].

While the additional information provided by subsequent higher tier studies cannot be discussed here, this example confirms that a heavy test burden is not a *qua non* to combine risk assessment and a preliminary evaluation of treatments' mechanisms of action. In the case of benzodiazepines, information on the latter point turned out to be of potential value when recent data showed borderline neuropsychological deficits in children exposed prenatally to these agents (37).

As concerns the choice of an appropriate strategy in behavioral teratology studies, two factors must be considered jointly, i.e., the goals of a particular project and the size and type of available resources. On one hand, reactive studies concerning agents for which human data are already available (as was the case when animal studies on the effects of MM and lead were started) are best focused on end points that correspond as closely as possible to the profile of the effects observed (or suspected) in humans.

On the other hand, proactive studies concerning agents without a previous history of human exposure (or for which reliable data concerning possible effects in exposed humans are not yet available) must use tests that can pick up a wide variety of different effects at successive developmental stages. In general, these tests include an observational battery assessing postnatal neurobehavioral development, a simple activity test administered at successive developmental stages that are characterized by different response patterns, and one or more additional economical tests concerning end points with broad functional significance, such as pain reactivity and analgesic drug responses.

Subsequent decisions concern the choice of one or the other version of a test, which can be strongly influenced by the resources available. In particular, automated and nonautomated versions of any given test should be carefully weighed on the balance of potential advantages and disadvantages. In fact, automated tests can reduce the workload by one or more orders of magnitude and can also reduce the load of experimenter biases. (These biases, it should be noted, can never be entirely eliminated; consider, for example, the different amounts of stress produced by different handling styles, which can influence the animal's behavioral performances.) On the other hand, tests relying on direct observation (or, preferably, on the scoring of videorecordings) can provide information on a wider range of response end points than automated tests. This can allow the experimenter to pick up effects not recorded by an automated apparatus, identifying differences between the profiles of different treatments which may have similar effects on responses that can be automatically recorded (e.g., an overall increase or decrease of locomotor activity). At this point, the availability of either hard currency for the purchase of automated apparatus or low-cost manpower with adequate technical education (the simultaneous availability of both being the exception rather than the rule) can tip the balance toward the use of either automated or nonautomated tests.

When access to automated apparatus is limited and heavy reliance is made on trained observers (see section "Spontaneous Motor Activities" for specific guidelines), the following minimal requirements for the operation of a behavioral teratology laboratory need to be fulfilled. The first and foremost requirement is an adequate small animal maintenance facility with

internationally accepted standards of air quality, temperature, and humidity control. Since the same conditions must be guaranteed in testing rooms, this is likely to be the most expensive requirement with respect both to investment and to operating costs.

Additional requirements are a small shop for building, maintaining, and modifying a simple apparatus, one or more videorecording units, and one or more PCs, depending on the size of the research activity. PCs capacities are best exploited when the machines are used both for statistical analyses, relying on the highly effective software now available and on expert advice (so as to avoid the all-too-frequent statistical monstrosities), and for processing the input from manually operated keyboards used to score response events, relying on updated versions of *ad hoc* software, such as The Observer (38).

In the present overview, the space devoted to the description of the various tests has been limited as far as possible in order to assign more space to conditions of use, type of information expected, test limitations, and possible biases, including some of the more important (and sometimes unavoidable) biases in data analysis. This is to encourage readers, who can look elsewhere for more technical details, to adopt a critical and self-critical attitude, the only effective protection against those traps and pontifications that are all too frequent in behavioral research (39).

### Postnatal Sensory and Motor Development

The rat and the mouse are altricial species, that is, the pups are born in a highly immature condition after a short pregnancy (18–22 days, depending on species and strain). At birth, the eyes and ears are closed, the pup is able to crawl, and to get attached to a nipple and to suckle, while it needs close body contact with the mother for purposes of thermoregulation. Several reflexes and responses appear at successive postnatal stages in parallel with somatic changes, progressively increasing the pup's sensory and motor capabilities and, afterwards (particularly after eye opening around the end of the second postnatal week), its ability to procure food and fluid. Weaning in laboratory breeding units is performed at a fixed time after birth, generally at the end of the third postnatal week.

The time of occurrence of specified somatic changes and the time of first appearance and subsequent complete maturation of various reflexes and responses show a

remarkable regularity. This provides the experimenter with an effective tool to assess whether somatic and neurobehavioral development is modified by prenatal and/or early postnatal treatment.

### General Indications

Any assessment of early treatment effects, such as those illustrated in this and subsequent sections, should attempt to minimize biases in design and data analysis (to be mentioned later). It should collect adequate information both on maternal toxicity (including reproductive end points) and on somatic end points of offspring development, in addition to those included in test batteries for postnatal neurobehavioral assessment (22).

Possible confounding variables must be adequately controlled. As concerns effects on the dam, for example, the experimenter can obtain useful information by simple measurements of food and water consumption in addition to measurements of body weight, as has been shown by recent data on gestational ozone (O<sub>3</sub>) exposure (40); subsequently, pair-fed (yoked) control groups may have to be used if a treatment has marked effects on maternal food or water intake.

In addition, cross-fostering procedures aimed at controlling for postnatal maternal effects are a *qua non* after prenatal treatments (6,35,41,42). Most studies use unidirectional fostering of both control and treated litters to untreated dams, eliminating postnatal maternal effects such as those produced by changes in maternal functions (e.g., milk production) or behavior, but not allowing the experimenter to know whether or not these effects might have occurred. Some cues to understanding the loss of information that can occur with any cross-fostering procedure were obtained with a complex design in which prenatal benzodiazepine treatment was followed by nursing of control and treated litters by either their own dams, different dams of the same group, or dams of the other group [(43); see in (41) a response model analysis for cross-fostering studies based on comparisons between different fostering procedures in an experiment with prenatal MM treatments].

The most commonly employed end points and their respective developmental time tables are essentially the same in rats and in mice. The end points illustrated below have been used in mice by our group for several years and are based on the well-known Fox battery (44) with a few modifications and additions; equivalent illustrations concerning the rat can be

readily found in the literature (2,9). Experiments performed in mice indicate that the end points in question are adequate to reveal both acceleration and retardation as a result of early treatment conditions (45,46). Accelerated development of reflexes and responses has also been shown in rats after early treatments with antidepressants (47), cocaine (48), and naloxone (49).

In most behavioral teratology studies, the experimenter questions *a*) whether treated animals are significantly retarded relative to their appropriate controls; *b*) whether neurobehavioral retardation, if any, is specific or is associated with significant changes in somatic development (such as reduced weight at birth, slowing of postnatal weight gain, delayed ear and eye opening, or other indications of overt toxicity); and *c*) whether the retardation is selective or widespread, i.e., involving a limited spectrum of reflexes and responses or a wide variety of end points.

With respect to the question in *b*, study designs should allow calculation of dose-response functions. For example, a prenatal sulfur dioxide (SO<sub>2</sub>) inhalation study (50) has shown neurobehavioral impairment in mice at both 32 and 64 ppm, but only the latter concentration reduced pup weight at birth.

Most simple somatic and neurobehavioral end points lend themselves to ordinal scoring. To escape the problem of litter confounding, and to make an effective use of the available experimental animals when several tests are planned, only one animal (or at most one male and one female) from each litter is assigned to each test. With this procedure, the litter random factor and the subject random factor become one and the same thing, eliminating a dangerous bias that is difficult to handle in the case of nonparametric data and greatly simplifying data analysis (42).

### Testing Mice by the Fox Scale

This scale has been designed to include a number of end points that are representative of various components of neural and behavioral development in the first postnatal weeks, without imposing an excessive test burden on either the pup or the experimenter (44). The description given below represents a slightly modified version of the original scale to include some additional measures (strong and weak tactile stimulation tests) (35).

The most common procedure is to test each animal daily to avoid time lags in the

detection of a maturational event. However, little is known about the effects of different handling and testing burdens on subsequent maturation of the various reflexes and responses, although handling is recognized as a crucial variable.

In the list that follows, the postnatal ages at which a reflex or response first appears and subsequently shows complete maturation (i.e., adultlike characteristics) have been intentionally omitted [for representative illustrations of the absence or presence of a delay in postnatal neurobehavioral development after prenatal O<sub>3</sub> or benzodiazepine exposure, respectively, see (40,51)]. In fact, it appears essential that each experimenter establish and repeatedly verify control baselines specific for the conditions under which the work is conducted, including animal strain, types of control treatments (depending on exposure schedules), and scoring system adopted (see above). In addition to assessing the neurobehavioral end points indicated below, it is also essential to perform a parallel assessment of somatic developmental end points, including at least body weight gain and time of eye opening, ear opening, and incisor eruption. These neurobehavioral end points of the Fox Scale are as follows: *a*) righting reflex—pup returns to its feet when placed on its back; *b*) cliff aversion—pup withdraws from the edge of a flat surface when its snout and forepaws are placed over the cliff; *c*) forelimb and hindlimb stick grasp reflex—pup grasps the shaft of a toothpick when it is touched to the palm of each paw; *d*) vibrissa placing reflex—pup places its forepaw on a cotton swab stroked across its vibrissae; *e*) level or vertical screen test—pup holds onto a wire mesh (5×5 mm) when dragged across it horizontally or vertically by the tail; *f*) screen climbing test—pup climbs up the vertical screen using both fore- and hindpaws; *g*) pole grasping—pup grips a wooden pencil with its forepaws; *h*) auditory startle response—pup shows a whole-body startle response when a loud clap of the hands occurs less than 10 cm away; and *i*) strong and weak tactile stimulation tests—a head-turning response is triggered by the application of tactile stimuli (von Frey hairs of 0.35 or 0.05 g) in the perioral area on both sides of the head.

No attempt can be made here to review the ample literature concerning the effects of various early treatments on postnatal neurobehavioral development as assessed by functional observational batteries like the one illustrated above. One problem is that batteries with more than a few end

points, while yielding consistent results in the same laboratory over an extended period of time, have apparently not been included in interlaboratory comparisons.

As concerns other problems, one can mention the differences in treatment effects between strains of the same species. In one study, phenobarbital treatment on days 10 to 16 of gestation delayed postnatal sensory-motor development in mice of the DBA strain, whereas drug-exposed mice of the C57BL/6J strain were unaffected or slightly quicker in attaining mature responses in some tests (52). The considerable strain difference in response to treatment was also confirmed by the fact that only the DBA mice developed hyperactivity (up to three times the control level at 18 days); in addition, the profile of changes in uptake of neurotransmitters by cerebral cultures established with tissue removed shortly after the end of prenatal treatment differed markedly between the two strains.

### Other Tests

A variety of other tests that have been used to assess the pup's neurobehavioral capabilities at successive developmental stages cannot be described and evaluated here in any detail; these tests are often aimed at a more thorough evaluation of maturation phenomena that are subjected to quick and approximate assessment in batteries like the Fox scale. This applies, for example, to conventional assessments of motor coordination such as the rotarod test, which has a long tradition of use in pharmacology-toxicology, as well as to assessments of startle responses and startle habituation, which were included in the Collaborative Behavior Teratology Study (22) [for a review of reflexive measures, see (53)].

Other tests of quite different types have proven useful, such as the swimming test for which detailed normative data at successive ages are available (9), as well as various assessments of seizure sensitivity and thermoregulatory responses; the latter tests are clearly described in any manual of pharmacological methods. As concerns thermoregulation, it is worth underlining that the value of conventional temperature measurements, including responses to drugs that lower or raise body temperature, can acquire greater functional value when accompanied by simple tests of thermotactic behavior. For example, the finding of retarded development of thermoregulation after prenatal alcohol exposure (54) was strengthened by the finding

that alcohol-exposed pups moved closer to the warm end of a thermocline than did control pups; their body temperature, however, did not rise concurrently (55).

Still other tests cannot claim the renown created by extensive use, but two of them deserve a brief description because they are quite economical and are apparently able to provide information at least somewhat different from that obtained by the tests so far discussed.

**Suckling Test.** This procedure has been extensively used in rats for a variety of developmental psychopharmacological assessments (56); more recently, it has been used in mice in order to assess the effects of a previous NGF treatment, including those on psychopharmacological reactivity (57). The test can be performed at different ages after birth and uses the anesthetized pup's own dam in order to assess the responses to the suckling stimulus; these include attachment to the nipple and several other behaviors such as paddling with forelimbs, treading with hindlimbs, nipple shifting, and displacing a sibling from a nipple.

**Homing Test.** This test exploits the strong tendency of the immature pup to maintain body contact with the dam and the siblings, which requires adequate sensory (olfactory) and motor capabilities as well as the associative and discriminative capabilities that allow the pup to become imprinted by the mother's odor, to remember it, and to recognize it among others. In a version evolved for 10-day-old mice (51), the pup is placed at one end of a rectangular arena (36×22.5 cm) with a wire mesh floor and a goal area at the opposite end (14×22.5 cm). Shavings from the home litter are evenly spread under the floor of the latter area. The score is the time taken by the pup to place both forelimbs above the goal area. In the version most often used in rats, the pup is placed on the divide between the areas over home and clean bedding, and the time spent over each area is measured (58,59).

### Pain Reactivity and Analgesic Drug Effects

Appropriate adaptive responses to painful stimuli are a vital part of the organism's repertoire and lend themselves to assessment by simple and reliable methods such as the tail flick test, the hot plate test, and modified versions for neonates (e.g., the tail immersion test). In second-tier and higher behavioral teratology studies, if not in first-tier studies, any evaluation of treatment effects that does not include this area

of assessment must be considered as incomplete. This is an important notation since in some countries the use of any test involving even moderate pain and stress is subject to *ad hoc* authorization for ethical reasons; in other words, these tests cannot be performed unless they are supported by a strong rationale whose validity is acknowledged by the regulatory authority. On the other hand, appropriate versions of titration schedules, which are aimed at assessing aversive thresholds while the animal maintains control over the magnitude of the stimulus and therefore minimizes pain and stress (60,61), have apparently not been evolved for application to immature subjects.

The tail flick test measures the latency of the flick response to a thermal stimulus focused on the animal's tail. The test requires restraint of the animal, but such restraint need not be a drastic one.

The hot plate test usually measures the latency to the first paw-licking response after the rat or mouse is placed on a metal plate with temperature maintained at a constant level. The temperature used is generally around 55°C to ensure that all animals experience moderate pain, but not more intense pain.

The behavioral response measured in the hot plate test, being obviously more elaborate than the tail flick end point, is highly sensitive to a variety of influences, particularly experience factors. In any event, exposure to any painful stimulus, as shown by the literature, has both short- and long-term effects on responses to the same and to other painful stimuli. This is often in the direction of reduced pain reactivity, which, depending jointly on organismic, exposure, and time factors, can be mediated more by opioid or more by non-opioid (particularly cholinergic) mechanisms (see below for an example of differential early treatment effects on these two types of analgesia).

Therefore, experimental protocols, including sources of experimental animals, must be highly standardized to avoid a variety of possible biases such as those that may be produced by early rearing variables and other environmental influences besides the well-known effects of organismic variables. (In most analgesia tests, for example, male rats tend to be less reactive than females to pain and to display significantly greater magnitudes of morphine analgesia.) In fact, pain reactivity (including analgesic drug responses) shows considerable developmental plasticity, as is shown by the long-term changes produced not only by prenatal and postnatal stress (62,63) but also by

nonstressful manipulations, such as variation of litter gender composition (64,65).

The interest for behavioral toxicology and teratology of these plasticity phenomena has been considerably enhanced by recent information concerning underlying mechanisms. In fact, hyperalgesia has been shown both in NGF-treated adult mice (66) and in transgenic mice that make excess NGF and have overgrown sympathetic neurons that make contact with sensory neurons [Davis et al., unpublished data; (67)]. Therefore, changes in the dynamics of growth factors should be considered when a toxicological profile includes modifications in pain sensitivity.

Without attempting a review of the literature on the effects of early treatments, it must be mentioned that several experiments have shown short- and long-term effects of prenatal and postnatal treatments with opiate drugs (particularly morphine and methadone) on pain reactivity and responses to analgesics, including opposite effects on sensitivity to morphine (68-71). One of these studies has documented an interaction in rats between treatment and sex (that is, morphine hypersensitivity in males and hyposensitivity in females) with tests performed an extended period of time after the termination of morphine treatment on postnatal days 1 to 7. Moreover, the specificity of these effects was supported by the finding that both opioid stress analgesia after intermittent cold-water swims and nonopioid analgesia after continuous swims were little affected by the early treatment (70).

The literature on other potential behavioral teratogens is more scattered but contains some interesting indications; for example, the analgesic effect of morphine was found to be enhanced by exposure to lead throughout pregnancy and lactation (72). On the other hand, as emphasized in a previous section, negative data like those concerning the development of morphine analgesia after prenatal benzodiazepine treatment can provide a useful contrast when attempting to understand the mechanisms of production of other treatment effects (35).

### Spontaneous Motor Activities

Innumerable combinations can be found in the literature of activity test types and subtypes, test conditions, test schedules, and other local paraphernalia that are treasured in any respectable behavioral laboratory. A considerable portion of these test situations originally evolved to meet the need

for neuropsychopharmacological assessments (73,74), offering an ample choice when behavioral teratology assessments were started in relatively recent times.

On the other hand, as shown by the representative examples given in a previous section, the profile of early treatment effects on motor activity can show considerable variation as a function of test type and test conditions. Moreover, treatment-test interactions have a perverse tendency to escalate into higher order interactions whenever other variables are also considered, e.g., animal strain, sex, and age of testing, to mention only some of the more important ones.

The aim of the present section is therefore to provide basic information concerning activity tests that have proved effective in detecting developmental effects of toxicological interest as well as testing strategies that are likely to minimize the risk of both false positives and false negatives.

### General Indications for the Use of Activity Tests

As already mentioned, the range of available motor activity tests is so wide that it would be arbitrary to propose a hierarchy of cost-effectiveness ratios aimed at orienting test choice. Some general indications, however, can be given here concerning either requirements that need to be fulfilled or test conditions from among which the experimenter must make choices related to research goals and available resources.

First, an obvious requirement is the reproducibility of results both within laboratories (hence the importance of building a database of historical controls that serves to locate anomalies in individual experiments) and between laboratories. Experimenter biases are more easily minimized by the use of automated versions of activity tests, several of which are commercially available. On the other hand, some advantages of nonautomated tests, especially those designed to allow the scoring of several responses from videorecordings, can be more pronounced in activity assessments than in other types of assessments. This applies in particular to treatments that tend to produce a mixed bag of response enhancements (stimulation) and response reductions (depression), i.e., effect profiles not easily assimilable to those of reference agents, at least when the trend of dose-response functions is considered.

Second, the choice of appropriate ages for activity testing must be based on the normal developmental pattern, which includes widely differing activity profiles at

successive ages. In altricial species like the rat and the mouse, whose offspring are highly immature at birth, the activity level is low for 10 or more postnatal days and then increases rapidly around or shortly after the end of the second postnatal week (i.e., in relation with eye and ear opening). The typical adultlike habituation pattern—namely, the response reduction either at successive times during a single test of at least 30 min duration (within-session habituation) or in successive tests (between-session habituation)—emerges only several days later (at about the time of weaning around the end of the second postnatal week or even after one or two additional weeks, depending on the strain or the test used). Since the phenomena just mentioned provide useful information on the development of several important regulatory mechanisms (monoaminergic, cholinergic, etc; see the example of prenatal benzodiazepine effects in the section “Inference Strategies, Research Goals, and Resources”), activity should be tested at three successive ages: shortly after the activity increase related to eye and ear opening, shortly after the time when controls first show a typical adultlike habituation pattern, and at the young adult stage, i.e., after sexual maturation. The use of the same animals at different ages is obviously more economical than the use of naive animals at each age. Test experience, as is well known, can influence subsequent performances; however, data suitable to determining whether early treatment effects can be substantially affected by repeated testing are apparently not available.

Third, a choice must be made of testing time relative to the circadian illumination cycle. Both rats and mice are predominantly nocturnal species and therefore show higher activity levels during periods of darkness than during light periods; in relation to its smaller body size, however, the mouse has more frequent bouts of activity, feeding, and drinking during light hours than the rat. Experiments are generally performed during regular working hours, i.e., during the light period when the normal light-dark cycle is used in animal and test rooms (which is apparently the case in the majority of published experiments), or, vice versa, during the dark period if the cycle is reversed. Direct comparisons between the effects of treatments of behavioral teratological interest under the two conditions have apparently not been performed. According to our experience, reversed cycle conditions, which allow us to test animals during the dark period without requiring a reversal of

the experimenters' circadian cycle, are preferable to other conditions. During this period, small rodents show higher baselines of overall activity and a fuller display of various components of their repertoire, which improves the quality of the data when the scoring of more than one or two responses is desirable.

In this context, it is recommended that at least an approximate balance be achieved between experimental groups in the assignment of animals to different times of testing within the chosen period. In fact, activity levels tend to fluctuate within both the light and dark periods; bouts of high activity, for example, are more frequent during the initial than during the later portions of the dark period, which can fake stimulating or depressant effects of a treatment if balancing for time of testing is not performed.

Fourth, activity profiles and treatment effects thereon can be influenced by the greater or lesser familiarity of the test environment, if nothing else because of the effects of habituation and different amounts of stress involved (for some examples drawn from behavioral teratology studies, see the section “Inference Strategies, Research Goals, and Resources”). At the two opposite extremes, a home cage test provides the most familiar environment, and an exposed test area such as the open field arena provides the most unfamiliar environment. Any test, however, can be used both ways; the open field, for example, can become a highly familiar environment after prolonged or repeated exposure, while the home cage can be turned into an unfamiliar environment when the rat or mouse is shifted from a social to an isolated condition (as is usual) and the cage is freshly cleaned and furnished with clean bedding.

Most activity data are parametric and therefore amenable to statistical analyses that range from comparisons between two groups by the Student's *t*-test to analysis of variance (ANOVA) designs of increasing complexity (75). Under the condition of having access to expert statistical advice and appropriate software, mixed-model ANOVA designs have special value in the analysis of parametric data from behavioral teratology experiments, not being subject to the limitations of nonparametric analyses. Provided that they are balanced (since unbalancing imposes the use of special, burdensome procedures), mixed-model ANOVA designs can cope with any combination of fixed factors (such as treatment, sex, etc.) and random factors (such as litters within each treatment-test age combination

and subjects within each final group undergoing repeated testing). Without having to bother with the algebra and computational procedures, an experimenter needs only limited training to grasp the meaning of various effects and interactions and to learn to locate their sources by the minimal number of between-group comparisons within logical sets of means.

Clarity of statistical inferences and minimization of both type I and type II errors (false positives and false negatives, respectively) are thus achieved much more effectively than by the repetitive application of apparently simpler procedures to large databases with many group and subgroup means. The frequency of false negatives, which can be quite costly and highly misleading in experiments of long duration, as is the case in behavioral teratology, can be further reduced by the adoption of a post hoc test such as Tukey's test, whose application is permissible (or even recommended) in the absence of significant ANOVA results (76). In fact, *F* tests on main effects and interactions can often be blind to meaningful deviations of a limited number of means from the overall average within a logical set of means; therefore, the experimenter, who is well acquainted with the nature of the data, rather than the statistics advisor, must learn to single out instances in which negative ANOVA results are highly likely to constitute type II errors and to identify significant between-group differences by appropriate post hocs.

In the case of multiple response variables, separate univariate analyses performed on individual end points are indicated and mostly sufficient to characterize the effect profile. However, the proliferation of univariate analyses on data from the same experiment with several response end points increases the risk of spurious statistical significance (i.e., false positives); therefore, one should discount isolated instances of ANOVA *F*s and post hoc test values above the *p* cut-off point chosen *a priori*. Specifically, the expectation is to obtain an average of one such value out of 20 tests, in the absence of any real effect, if the cut-off point is *p* = 0.05.

Given the scope of the present review, no attempt can be made to discuss when and how to escalate from univariate to multivariate analyses in experiments with multiple response end points. The experimenter performing first-tier risk assessments must know that the data obtained in these studies can gain considerable additional value when the statistical advisor is able to

single out databases whose features make multivariate analysis worth the additional burden (to be borne by the advisor, of course, not by the experimenter). In fact, the art of characterizing and comparing effect profiles produced by behavioral teratology studies is still in its infancy. Therefore, the experimenter involved in first-tier risk assessments on poorly known agents is an ideal provider of new and original databases that can serve to verify the value of alternative multivariate models for the purpose of subsequent standardization.

### Open Field

The open field is a typical all-purpose observational test, which imposes a considerable workload; therefore its cost-effectiveness ratio depends jointly on labor cost and the value attached to information provided by multiple response end points. The test is performed in a circular or square arena with a washable floor that needs to be thoroughly cleaned after each test. The size of the arena must be adjusted to both animal species and age. In the case of mice, for example, we now use a square (40 × 40 cm) arena for 10- to 30-day-old subjects (77). In the case of rats, a typical size is 60 × 60 cm, which has proved suitable for testing animals of different ages (78).

Whatever species and test conditions are used, open field tests must be performed under closely controlled conditions of illumination, background noise, and layout of landmarks outside the arena (e.g., the room furniture, any other object of more than minimal size, and the experimenter's location). As in any observational test, experimenter bias in the scoring of various responses must be closely controlled whether the scoring is made directly during the test or later on, from video-recordings. The ideal situation is to have two observers, both blind to the animal's treatment condition (although unavoidably not to other conditions such as age and sex) and to the fellow observer's scoring. In research teams with adequate experience, these precautions can be somewhat relaxed if the reliability of each observer is verified by periodic checks, but blindness to treatments remains essential. Blindness, however, may prove impossible when early treatments produce stunting or other visible effects; the same applies to pretest pharmacological challenges (drug vs vehicle), since agents producing typical behavioral effects must be used.

One measurement that is always taken in open field tests is that of locomotor

activity (deambulation); squares are drawn on the floor (of 8 × 8 cm and 12 × 12 cm, respectively, in the mouse and the rat arenas mentioned above), and the number of crossings between adjacent squares are counted. The assessment of activity trends as a function of prolonged or repeated exposure (generally in the direction of response decrements, or habituation) is at least as important as that of overall activity levels, particularly when comparing successive developmental phases. In fact, as already mentioned, trends over age of overall activity and of habituation are ontogenetically dissociated from each other. In this respect, it should be noted that the two types of habituation—namely, that occurring during the same extended exposure (within-session habituation) and that occurring between successive exposures, generally at 24-hr intervals (between-session habituation)—are not identical phenomena and are likely to be served by neural mechanisms that are at least partly separate from each other. The strongest evidence on this point has been provided by ontogenetic dissociations consisting of age-dependent differences both in habituation trends and in *d*-amphetamine and scopolamine effects, depending on whether rats were given a single 30-min open field test or three 5-min tests at 24-hr intervals, with all other conditions being equal, including the use of different animals for tests at different ages (78).

As concerns the scoring of responses other than locomotor activity and the use of multiple response data, the list of potential end points is a fairly long one, including locomotion (crossings between squares), rearing (wall rearing and nonsupported), grooming, exploratory sniffing, freezing, time spent lying still (without freezing), time spent in contact or not in contact with the walls, various stereotypes (i.e., increase to above a zero or very low control level of acts such as face washing, gnawing, circling, jumping, head scanning, and focused sniffing), digging and push digging (if the floor is covered with bedding material), urination, and defecation.

If responses are scored during the test, no more than three to four of them should be included in the protocol to minimize the risk of observer errors; the suggestion is to focus the attention on locomotion, rearing, grooming, and stereotypes that can reveal noxious effects of the treatment under study or must be scored as part of a pharmacological reactivity assessment [see Laviola et al. (79) for normative data on *d*-amphetamine effects in mice at different

developmental stages]. In fact, these end points generally suffice to verify whether a treatment produces qualitatively or quantitatively different effects depending on the response variable.

If the scoring is performed on video-recordings, any number of responses can be considered. A high number of end points however, can create considerable problems in both data analysis (by increasing the burden as well as the risk of spurious statistical significance) and data interpretation (for the exploitation of databases in subsequent multivariate analyses, see the previous section). In first-tier analyses, it is preferable to limit the number of response variables to the minimum that appears to be necessary to characterize a profile of mixed stimulant and depressant effects, i.e., to avoid its confounding with either overall depression or other mixed profiles, including those which are typical of psychostimulants such as amphetamine and cocaine. In fact, uniform enhancement of several responses that are partly or totally incompatible with each other (so-called response competition) cannot occur, while the specific features of a mixed profile can provide significant information on the nature of the changes produced by a treatment. For example, a high frequency of grooming and stereotyped movements accompanying a reduction of locomotion and rearing can point to the need for biochemical and pathological investigations aimed at identifying the nature and location of subcortical changes.

### Activity in Closed Test Environments

Many tests are in use for the assessment of activity in closed environments. These range from a thoroughly familiar home cage with the animal's own soiled bedding (but generally with a shift from a grouped to an isolated condition for the time of the test) to a variety of less familiar situations, e.g., cages and boxes of quite different shapes and sizes. No particular version of this type of test can be considered superior to others so as to deserve to be recommended and described in detail.

Most of the tests we considered are used with automated recording of responses that exploits physical events such as movement of the floor, vibration of the cage or box, interruption of photocell beams, and alteration of low energy radio frequencies. In most types of commercial and noncommercial automated apparatus, appropriate adjustment of the location and sensitivity of the sensors can allow the experimenter to record two types of movements—i.e., one

in the horizontal plane (locomotion) and the other in the vertical plane (generally rearing)—and also to pick up gross neurological symptoms such as tremor. In many instances, this arrangement can suffice to reduce bias in the assessment of a variety of treatment effects on activity that either can consist of different (or even opposite) changes in the two types of responses or can be secondary to neurological manifestations. The risk of confounding between quite different behavioral syndromes, however, is not entirely eliminated. The classic example in psychopharmacology is that of the similar scores provided by an automated apparatus that measures only ambulation and rearing after treatment with neuroleptics, various sedatives, tremorogenic agents, various kinds of cholinergic agonists, and high (stereotypy-inducing) doses of monoaminergic stimulants such as amphetamine.

Therefore, observational versions of the tests may have to be considered when it appears desirable to analyze treatment effects on a wider range of responses; this is best achieved by scoring from videorecordings taken through a transparent wall of the box or cage. Observation is also needed when testing of nonisolated animals is desirable to characterize the social interaction components of the repertoire during development, particularly after early exposure to agents whose effects on social responses are documented in the adult (80).

Activity assessments are also performed in test environments with greater complexity and more elaborate scoring systems that consider both different responses and different sequences of the same response. For example, extensive use has been made of the Figure-8 maze, which consists of several interconnected alleys forming the figure and intended to mimic the burrows of the rat's natural habitat (81). This test environment, which tends to elicit high levels of spontaneous motor activity, has been exploited in several studies, including the Collaborative Behavior Teratology Study (22). In the Collaborative Behavior Teratology Study, the profiles of prenatal treatment effects on activity were assessed by testing at different ages and for different durations (including responses to an amphetamine challenge); these profiles have shown considerable variation among different laboratories, particularly in prenatal *d*-amphetamine treatment conditions. Comparable attempts to analyze interlaboratory variation in the results of other activity tests are not available; therefore, the variability of Figure-8 maze results in the study just mentioned should

not be construed as an indication against the further use of this test.

Other tests also exploit complex environments, as is the case with the hole-board test, which is aimed at separating activity (mainly locomotion) from exploration (dipping the head into the holes drilled in the cage floor). This and other similar distinctions are essentially anthropomorphic ones and fail to provide adequate cues to the understanding of response-related differences in treatment effects, which need to be analyzed and interpreted by more objective criteria.

### Representative Treatment Effects and Interactions

This section provides a selection of representative treatment effects and interactions (with test factors per se on one side and organismic factors such as animal strain, sex, and age of testing on the other), including information on responses to drug challenges after treatments at early developmental stages (35,82–85).

A preliminary consideration is that, before any behavioral teratology study, developmental trends of activity and responses to drug challenges need to be assessed and verified for replicability more thoroughly than other end points that have a lesser tendency to vary, like those illustrated in the section on postnatal sensory and motor development. Later on, repeated cross-checks between controls in a particular experiment and in previous experiments (historical controls) are equally essential. In fact, several data document both reassuring analogies between developmental trends in the presence of potentially strong sources of variation and unexpected deviations from apparently well-established developmental profiles.

At one extreme, for example, five inbred mouse strains with quite different activity profiles were shown to be highly uniform with respect to the appearance of antimuscarinic (scopolamine) hyperactivity at the end of the third postnatal week (86). In my group's experience, some phenomena, like the simultaneous appearance in mice of habituation and scopolamine hyperactivity have been observed over and over again, under both the same and quite different conditions, such as repeated brief exposures to the open field (yielding between-session habituation) and a single extended exposure to a house cage made unfamiliar by clean bedding (yielding within-session habituation).



At the opposite extreme, the experimenter can be faced with unexpected dissociations between developmental events that in several previous experiments had always been closely related to each other in spite of differences in species and test conditions. For example, in an experiment using different test schedules in the same open field, rat pups exposed to a single 30-min session showed only minimal within-session habituation at 3 weeks and full-fledged habituation at 4 weeks; at the latter time, however, scopolamine was still without effect and full-fledged drug hyperactivity was found only 2 weeks later. Even more surprisingly, exposure to three brief (5-min) sessions at 24-hr intervals allowed a near-complete between-session habituation to appear at 3 weeks, but no scopolamine effect was seen at 3, 4, or 6 weeks (78).

Previous data had indicated that the maturation of scopolamine effects on other responses could occur much later than that of its effects on locomotor hyperactivity (87). However, the data now available on the age- and test-dependent effects of this and other drug challenges, particularly amphetamine (78), strengthen the notion that the response factor can account for only part of the observed variation in developmental activity profiles and drug reactivity.

This indicates two things: the maturation of a given neural mechanism can be revealed by the earliest age at which a stimulant or an antimuscarinic treatment produces marked locomotor hyperactivity, but functional reliance on the same mechanism for the modulation of the same response under different conditions (or of other responses) can develop only later. Ontogenetic dissociations of the two types of phenomena probably depend on the developmental pace of other interacting mechanisms whose role varies from one situation to the other.

This tentative model can help us understand the maturation phenomena exploited in behavioral teratology studies, which are related to specific new needs of the developing organism each time it moves on from one ecological niche to a new one (in the case of altricial rodents, for example, at the time of eye and ear opening and at the time of weaning). The methodological implications are not only for higher tier studies but also to first-tier risk assessments, for reasons of efficacy and economy. In fact, the experimenter should verify that the chosen combinations of activity test conditions and drug challenges are suitable for picking up critical developmental events at the time of

their earliest possible occurrence, such as an activity surge and an increase in sensitivity to monoaminergic psychostimulants around the end of the second postnatal week and the appearance of adultlike habituation and scopolamine hyperactivity about 1 week later. In fact, any treatment effect on the development of the neural mechanisms responsible for these changes could be missed if test conditions delay the onset of functional reliance on these mechanisms and the changes produced by the treatment under study tend to be attenuated as the animal grows older.

Another methodological implication involves some important gaps in our knowledge concerning early treatment effects on developmental trends of drug reactivity. Specifically, several behavioral teratology studies have verified that useful information can be obtained by the use of monoaminergic drugs (particularly amphetamine, apomorphine, and cocaine), cholinergic-muscarinic agonists and antagonists, opiate drugs, and benzodiazepine receptor ligands. By contrast, the properties of other classes of agents have not yet been effectively exploited; the most obvious example is represented by 5-HT agonists and antagonists, whose effects at successive developmental stages have been carefully studied (88,89). This may be due to the fact that the wide range of types and subtypes in these drug classes makes it difficult to effect a rational choice of a highly restricted number of drug challenges under the constraints of behavioral teratological assessments. The most obvious of these constraints is the need to have all challenges and appropriate controls represented within each litter in each treatment-test age combination, in order to avoid the confounding of treatment and litter factors.

Overall, the increasing diversification within all classes of model drugs awaits exploitation in a more thorough analysis of early treatment effects and mechanisms of action on the basis of normative data concerning psychopharmacological reactivity during development; the effectiveness of this analysis is shown, for example, by the different effects of selective  $\mu$ -,  $k$ -, and  $\delta$ -opioid agonists in neonatal rats (90). The nature of the interactions observed when studying early treatment effects on activity and drug reactivity can be further illustrated by the data concerning two well-known neurotoxins, MM and ethanol, and an environmental pollutant,  $O_3$ , which is devoid of major neurotoxic effects but can produce subtle and selective behavioral changes.

**Methylmercury (MM).** As early as 1978, Hughes and Sparber (91) showed that, in the absence of prenatal MM effects on operant response levels (essentially a measure of spontaneous activity) and on acquisition of an autoshape task, MM subjects showed an attenuation of the disrupting amphetamine effects on autoshaped behavior.

In a series of studies by Annau, Cuomo, and co-workers (92-94), rats were given 5 to 8 mg/kg MM (po) on day 8 or 15 of pregnancy; open field activity at various postnatal ages was apparently not affected. The role of test factors, however, is suggested by the results of another study in the same series (95), which showed hyperactivity at 4 to 15 days with a different test using a closed activity cage. More importantly, the former studies showed an enhancement of amphetamine hyperactivity in the open field and an enhancement of both apomorphine sniffing and apomorphine depression of activity in the MM offspring. These effects tended to become attenuated after the second or third postnatal week and were related to an increased density of dopamine receptors, whereas clonidine effects on activity and the profile of cortical 2-adrenoceptors were not modified.

In the Collaborative Behavior Teratology Study (22), involving six laboratories, there was more variation in the effects of prenatal MM (2 or 6 mg/kg po on gestational days 6-9) on Figure-8 activity, which was assessed at different ages and by tests with different durations, than in the effects on other end points, particularly auditory startle habituation. Nevertheless, there was an overall significance of two fairly consistent effects, namely, increased Figure-8 maze activity in 1-hr tests and enhancement of the hyperactivity response to an amphetamine challenge.

Overall, the profile of MM prenatal effects on activity of the offspring and response to amphetamine appears to be adequately verified across laboratories and tests and consistent with the observed changes in dopamine receptors.

**Ethanol.** The effects of early (particularly prenatal) ethanol exposure on subsequent activity have been extensively investigated and have yielded fairly consistent results. In fact, several studies have shown hyperactivity in the treated offspring at different ages and with different tests (96-101), although some negative findings have also been reported (29,102). Because the latter study (102) was aimed mainly at the assessment of behaviors other than

activity, separate and combined the effects of prenatal and postnatal exposure were assessed in artificially reared rat pups.

The nature of ethanol effects has also been extensively investigated by the use of drug challenges. In several of the experiments mentioned above (97–100), responses at successive postnatal ages to *d*-amphetamine, alpha-methyl-*p*-tyrosine, *p*-chlorophenylalanine, and methysergide were apparently not modified, whereas the appearance of physostigmine depression and scopolamine hyperactivity was markedly delayed. It was concluded that the basic change produced by prenatal ethanol exposure was a delayed maturation of those cholinergic–muscarinic mechanisms that are deemed to modulate activity levels by counterbalancing the excitatory function of other (mainly monoaminergic) mechanisms.

Other data, however, contradict at least in part the inferences just mentioned. In one study, for example, ethanol-treated offspring showed an enhanced amphetamine hyperactivity (103); in another study a similar sensitization was found with a methylphenidate challenge (104). In addition, a dose–response shift to the left for locomotor activity has been found in ethanol-treated rat offspring with apomorphine (105).

In the first of these studies (103), males at 4 weeks of age clearly showed the change in amphetamine reactivity but females did not; this is of considerable interest in the light of the growing evidence on early sex dimorphism of monoaminergic regulatory systems (79). Moreover, the same study achieved a separation of the effects of ethanol from those of an important confounding variable by the use of a yoked control group that received sucrose substituted isocalorically for ethanol. Specifically, the ethanol enhancement of amphetamine reactivity in 4-week-old males failed to occur in the yoked group, whereas both the ethanol-treated and the yoked animals of both sexes showed a reduced amphetamine response at 6 weeks of age.

Considering the extraordinary importance of adequately understanding the developmental effects of subteratogenic doses of ethanol, the data mentioned so far are of considerable interest. In spite of the numerous discrepancies whose sources have not yet been located, they provide a basis for planning further experiments aimed at identifying the combinations of organismic, treatment, and test factors that facilitate or prevent the behavioral expression of changes in regulatory mechanisms.

Some interesting indications are already available concerning possible interacting variables, particularly considering that the various activity tests involve different degrees of familiarity and different amounts of stress. For example, stress responses (grooming) were enhanced in ethanol-treated offspring after a forced swim but not after exposure to a less stressful environment. Moreover, when the less stressful exposure preceded the forced swim, the ethanol effect on swim-induced grooming disappeared; in addition, naloxone reduction of grooming was not modified (106).

**Ozone (O<sub>3</sub>).** Although O<sub>3</sub> is known mainly for its adverse effects on the respiratory system, it can also produce other types of somatic effects, e.g., on immune functions (107). Exposure to O<sub>3</sub> can cause behavioral disturbances in humans, such as impairment of vigilance level and decreased physical performance and can depress several responses in adult rats and mice, such as locomotion, feeding and drinking, and operant responding (40,108). Late-gestational exposure of rats to O<sub>3</sub> (pregnancy days 17–20, 1.0–1.5 ppm) has been shown to impair somatic and neurobehavioral development of the offspring, including a delay in the maturation of reflexes and responses such as righting, grooming, and locomotion in an open field (109); however, cross-fostering procedures aimed at ruling out postnatal maternal effects were not used in this study. A more recent experiment in mice, which included these procedures, failed to show any effect of mid- and late pregnancy O<sub>3</sub> exposure (days 7–17, 0.4–1.2 ppm) on either postnatal neurobehavioral development or activity, habituation, and response to amphetamine at 60 days; the responses were assessed by an automated unfamiliar-home-cage test suitable to yielding high initial activity scores followed by response reduction (40).

Subsequent studies in mice used more prolonged O<sub>3</sub> exposure, up to 0.6 ppm, from several days before the start of pregnancy until either day 17 of pregnancy (110) or weaning of the offspring 3 weeks after birth (77), with or without cross-fostering at birth, respectively. Both exposure schedules again failed to affect postnatal neurobehavioral development, in spite of a marked and long-lasting depression of postnatal body weight gain in the study with combined gestational and postnatal treatment. The offspring exposed prenatally but not postnatally also failed to show changes in young adult activity or habituation in the automated test mentioned above. However,

observational social interaction tests performed at 23 to 25 and 43 to 45 days on paired animals in the same test environment showed significant changes in response profiles; these consisted mainly of a reduction of locomotion and other exploratory movements (such as rearing and sniffing at the air, the walls, or the sawdust on the floor), which was paralleled by a significant increase in self-grooming. By contrast, several responses directed to the partner (such as sniffing, following, grooming, etc.) were not affected by prior O<sub>3</sub> exposure (110).

The effects of combined gestational and postnatal exposure on response profiles in two different observational tests and an automated test were more pronounced but equally selective (77). An open-field test (40×40 cm arena) at 24 days in O<sub>3</sub>-exposed mice showed an attenuation or elimination of sex differences, i.e., a reduction of the higher female scores to male levels in the case of rearing and a convergence of both female (lower) and male (higher) scores toward the overall average in the case of sniffing. The marked response changes produced by an antimuscarinic challenge (scopolamine 2 mg/kg) were not modified.

The second open-field test at 29 days, performed in a smaller arena (16×15 cm) using the mice that had received the saline (control injection) in the previous test, showed only an O<sub>3</sub> reduction in grooming. (Note that this effect, opposite to that observed after gestational but not postnatal exposure, was obtained in conditions quite different from those of the previous experiment—namely, in an open field rather than a closed environment, and in subjects with prior experience tested individually rather than in naive subjects tested in same-sex pairs.) The marked effects on various responses produced by a psychostimulant challenge (*d*-amphetamine, 3.3 mg/kg) were essentially unmodified. (Some isolated instances of statistical significance had to be discounted when taking into account the number of comparisons performed.) Treatment–sex interactions could not be adequately evaluated in this experiment since there were only four subjects in each combination of prior exposure, sex, and pretest treatment.

The third activity assessment exploited the final test exposure at 31 days of age on a conditioned place preference (CPP) schedule that included a single 3.3 mg/kg *d*-amphetamine treatment 2 days before; naive littermates of the subjects in the two previous tests were used, and activity was recorded by photocells located in the CPP

apparatus (essentially a rectangular 40×15 cm, 3-compartment open field). Ozone-treated mice showed a disappearance of the sex difference in overall activity, with the O<sub>3</sub> males' scores being significantly higher than those of the controls and similar to those of control and O<sub>3</sub> females, which did not differ from each other.

Overall, the effect profiles observed in the various situations with the different schedules of O<sub>3</sub> exposure appear to have the following methodological implications. First, these data, most of which essentially belong to first-tier risk assessments in the absence of previous appropriate studies, show that a considerable amount of information can be obtained by these assessments, i.e., prior to any higher tier study. The overall test burden was fairly large but quite reasonable relative to the burden created by the nature of the treatment (O<sub>3</sub> being a highly unstable molecule that requires considerable technical paraphernalia to maintain controlled concentrations), by the need to use different concentrations and durations of exposure, and by the maintenance of the animals.

Second, some significant O<sub>3</sub> effects on activity, such as the attenuation or elimination of sex differences, were picked up by quite different versions of the same test, namely, a conventional square open field with scoring from videorecordings of the responses of naive animals, and a rectangular 3-compartment open field with automated recording of the responses of experienced animals. (Of course, the use of the CPP apparatus is not recommended for routine activity testing, but in this experiment the recording of activity in the final session of the CPP schedule—clearly a higher tier test—provided additional data practically gratis).

Third, the failure to locate treatment–sex interactions in the second observational test warns against an excessive reduction of final group size. In fact, as is well known, the power of ANOVA *F* tests on interaction effects is much less than that of *F* tests on main effects; furthermore, post hoc on differences between the means of final groups cannot circumvent this obstacle when the number of degrees of freedom is strongly reduced.

Fourth, the negative results, including the absence of significant changes in reactivity to different drug challenges, apparently rule out major or widespread neuroteratogenic effects of the pollutant. On the

other hand, the selective response changes in different activity tests were suitable to generate specific working hypotheses concerning possible mechanisms of the treatment's action. Specifically, the attenuation of sex differences bears a strong similarity to that observed after developmental exposure to a wide variety of stressors, which suggests that O<sub>3</sub> may produce long-lasting neural and endocrine changes like those documented by studies on early stress effects (77,110); this in turn could account for selective medium- and long-term behavioral changes in the absence of direct neuroteratogenic effects.

The data available so far cannot indicate whether the changes produced by O<sub>3</sub> exposure resulted from a direct effect on the fetus and the newborn or were the consequence of somatic and behavioral changes in the dams; these data, however, suffice to reject the null (no effect) hypothesis and to characterize developmental O<sub>3</sub> exposure as a treatment with a possible risk of borderline neurobehavioral pathology. The significance of this inference obviously depends on both the sensitivity of developing humans relative to that of altricial rodent species and the number of human subjects exposed to concentrations of the same order of magnitude. As concerns the latter, a well-known WHO meta-analysis published in 1987 pointed to peak O<sub>3</sub> concentrations often exceeding 0.2 ppm in large, polluted urban areas and to lowest-observed-effect levels in the 0.08 to 0.2 ppm range (111). A more recent survey in the Paris area showed that 0.05 ppm can produce a significant increase in the frequency and severity of respiratory illnesses in children (112); however, information concerning neurobehavioral development of exposed children is apparently not yet available.

## Conclusion

This review has attempted to provide both general and specific indications for the effective use of economical test methods in behavioral teratology assessments; in parallel, an attempt has been made to discuss the nature of the information that can reasonably be expected when using one or the other test strategy.

Some tests, particularly functional observational batteries that can exploit the regularity in postnatal development of several reflexes and responses, have been standardized and validated to the point that

their use can be strongly recommended both in proactive studies performed prior to any human exposure and in reactive studies. Other tests, particularly those aimed at assessing motor/exploratory activities, have been and continue to be used in many different versions, often resulting in different profiles of treatment effects as a function of age at testing and other variables.

In spite of this variation of treatment effects, behavioral teratology is coming close to defining the conditions that are most suitable for risk assessment in first-tier proactive studies, with increasing control over possible biases and confounding variables and therefore a progressive abatement of the probability of both false negatives and false positives. In this context, differences in treatment effects as a function of other variables should not be discounted as pure noise or as a nuisance. In fact, they often provide useful preliminary information on treatments' mechanisms of action. In addition, they help in evaluating the functional significance of different CNS changes assessed by neurobiological methods (e.g., in receptor profile, enzyme activity, or neurotransmitter metabolism), which are otherwise difficult to interpret. This applies in particular to ontogenetic dissociations, i.e., differential treatment effects depending jointly on developmental stage at the time of exposure, age of testing, and response end point.

On the other hand, appropriate reactive studies appear to have special value when there is a risk that chemical exposures result in borderline pathology among large cohorts of developing human subjects in the absence of hard clinical and pathological evidence of neuroteratogenicity, as has been shown to be the case with low doses of lead and ethanol and with prenatal benzodiazepine exposure. Specifically, animal data can not only be a useful complement to epidemiological and clinical–psychological data that are already available, but can also contribute to the rational designing of further human studies. In fact, human studies have a long duration and a high cost, and experimenters must avoid making random choices among a wide variety of possible end points and confounding variables. These choices could be made more effectively by taking into account the indications provided by animal experiments concerning behavioral responses or repertoires and regulatory mechanisms that are affected or unaffected by a particular treatment.

## REFERENCES

- Coyle I, Wayner MJ, Singer G. Behavioral teratogenesis: a critical evaluation. *Pharmacol Biochem Behav* 4:191-200 (1976).
- Zbinden G. Experimental methods in behavioral teratology. *Arch Toxicol* 48:69-88 (1981).
- WHO-IPCS. Principles for Evaluating Health Risks to Progeny Associated with Exposure to Chemicals during Pregnancy. *Environmental Health Criteria* 30. Geneva:World Health Organization, 1984.
- Leonard BE. Behavioral teratology and toxicology. In: *Psychopharmacology 2, Part 1: Preclinical Psychopharmacology* (Grahame-Smith DG, ed). New York:Elsevier, 1985;239-282.
- Tucker JC. Benzodiazepines and the developing rat: a critical review. *Neurosci Biobehav Rev* 9:101-111 (1985).
- Annau Z, Eccles CU. Prenatal exposure. In: *Neurobehavioral Toxicology* (Annau Z, ed). Baltimore:The Johns Hopkins University Press, 1986;153-169.
- Ruppert PH. Postnatal exposure. In: *Neurobehavioral Toxicology* (Annau Z, ed). Baltimore:The Johns Hopkins University Press, 1986;170-189.
- Elia J, Katz IR, Simpson GM. Teratogenicity of psychotherapeutic medications. *Psychopharmacol Bull* 23:531-586 (1987).
- Vorhees CV. Reliability, sensitivity and validity of behavioral indices of neurotoxicity. *Neurotoxicol Teratol* 9:445-464 (1987).
- Sonderegger T, Zimmermann E, eds. Developmental effects of drug dependence. *Neurobehav Toxicol Teratol* 6:259-331 (1984).
- Cuomo V, Annau Z, eds. Perinatal actions of psychotropic drugs and environmental chemicals and their behavioral effects in later life. *Neurobehav Toxicol Teratol* 7:469-525 (1985).
- Cranmer JM, Tilson HA, eds. Neurotoxicology in the fetus and child. *Neurotoxicology* 7:1-669 (1986).
- Sonderegger TB, ed. Perinatal substance abuse: research and clinical perspectives. *Neurobehav Toxicol Teratol* 8:325-419 (1986).
- Sonderegger TB, ed. Developmental effects of drug dependence. *Neurotoxicol Teratol* 9:287-327 (1987).
- Boer GJ, Feenstra MGP, Mirmiran M, Swaab DF, Van Haaren F, eds. Progress in brain research: biochemical basis of functional neuroteratology. In: *Permanent Effects of Chemicals on the Developing Brain, Vol 73*. Amsterdam:Elsevier, 1988;1-565.
- Cranmer JM, Wiggins RC, eds. Drug abuse and brain development. *Neurotoxicology* 10:305-634 (1989).
- Hutchings DE, ed. Prenatal abuse of licit and illicit drugs. *Ann NY Acad Sci* 562:1-388 (1989).
- Rees DC, Francis EZ, Kimmel CA, eds. Qualitative and quantitative comparability of human and animal developmental neurotoxicity. *Neurotoxicol Teratol* 12:173-292 (1990).
- Mutti A, Costa LG, Manzo L, Cranmer JM, eds. Proceedings of the third meeting of the International Neurotoxicology Association. *Neurotoxicology* 13:1-308 (1992).
- Nelson BK. Origins of behavioral teratology and distinctions between research on pharmaceutical agents and environmental/industrial chemicals. *Neurotoxicol Teratol* 12:301-305 (1990).
- Spyker JM. Behavioral teratology and toxicology. In: *Behavioral Toxicology* (Weiss B, Laties VG, eds). New York:Plenum Press, 1975;311-349.
- Buelke-Sam J, Kimmel CA, Adams J, eds. Design considerations in screening for behavioral teratogens: results of the Collaborative Behavior Teratology Study. *Neurobehav Toxicol Teratol* 7:537-789 (1985).
- 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. *Neurotoxicol Teratol* 10:155-167 (1988).
- Barr HM, Streissguth AP, Darby BL, Sampson PD. Prenatal exposure to alcohol, caffeine, tobacco, and aspirin: effects on fine and gross motor performance in 4-year-old children. *Devel Psychol* 26:339-348 (1990).
- Fried PA, Watkinson B. 36- and 48-month neurobehavioral follow-up of children prenatally exposed to marijuana, cigarettes, and alcohol. *J Dev Behav Pediatr* 11:49-58 (1990).
- Needleman HL. The health effects of low level exposure to lead. *Annu Rev Public Health* 12:111-140 (1991).
- Driscol CD, Streissguth AP, Riley EP. Prenatal alcohol exposure: comparability of effects in humans and animal models. *Neurotoxicol Teratol* 12:231-237 (1990).
- Davis JM, Otto DA, Weil DE, Grant LD. The comparative developmental neurotoxicity of lead in humans and animals. *Neurotoxicol Teratol* 12:215-229 (1990).
- Nelson BK, Brightwell WS, Mackenzie-Taylor DR, Burg JR, Massari VJ. Neurochemical, but not behavioral, deviations in the offspring of rats following prenatal or paternal inhalation exposure to ethanol. *Neurotoxicol Teratol* 10:15-22 (1988).
- Stamper CR, Balduini W, Murphy SD, Costa LG. Behavioral and biochemical effects of postnatal parathion exposure in the rat. *Neurotoxicol Teratol* 10:261-266 (1988).
- Clemens GR, Hartnagel RE, Bare JJ, Thyssen JH. Teratological, neurochemical, and postnatal neurobehavioral assessment of METASYSTOX-R, an organophosphate pesticide in the rat. *Fundam Appl Toxicol* 14:131-143 (1990).
- Weisenburger WP, Minck DR, Acuff KD, Vorhees CV. Dose-response effects of prenatal phenytoin exposure in rats: effects on early locomotion, maze learning, and memory as a function of phenytoin-induced circling behavior. *Neurotoxicol Teratol* 12:145-152 (1990).
- Hartley P, Neill D, Hagler M, Kors D, Vogel G. Procedure- and age-dependent hyperactivity in a new animal model of endogenous depression. *Neurosci Biobehav Rev* 14:69-72 (1990).
- Dallemagne G, Weiss B. Altered adult behavior of mice following postnatal treatment with haloperidol. *Pharmacol Biochem Behav* 16:761-767 (1982).
- Bignami G, Alleva E, Chiarotti F, Laviola G. Selective changes in mouse behavioral development after prenatal benzodiazepine exposure: a progress report. *Prog Neuropsychopharmacol Biol Psychiatry* 16:587-604 (1992).
- Gruen RJ, Deutch AY, Roth RH. Perinatal diazepam exposure: alterations in exploratory behavior and mesolimbic dopamine turnover. *Pharmacol Biochem Behav* 36:169-175 (1990).
- Viggedal G, Hagberg BS, Laegreid L, Aronsson M. Mental development in late infancy after prenatal exposure to benzodiazepines—a prospective study. *J Child Psychol Psychiatry* 34:295-305 (1993).
- Noldus LPJJ. The Observer: a software system for collection and analysis of observational data. *Behav Res Methods Instrum Comput* 23:415-429 (1991).
- Weiskrantz L. Some traps and pontifications. In: *Analysis of Behavioral Change* (Weiskrantz L, ed). New York:Harper and Row, 1968;415-429.
- Bignami G, Musi B, Dell'Omo G, Laviola G, Alleva E. Limited effects of ozone exposure during pregnancy on physical and neurobehavioral development of CD-1 mice. *Toxicol Appl Pharmacol* 129:264-271 (1994).
- Spyker DA, Spyker JM. Response model analysis for cross-fostering studies: prenatal versus postnatal effects on offspring exposed to methylmercury dicyandiamide. *Toxicol Appl Pharmacol* 40:511-527 (1977).
- Chiarotti F, Alleva E, Bignami G. Problems of test choice and data analysis in behavioral teratology: the case of prenatal benzodiazepines. *Neurotoxicol Teratol* 9:179-186 (1987).
- Laviola G, Bignami G, Alleva E. Interacting effects of oxazepam in late pregnancy and fostering procedure on mouse maternal behavior. *Neurosci Biobehav Rev* 15:501-504 (1991).
- Fox WM. Reflex-ontogeny and behavioural development of the mouse. *Anim Behav* 13:234-241 (1965).

45. Calamandrei G, Alleva E. Epidermal growth factor has both growth-promoting and growth-inhibiting effects on physical and neurobehavioral development of neonatal mice. *Brain Res* 477:1-6 (1989).
46. Santucci D, Calamandrei G, Alleva E. Neonatal exposure to bFGF exerts NGF-like effects on mouse behavioral development. *Neurotoxicol Teratol* 15:131-137 (1993).
47. Drago F, Continella G, Alloro MC, Scapagnini U. Behavioral effects of perinatal administration of antidepressant drugs in the rat. *Neurobehav Toxicol Teratol* 7:493-497 (1985).
48. Sobrian SK, Burton LE, Robinson NL, Ashe WK, James H, Stokes DL, Turner LM. Neurobehavioral and immunological effects of prenatal cocaine exposure in rat. *Pharmacol Biochem Behav* 35:617-629 (1990).
49. Shepanek NA, Smith RF, Tyler ZE, Royall GD, Allen KS. Behavioral and neuroanatomical sequelae of prenatal naloxone administration in the rat. *Neurotoxicol Teratol* 11:441-446 (1989).
50. Singh J. Neonatal development altered by maternal sulfur dioxide exposure. *Neurotoxicology* 10:523-528 (1989).
51. Alleva E, Laviola G, Tirelli E, Bignami G. Short-, medium-, and long-term effects of prenatal oxazepam on neurobehavioural development of mice. *Psychopharmacology* 87:434-441 (1985).
52. Sedowofia SK, Innes J, Peter A, Alleva E, Manning A, Clayton RM. Differential effects of prenatal exposure to phenobarbital on the behaviour and neurochemistry of CBA and C57BL/6J mice. *Psychopharmacology* 97:123-130 (1989).
53. Fechter LD, Young JS. Reflexive measures. In: *Neurobehavioral Toxicology* (Annau Z, ed). Baltimore: The Johns Hopkins University Press, 1986;23-42.
54. Zimmerberg B, Ballard GA, Riley EP. The development of thermoregulation after prenatal exposure to alcohol in rats. *Psychopharmacology* 91:479-484 (1987).
55. Zimmerberg B, Beckstead JW, Riley EP. Prenatal alcohol exposure and thermotaxic behavior in neonatal rats. *Neurotoxicol Teratol* 9:283-286 (1987).
56. Ristine LA, Spear LP. Effects of serotonergic and cholinergic antagonists on suckling behavior of neonatal, infant, and weanling rat pups. *Behav Neural Biol* 41:99-126 (1984).
57. Calamandrei G, Valanzano A, Alleva E. NGF and cholinergic control of behavior: anticipation and enhancement of scopolamine effects in neonatal mice. *Dev Brain Res* 61:237-241 (1991).
58. Adams J, Buelke-Sam J, Kimmel CA, Nelson CJ, Reiter LW, Sobotka TJ, Tilson HA, Nelson BK. Collaborative Behavioral Teratology Study: protocol design and testing procedures. *Neurobehav Toxicol Teratol* 7: 579-586 (1985).
59. Schlumpf M, Ramseier H, Abriel H, Youmbi M, Baumann JB, Lichtensteiger W. Diazepam effects on the fetus. *Neurotoxicology* 10:501-516 (1989).
60. Weiss B, Laties VG. Fractional escape and avoidance on a titration schedule. *Science* 128:1575-1577 (1958).
61. Weiss B, Laties VG. Comparative pharmacology of drugs affecting behavior. *Fed Proc* 26:1146-1156 (1967).
62. Kinsley CH, Mann PE, Bridges RS. Prenatal stress alters morphine- and stress-induced analgesia in male and female rats. *Pharmacol Biochem Behav* 30:123-128 (1988).
63. Shimada C, Kurumiya S, Noguchi Y, Umamoto M. The effects of neonatal exposure to chronic footshock on pain-responsiveness and sensitivity to morphine after maturation in the rat. *Behav Brain Res* 36:105-111 (1990).
64. Alleva E, Caprioli A, Laviola G. Postnatal social environment affects morphine analgesia in male mice. *Physiol Behav* 36:779-781 (1986).
65. Laviola G, Loggi G. Sexual segregation in infancy and bi-directional benzodiazepine effects on hot-plate response and neophobia in adult mice. *Pharmacol Biochem Behav* 42:865-870 (1992).
66. Della Seta D, De Acetis L, Aloe L, Alleva E. NGF effects on hot plate behaviors in mice. *Pharmacol Biochem Behav* 49: 701-705 (1994).
67. Barinaga M. Watching the brain remake itself. *Science* 266:1475-1476 (1994).
68. Zagon IS, McLaughlin PJ. Perinatal exposure to methadone alters sensitivity to drugs in adult rats. *Neurobehav Toxicol Teratol* 6:319-323 (1984).
69. Alleva E, Laviola G. Short-term and delayed behavioral effects of pre- and post-weaning morphine in mice. *Pharmacol Biochem Behav* 26:539-542 (1987).
70. Arjune D, Bodnar RJ. Post-natal morphine differentially affects opiate and stress analgesia in adult rats. *Psychopharmacology* 98:512-517 (1989).
71. Enters EK, Guo H, Pandey U, Ko D, Robinson SE. The effect of prenatal methadone exposure on development and nociception during the early postnatal period of the rat. *Neurotoxicol Teratol* 13:161-166 (1991).
72. Baraldi M, Zanoli P, Borella P, Caselgrandi E, Petraglia F. Neurobehavioral and neurochemical abnormalities of pre- and postnatally lead-exposed rats: zinc, copper and calcium status. *Neurobehav Toxicol Teratol* 7:499-509 (1985).
73. Robbins TW. A critique of the methods available for the measurement of spontaneous motor activity. In: *Handbook of Psychopharmacology*, Vol 7 (Iversen LL, Iversen SD, Snyder SH, eds). New York: Plenum Press, 1977;37-77.
74. Munn NL. *Handbook of Psychological Research on the Rat. An Introduction to Animal Psychology*. Boston: Houghton Mifflin, 1950.
75. Winer BJ. *Statistical Principles in Experimental Design*, 2d ed. New York: McGraw-Hill, 1971.
76. Wilcoxon RR. *New Statistical Procedures for the Social Sciences*. Hillsdale, NJ: Erlbaum, 1987.
77. Dell'Omo G, Fiore M, Petrucci S, Alleva E, Bignami G. Neurobehavioral development of CD-1 mice after combined gestational and postnatal exposure to ozone. *Arch Toxicol* (in press).
78. Laviola G, Renna G, Bignami G, Cuomo V. Ontogenetic and pharmacological dissociation of various components of locomotor activity and habituation in the rat. *Int J Dev Neurosci* 6:431-438 (1988).
79. Laviola G, Dell'Omo G, Chiarotti F, Bignami G. *d*-Amphetamine conditioned place preference in developing mice: relations with changes in activity and stereotypes. *Behav Neurosci* 108:1-11 (1994).
80. Terranova ML, Alleva E, Laviola G. Affiliation and neophobia in developing mice prenatally exposed to oxazepam. *Behav Pharmacol* 5:52-60 (1994).
81. Ruppert PH, Dean KF, Reiter LW. Development of locomotor activity of rat pups in figure-eight mazes. *Dev Psychobiol* 18:247-260 (1985).
82. Zenick H. Use of pharmacological challenges to disclose neurobehavioral deficits. *Fed Proc* 42:3191-3195 (1983).
83. Walsh TJ, Tilson HA. The use of pharmacological challenges. In: *Neurobehavioral Toxicology* (Annau Z, ed). Baltimore: The Johns Hopkins University Press, 1986;244-267.
84. Hannigan JH, Blanchard BA. Commentary: psychopharmacological assessment in neurobehavioral teratology. *Neurotoxicol Teratol* 10:143-145 (1988).
85. Spear LP. Neurobehavioral assessment during the early postnatal period. *Neurotoxicol Teratol* 12:489-495 (1990).
86. Bond NW. Ontogeny of activity in inbred and random-bred strains of mice: effects of scopolamine. *Behav Neurosci* 102:887-893 (1988).
87. Bauer RH. Differential effects of *d*-amphetamine and scopolamine on the ontogeny of rearing. *Pharmacol Biochem Behav* 21:321-323 (1984).
88. Enters EK, Spear LP. Ontogenetic transitions in the psychopharmacological response to serotonergic manipulations. *Psychopharmacology* 96:161-168 (1988).
89. Kirstein CL, Spear LP. 5-HT<sub>1A</sub>, 5-HT<sub>1B</sub> and 5-HT<sub>2</sub> receptor agonists induce differential behavioral responses in neonatal rat pups. *Eur J Pharmacol* 150:339-345 (1988).
90. Jackson HC, Kitchen I. Behavioural effects of selective  $\mu$ -,  $k$ -, and  $\delta$ -opioid agonists in neonatal rats. *Psychopharmacology* 97:404-409 (1989).
91. Hughes JA, Sparber SB. *d*-Amphetamine unmasks postnatal consequences of exposure to methylmercury *in utero*: methods

- for studying behavioral teratogenesis. *Pharmacol Biochem Behav* 8:365-375 (1978).
92. Hughes JA, Annau Z. Postnatal behavioral effects in mice after prenatal exposure to methylmercury. *Pharmacol Biochem Behav* 4:385-391 (1976).
  93. Cuomo V, Ambrosi L, Annau Z, Cagiano R, Brunello N, Racagni G. Behavioural and neurochemical changes in offspring of rats exposed to methyl mercury during gestation. *Neurobehav Toxicol Teratol* 6:249-254 (1984).
  94. Cagiano R, De Salvia MA, Renna G, Tortella E, Braghiroli D, Parenti C, Zanoli P, Baraldi M, Annau Z, Cuomo V. Evidence that exposure to methyl mercury during gestation induces behavioral and neurochemical changes in offspring of rats. *Neurotoxicol Teratol* 12:23-28 (1990).
  95. Eccles CU, Annau Z. Prenatal methyl mercury exposure. I: Alterations in neonatal activity. *Neurobehav Toxicol Teratol* 4:371-376 (1982).
  96. Abel EL, Dintcheff BA. Effects of prenatal alcohol exposure on behavior of aged rats. *Drug Alcohol Depend* 16:321-330 (1986).
  97. Bond NW. Prenatal ethanol exposure and hyperactivity in rats: effects of *d*-amphetamine and alpha-methyl-*p*-tyrosine. *Neurobehav Toxicol Teratol* 7:461-467 (1985).
  98. Bond NW. Prenatal alcohol exposure and offspring hyperactivity: effects of scopolamine and methylscopolamine. *Neurobehav Toxicol Teratol* 8:287-292 (1986).
  99. Bond NW. Prenatal alcohol exposure and offspring hyperactivity: effects of *para*-chlorophenylalanine and methysergide. *Neurobehav Toxicol Teratol* 8:667-773 (1986).
  100. Bond NW. Prenatal alcohol exposure and offspring hyperactivity: effects of physostigmine and neostigmine. *Neurotoxicol Teratol* 10:59-63 (1988).
  101. Becker HC, Randall CL. Effects of prenatal ethanol exposure in C57BL mice on locomotor activity and passive avoidance behavior. *Psychopharmacology* 97:40-44 (1989).
  102. Wigal T, Amsel A. Behavioral and neuroanatomical effects of prenatal, postnatal, or combined exposure to ethanol in weanling rats. *Behav Neurosci* 104:116-126 (1990).
  103. Blanchard BA, Hannigan JH, Riley EP. Amphetamine-induced activity after fetal alcohol exposure and undernutrition in rats. *Neurotoxicol Teratol* 9:113-119 (1987).
  104. Means LW, Medlin CW, Hughes VD, Gray SL. Hyperresponsiveness to methylphenidate in rats following prenatal ethanol exposure. *Neurobehav Toxicol Teratol* 6:187-192 (1984).
  105. Hannigan JH, Blanchard BA, Horner MP, Riley EP, Pilati ML. Apomorphine-induced motor behavior in rats exposed prenatally to alcohol. *Neurotoxicol Teratol* 12:79-84 (1990).
  106. Hannigan JH, Blanchard BA, Riley EP. Altered grooming responses to stress in rats exposed prenatally to ethanol. *Behav Neural Biol* 47:173-185 (1987).
  107. Becker S, Quay J, Koren HS. Effects of ozone on immunoglobulin production by human B cells *in vitro*. *J Toxicol Environ Health* 34:353-356 (1991).
  108. Weiss B. Behavioral toxicology and environmental health science. *Am Psychol* 38:1174-1187 (1983).
  109. Kavlock R, Meyer E, Grabowski CT. Studies on the developmental toxicity of ozone: postnatal effects. *Toxicol Lett* 5:3-9 (1980).
  110. Petruzzi S, Fiore M, Dell'Omo G, Bignami G, Alleva E. Medium and long-term behavioral effects in mice of extended gestational exposure to ozone. *Neurotoxicol Teratol* 17:463-470 (1995).
  111. WHO. Air Quality Guidelines for Europe. WHO Regional Publications, European Ser 23. Copenhagen:World Health Organization, 1987.
  112. Boissavy-Vinau M. L'ozone des villes et l'ozone des champs. *Recherche* 26:86-87 (1995).