SYMPOSIUM OVERVIEW

Biologically Based, Quantitative Risk Assessment of Neurotoxicants

WILLIAM SLIKKER, JR.,* KENNY S. CRUMP,† MELVIN E. ANDERSEN,‡ AND DAVID BELLINGER§

*Division of Neurotoxicology, National Center for Toxicological Research/Food and Drug Administration, Jefferson, Arkansas 72079; †Kaes Crump, ICF Kaiser Engineers, Ruston, Louisiana 71270; ‡ICF Kaiser, Crump Division, Morrisville, North Carolina 27560; and §Neuroepidemiology Unit, Children’s Hospital, Harvard Medical School, Boston, Massachusetts 02115

Received June 13, 1995; accepted July 5, 1995

The need for biologically based, quantitative risk assessment procedures for noncancer endpoints such as neurotoxicity has been discussed in reports by the United States Congress (Office of Technology Assessment, OTA), National Research Council (NRC), and a federal coordinating council. According to OTA, current attention and resources allocated to health risk assessment research are inadequate and not commensurate with its impact on public health and the economy. Methods to include continuous rather than dichotomous data for neurotoxicity endpoints, biomarkers of exposure and effects, and pharmacokinetic and mechanistic data have been proposed for neurotoxicity risk assessment but require further review and validation before acceptance. The purpose of this symposium was to examine procedures to enhance the risk assessment process for neurotoxicants and to discuss techniques to make the process more quantitative. Accordingly, a review of the currently used safety factor risk assessment approach for neurotoxicants is provided along with specific examples of how this process may be enhanced with the use of the benchmark dose approach. The importance of including physiologically based pharmacokinetic data in the risk assessment process and specific examples of this approach is presented for neurotoxicants. The role of biomarkers of exposure and effect and mechanistic information in the risk assessment process are also addressed. Finally, quantitative approaches with the use of continuous neurotoxicity data are demonstrated and the outcomes compared to those generated by currently used risk assessment procedures. © 1996 Society of Toxicology.

Introduction (William Slikker, Jr.)

Over the past 5 years, significant progress has been made in defining the principles of neurotoxicity risk assessment and the limitations of currently used procedures (OTA, 1990; NRC, 1992; Reiter et al., 1994). Recently, guidelines for defining neurotoxicity in the context of traditional qualitative, noncancer endpoint procedures (no-observed-adverse-effect level, NOAEL, or lowest-observed-adverse-effect level, LOAEL) have been proposed (Toxicological Principles, 1993; Tilson et al., 1995). These approaches to risk assessment provide yes/no decision-making capability but do not provide ample guidelines on quantitative procedures. As our nation moves toward the 21st century, the time to ban or not to ban is past and the need for quantitative risk assessment procedures is ever increasing. At what level of exposure to an agent is an individual at risk and what is that risk, 1 in 1000 or 1 in 1,000,000? This symposium was developed to address these quantitative approaches to neurotoxicity risk assessment and to continue the discussion on the validation of risk assessment procedures.

Four primary topics focusing on quantitative neurotoxicity risk assessment are presented including currently used and benchmark approaches of noncarcinogen risk assessment, the role of physiologically based pharmacokinetic modeling in the risk assessment process, the use of biomarkers and mechanistic data, and the use of continuous data in biologically based quantitative risk assessment procedures.

Currently Used Risk Assessment Procedures for Neurotoxicants

The most frequently used approach for risk assessment of neurotoxicants and other noncancer endpoints is the uncertainty or safety factor approach (Barnes and Dourson, 1988; Kimmel, 1990). Within the Environmental Protection Agency (EPA), for example, this approach involves the determination of reference doses (RfDs) by dividing a NOAEL by uncertainty factors that presumably account for interspecies and intraspecies differences in sensitivity or for extrapolation from subchronic to chronic exposures (Barnes and Dourson, 1988). Generally, an uncertainty factor of 10 is used to allow for the presumed greater sensitivity in humans than in animals and another uncertainty factor of 10 is used to allow for variability in sensitivity among humans. In this case, the RfD is equal to the NOAEL divided by 100. If the
NOAEL cannot be established, it is replaced by the LOAEL in the RfD calculation and an additional uncertainty factor of 10 is generally introduced (i.e., the RfD equals the LOAEL divided by 1000).

**Limitations and assumptions.** There are several features of this RfD or safety factor approach which deserve consideration. First, the method assumes a theoretical threshold dose below which no biological effects of any type are observed in a heterogeneous population. Not only is the determination of a threshold dose influenced by the sensitivity of the analytical methods employed, but also the theoretical bases of a threshold dose may be questioned. Unfortunately, less sensitive experiments can result in higher RfDs. If, due to normal variation in cellular function, an adverse effect can occur in untreated control subjects, then endogenous or exogenous factors may already be supplying a stimulus which is equivalent to a dose above the threshold dose. If exposure to an agent augments this stimulus, then an additional risk is expected and no threshold dose exists for that agent (Gaylor and Slikker, 1990). Second, the RfD approach relies on a single experimental observation (the NOAEL or LOAEL), instead of using complete dose–response curve data in the calculation of risk estimates. Because chemical interactions with biological systems are often specific, stereoselective, and/or saturable, a chemical’s dose–response curve may not be linear. Third, the magnitudes of the safety factors used to determine RfDs [interspecies extrapolation (10) and intraspecies extrapolation (10)] are based more on best estimates than actual data (Reiter et al., 1994; Slikker et al., 1992). Even though the RfD or safety factor approach has limitations, it will, by necessity, continue to be used until more quantitative risk assessment procedures have been developed and validated.

**Use of the Benchmark Dose Methodology in Determining Acceptable Exposure Levels (K. S. Crump)**

To set an acceptable level of a noncarcinogen in food, water, or the air, regulatory agencies often determine a NOAEL of exposure from toxicological or epidemiological studies and then divide the NOAEL by uncertainty or modifying factors that account for limitations in the data base or other uncertainties. A dose derived using the benchmark methodology (Crump, 1984; Gaylor and Slikker, 1990; Kodell and West, 1993; Crump, 1995) has been proposed as an alternative to the NOAEL in this procedure. The benchmark dose (BMD) is the dose that corresponds to a specified level of increased response (the benchmark response or BMR). A statistical lower confidence bound on the BMD (BMDL) has been specifically proposed as a replacement for the NOAEL in setting acceptable levels of human exposure.¹ The BMR is typically set at the lower end of the range of responses that can be detected experimentally, in order to avoid uncertainties associated with low-dose extrapolation using models that may not reflect biological realities.

Several potential advantages of the benchmark over the NOAEL have been identified (Crump, 1984):

1. The benchmark makes better use of dose–response information. Whereas the determination of a NOAEL generally involves comparing responses at discrete doses to responses in control subjects, the benchmark approach involves fitting a dose–response model to all of the dose–response data.
2. The benchmark reflects sample size more appropriately than a NOAEL. Smaller studies tend to result in smaller BMDLs, whereas the opposite is true for NOAELs.
3. Although the NOAEL is constrained to be one of the experimental doses, this is not the case with the benchmark approach. One consequence of this is that the benchmark approach is less likely to involve difficult and argumentative decisions that often arise concerning whether a particular experimental group defines a NOAEL.
4. A benchmark can be defined from a data set in which a NOAEL does not exist.
5. The calculation of a NOAEL generally utilizes data that are categorized into distinct dose groups. However, in some studies, including most epidemiological studies, categorization of subjects into dose groups is an arbitrary process. Such categorizations are not required in the benchmark approach.

The U.S. EPA is considering the adoption of the benchmark methodology for determining RfDs and reference air concentrations (RfCs) (EPA, 1990). A workshop convened by the EPA and the American Industrial Health Council (AIHC) endorsed the benchmark approach, but identified as an obstacle to implementing the approach the fact that essentially different quantitative methods must be applied to continuous and quantal data (AIHC, 1993).

Quantal data consist of “yes–no” information specifying whether a subject has an abnormal response and are modeled using a mathematical expression for the probability of an effect; i.e., the BMD is defined as the solution of the equation

$$P(BMD) - P_0 = BMR,$$

¹ This terminology differs from that in Crump, where the statistical lower bound was referred to as the BMD, rather than the BMDL.
² Although only additional risk, P(d)-P(0), is considered in this paper, the approach can easily be adapted to handle extra risk, defined as \([P(d)-P(0)]/[1-P(0)]\).
where $P_o$ is the probability of an abnormal response in an unexposed subject.

In contrast to quantal data, continuous data consist of measurements (such as organ weights, times, or ages) that can theoretically assume any value in a range. One approach for applying the benchmark approach to continuous data is to assume a particular distribution for the continuous response and to assume that the mean response is a function $\mu(d)$ of dose. Gaylor and Slikker (1990) and Crump (1995) discussed the application of the benchmark to continuous data. One proposed approach for continuous data is that the BMD be defined as the dose that gives a prescribed change in the mean response relative to the standard deviation $\sigma$, i.e., defining the BMD by

$$\pm [\mu(\text{BMD}) - \mu(0)]/\sigma = \text{BMR},$$

where BMR is specified in advance, and the plus (minus) sign is used if increases (decreases) in the mean response are considered detrimental. Crump (1995) discusses an approach for making BMDs comparable when they are calculated from both quantal and continuous data. Both quantal and continuous formulations require the specification of a dose–response model. It is recommended that the BMR be selected large enough so that the BMD is relatively insensitive to the choice of model. Allen et al. (1994) studied the relationship between the BMD and the NOAEL and the effect of various choices of the BMR and the dose–response model upon the BMD.

Prospects for Physiologically Based Pharmacokinetic (PB-PK) Modeling for Neurotoxicant Risk Assessments (M. E. Andersen)

Over the past 10–15 years there has been steady evolution and improvement of the scientific basis of cancer risk assessment. These advances are largely related to enhanced understanding of dose to target tissues and of the mode of action of particular carcinogens. Two of the more important extrapolations with carcinogens are high dose to low dose and interspecies. PB-PK models have been used to define the differences expected in tissue dose over a wide dose range and between species. Arguments related to knowledge of the mode of action tell us if the equivalent dose in a human is expected to have the same effect as it does in animals. By focusing on dose to tissue and mode of action, the various carcinogen assessment default procedures have been put into much sharper perspective.

With neurotoxicant risk assessment, responses are evaluated to develop NOAEL or LOAEL. The calculated value is then divided by a series of uncertainty factors to account for subchronic to chronic exposure extrapolations, interspecies differences, intraindividual differences, the robustness of the overall data base, and the presence or absence of a no-effect concentration among the test groups. While each factor has some qualitative basis, their application to individual compounds fails to take into account knowledge of tissue dosimetry or mode of action. Eventually, some of the advances in our treatment of carcinogens should permeate the noncancer risk assessment field and permit a more analytical approach to deriving human exposure standards from test results obtained in experimental animals. Perhaps, the first area in which mechanistic information can be introduced is related to improved definition of the concept of target tissue dose for these neurotoxicants.

Tissue dosimetry. PB-PK models realistically describe routes of administration, storage tissues, sites of metabolism/excretion, and tissue interactions. Organs are defined with respect to their volume, blood flow, and binding/metabolism of the toxic compound. Physiological, biochemical, and binding parameters are combined with mass-balance equations, one per tissue compartment, which can be solved to predict concentrations or tissue interactions in various sites within the body (Fig. 1). These models are readily scaled between animal species by changing the appropriate physiological and biochemical parameters in equations. PB-PK models for chemicals active in the nervous system have been designed to predict parent chemical concentration in the brain (for gaseous anesthetics), metabolite concentrations in peripheral nerves (for 2,5-hexanedione derived by oxidation from hexane), critical macromolecular adducts (for inhibition of cholinesterase by organophosphates), or reversible complexes with important nervous system receptors (for nicotine binding to cholinergic receptors in the nervous system).

Each of these examples is discussed briefly in the following sections.
Metabolism and toxicity. For systemic effects related to chemical administration, there is an interaction between the chemical and a biological substrate at the target tissue that initiates a series of events leading to alterations in biological function, i.e., an untoward or beneficial response. When assessing expected risks in people, the first step is to relate the appropriate measure of target tissue dose to the toxic response in the animal and predict the exposure conditions under which a similar dose would be achieved in exposed people. PB-PK modeling is used to evaluate the relationship between external exposure and target tissue dose for various exposure scenarios and various animal species. Classic examples here are related to those chemicals that are metabolized before they express any toxicity. Vinyl chloride (VC) was once proposed for use as a gaseous anesthetic. However, at concentrations well below its effective anesthetic range, it is converted to reactive metabolites that interact with DNA, form adducts, generate mutations, and eventually produce a rare tumor, hemangiosarcoma, in test animals and in workers exposed occupationally. The dependence of toxicity on metabolism also leads to potential complications when other substrates are present that compete with the toxic compound for a metabolizing enzyme. Vinylidene chloride (VDC) is highly toxic, following metabolism. Coexposures between VDC and trichloroethylene blocks VDC toxicity (Andersen et al., 1983). Risk assessments with these chloroethylenes require knowledge of the relation between amount metabolized and toxic outcome for different concentrations and in different animal species (cf. Reitz et al., 1995). These same considerations regarding metabolism also apply to some neurotoxicants.

Hexane. Hexane is relatively nontoxic but is metabolized to 2,5-hexanedione (2,5-HD) which causes peripheral neuropathies. 2,5-HD production requires consecutive oxidation of the parent alkane, proceeding from hexane to methyl-n-butyl ketone (M-n-BK) and to 2,5-HD. Both of these steps require the catalytic action of the same enzyme.

In early studies of the impairment in nerve conduction velocity by hexane exposure, Takeuchi et al. (1981) observed that exposure to 1000 ppm impaired nerve conduction more rapidly than exposure to 3000 ppm. A hint of the underlying basis of this effect was provided in dosimetry studies by Baker and Rickert (1981). They found that the disposition of 2,5-HD after inhalation exposure was concentration-dependent. At low inhaled hexane concentrations, there is an approximately linear relationship between hexane concentrations and blood 2,5-HD which is reduced and follows a complex time course.

A PB-PK model (Fig. 2) was used to evaluate the basis of these interactions in terms of the competition between hexane and M-n-BK for oxidation by cytochrome P450 (Andersen and Clewell, 1983). At concentrations below 1000 ppm, metabolism of both hexane and M-n-BK is proportional to their concentration. As concentrations increase metabolism becomes saturated and hexane reduces the rate of oxidation of M-n-BK during the exposure period. After exposure, hexane, which has lower blood solubility than M-n-BK, is rapidly eliminated by exhalation. The remaining M-n-BK becomes available for reaction and the 2,5-HD concentration increases even though the hexane exposure ceases.

The concentration response calculated based on a PB-PK model accounts for the competition by the two substrates and the optimum concentrations for producing maximum tissue 2,5-HD concentrations during hexane exposures. The model also clearly showed the complex behavior of 2,5-HD during and after the exposure of rats to 3000 ppm hexane. In assessing the dose response for 2,5-HD, the toxicity should be related to 2,5-HD in blood or in nervous tissue.

More complete simulation models of the toxicity of 2,5-HD require incorporation of factors related to addition of the ketone with neurofilament proteins, accumulation and clearance of the adducted proteins, and adaptation of the nervous system to the insult (see Fig. 3). The development of this type of biological information is especially important in assessing the subchronic to chronic exposure uncertainty factors used in some noncancer risk assessments.

Organophosphates (OPi). This class of compounds has been widely used as pesticides. Some of these pesticides, such as parathion, require metabolism to the active oxygenated form for activity. These compounds act by reacting with and blocking serine residues at the active site of acetylcholinesterase. Gearhart and colleagues (1990 and 1994) have developed comprehensive PB-PK models for diisopropyl fluorophosphate and parathion. These models integrate metabolic activation, reactivity, solubility, and anatomical information into a predictive model for the disposition of the parent compound and the inhibition of various esterases in blood cells, plasma, and tissues.

With these compounds, the main toxicologically significant interactions occur via second-order reactions of the ac-
tive cholinesterase inhibitor with the enzymes. In addition, the rates of hydrolysis or aging of the phosphate ester formed at the active site determine the duration of inhibition. The concentration of active toxicant reaching the target tissues is related to (1) rates of inactivation by reaction with esterases in the blood, (2) rates of inactivation by esterases that render the parent drug inactive, and (3) activation of the thiocompounds by oxidation in liver and other tissues. The PB-PK models integrate all of this information into coherent predictions of the time course of inhibition in various tissues. These OPi models may also find a role in exposure assessment by assessing previous exposure levels based on current observations of decreased plasma or red blood cell cholinesterase activities.

**Nicotine.** This alkaloid exerts effects on the nervous system due to interactions with cholinergic receptors. In humans, nicotine has clear physiological effects on heart rate mediated by central nervous system effects. In addition, it enhances performance in certain psychomotor tests, including enhancement in central information processing and improvements in sensorimotor performance. The psychomotor effects of nicotine are believed to be associated with its interaction with cholinergic receptors on the cell body of dopaminergic neurons. Plowchalk et al. (1992) developed a rat PB-PK model for nicotine that included high-affinity binding of nicotine to specific receptors in several tissues including the brain. The coupling of this dosimetry model with pharmacodynamic models is now in progress. The goal of this effort is the elaboration of the relationship between CNS-nicotine concentrations, expected nicotine occupancy of brain receptors, and altered physiological or psychomotor response. Ultimately, the goal of development of this nicotine model and of most PB-PK models is development of a refined understanding of the correspondence of tissue dose, expressed in increasingly sophisticated terms (Fig. 3), and the biological response of interest.

**Summary.** PB-PK models have begun to fulfill their promise in cancer risk assessment, where they serve to define target tissue dose, integrate diverse data within a risk assessment context, define important data gaps for investigation, and aid in prioritizing ongoing research studies (Andersen et al., 1995). Noncancer risk assessments have not yet made use of these techniques, partly at least due to the rigidity with which the uncertainty factors have been applied. However, the introduction of explicit consideration of appropriate measures of tissue dose in the RfC methodology (Jarabek, 1995) appears to open the door for increased use of dosimetry and mode-of-action models in these risk assessments. These PB-PK models appear to offer a good starting point for producing alternative risk assessment strategies for neurotoxicants and for compounds affecting other organ systems.

**The Role of Biomarkers in the Risk Assessment of Neurotoxicants (D. Bellinger)**

Valid measures of dose and effect are required in order to define exposure–outcome relationships in epidemiological...
As indicated in Fig. 4, for each point along the pathway individual differences may affect the likelihood of event occurrence. The basis for this could be an intrinsic ("host") characteristic or preexisting disease state that increases the internal dose or the biologically effective dose or that amplifies the effect at the target tissue by altering uptake, pharmacokinetics, or response. Because these differences may be very large, identifying markers of susceptibility is an important component of a comprehensive neurotoxicity risk assessment. For example, individual-specific threshold doses for postnatal neurological effects of fetal methyl mercury may span two to four orders of magnitude, depending on the health end point (Hattis, in Beck et al., 1993).

Biologic markers of exposure. Biologic markers are points on a continuum between health and disease. The distinction between a marker of exposure and a marker of effect is murky because an early biologic effect can also be viewed as a marker of exposure (Silbergeld, 1993). The accumulation of erythrocyte protoporphyrin (EP) represents a cellular response to the presence of lead in the red cell (an early biologic effect), but, until recently, it was used to screen children for elevated lead exposure. The location of markers on the continuum implies a causal model of disease pathogenesis and progression and is thus most valid if based on a thorough mechanistic understanding (Schulte, 1989). If it is not, misclassification of exposure or effect will occur, reducing the precision of quantitative estimates of dose–response relationships (Schulte and Mazzuckelli, 1991). For example, EP is a satisfactory exposure biomarker only over a limited portion of the lead dose range (>50 µg/dl). Approximately 50% of children with a blood lead level >30 µg/dl do not have an elevated EP level (Mahaffey and Annest, 1986), indicating a need to identify lead exposure biomarkers that have tighter relationships to internal dose than EP does, at least within the exposure range of greatest current interest to risk assessors.

Uncertainties in the assessment of prenatal lead exposure illustrate how mechanistic or pharmacokinetic data can enrich quantitative risk assessment. Although many indices of

---

**FIG. 3.** Processes of interest in the exposure: dose–response continuum for toxicological and pharmacological responses. Overall dose–response behaviors are dependent both on pharmacokinetic factors associated with particular chemicals (steps closer to the top of the flowchart) and with pharmacodynamic processes dependent on the intrinsic responses of the biological organism which are nearly independent of the nature of the particular chemical (steps closer to the bottom of the flowchart).

---

**FIG. 4.** A model of the classes of biologic markers and the progression (left to right) in a disease process from one class to the next. The dotted lines indicate that individual differences in susceptibility influence the rate and likelihood of progression. Source: National Research Council (1989).
prenatal lead exposure are available, they appear to sample body pools with different exposure averaging times and, quite possibly, different toxicologic significance. Only modest correlations (<0.4) hold among concentrations in different compartments (maternal blood at different stages of pregnancy, umbilical cord blood, placental body, placental membranes, cord tissue) (Baghurst et al., 1991). Not surprisingly, alternative exposure biomarkers may not be related in the same ways to a particular response. Although in one study the correlation between maternal and umbilical cord blood lead levels was substantial (0.80), only cord blood lead level was significantly associated with abnormal neurological soft signs in infants (Ernhart et al., 1986). This association was stronger when adjusted for maternal blood lead level, suggesting that soft sign scores were related to the portion of the variance in cord blood lead level that did not covary with maternal blood lead level.

In summary, quantitative risk assessment is improved by the availability of well-validated mechanistically based kinetic models that link alternative exposure biomarkers and that describe the relationships among them and their relative toxicologic significance for different effects. Without such a model, information about separate exposure markers cannot be integrated to estimate dose delivered to critical tissue sites. Individual exposure biomarkers may support different characterizations of dose–response and potentially different conclusions about critical dose, resulting in an unconvincing risk assessment.

**Biologic markers of effect.** The health end points of interest in much of occupational or environmental epidemiology are categorical variables (e.g., diseased/not diseased, dead/alive). In contrast, many end points used in neurotoxicity studies are continuously distributed indices of function (e.g., IQ in children). This creates a need to develop risk assessment strategies that exploit the multidimensional information in such indices.

Sometimes a quantitative score is “quantalized,” as when IQ is classified as “low” (i.e., $\leq 70$) or “normal” (i.e., $>70$). Any cutpoint reflects a strong assumption about what constitutes an “adverse” effect, a definition that may not be biologically meaningful in terms of the mechanism of neurotoxicity. Its use may also affect the likelihood that a particular toxicant is viewed as hazardous. If an IQ $\leq 70$ were used as a cutpoint, current evidence would provide little support for the hypothesis that low-level lead exposure affects children's cognition because the IQ decrements tend to be modest, only a few IQ points per 10 μg/dl increase in blood lead level (Bellinger, 1995). Furthermore, because scores $\leq 70$ are in the extreme lower tail of the IQ distribution (below the second percentile), a very large sample size would be needed to detect an increase in risk associated with a toxicant exposure. The value of any cutoff is thus design-dependent because the statistical significance of a difference in prevalence varies with sample size.

The use of an effect cutpoint also ignores the shape of the dose–response relationship at response levels above the cutpoint, i.e., variations that may be exposure-related. Dose–response estimation thus reduces to the relatively limited task of identifying the threshold associated with the particular cutpoint chosen. In contrast, with a continuously distributed effect biomarker one can evaluate whether effects occur anywhere along the response continuum. For example, even among 10 year olds with IQ scores 2 or more standard deviations above the population mean, academic skills were inversely related to blood lead level at 2 years of age (Bellinger, 1995). Although the performance of these children was well above average (at least in terms of IQ), the dose-related variation in academic scores clearly represents an “adverse” effect of early lead exposure. The general point is that the dose–effect relationship for a continuously distributed response variable can be decomposed into (1) the probability or incidence of response at a given dose (as with a quantalized outcome) and (2) the severity of response at a given dose, expressed as a distribution of responses around the average value (Wyzga, 1990). This permits examination of the distribution of responses at multiple points on the dose–effect curve, rather than at just one point. This approach uses more of the available information than is involved in deriving, for instance, a reference dose. Thus, one can assess exposure-related changes in the variability or dispersion of a distribution as well as in its central tendency. Usually viewed as power-reducing “noise” in neurotoxicity studies, variability in response to a given dose bears importantly on the issue of individual differences in susceptibility. For example, in one study, children with “high” cord blood lead levels had lower developmental scores in the first 2 years than did children with lower prenatal exposures (Bellinger et al., 1987, 1991). How did the individual estimates of “response severity” in the high lead group combine to produce a lower mean response? Was the underlying distribution of response severity highly skewed, perhaps bimodal, with most children showing no response but a small subset showing a “large” response (i.e., the “responders”)? Were most children responders, with the magnitude of response to a given dose normally distributed with random variation about a central value? The implications for risk characterization are straightforward. If only a small subset of children were responders, efforts to increase their host resistance by focused interventions would be preferable to primary prevention of exposure, which in turn would be preferable if most children were responders.

An “expected” IQ-type score was computed for each child on the basis of a variety of factors (e.g., maternal IQ, family social class, home environment, birth weight). The
distributions of residual scores (i.e., observed minus expected IQ) did not differ significantly from normal for any of the four doses (i.e., cord blood lead strata), thus failing to support the "small subset of responders" hypothesis.

In this same cohort, social class was related both to the apparent dose at which a performance decline was first apparent and to the overall severity of response (Bellinger et al., 1988). Children in the lower portion of the social class distribution appeared to be more vulnerable from the standpoint of the threshold, whereas children from the upper portion of the distribution appeared to be more vulnerable from the standpoint of the magnitude of performance decline over the dosing range in the cohort. With a continuously distributed effect biomarker, unique patterns of correlates can be identified for different aspects of the dose–effect relationship.

Several investigators have described the importance of distinguishing the impact of an exposure on the central tendency of a distribution from its impact on the percentage of a population falling into the extreme tails of a distribution (Weiss, 1990; Feingold, 1995). The population implications of this distinction may be profound when the implicated exposure is prevalent.

Using continuously distributed markers of neurological function also affords an opportunity to analyze temporal characteristics of performance, something that is not possible with dichotomous end points that are not time-varying (e.g., "final" states such as death or malformation). Change in a growth curve function can be decomposed into level of performance and rate of change in performance, which may vary independently. Rate of development may be the first dimension of performance to be affected by a toxicant exposure, so modeling rate changes would reveal toxicant effects earlier in a disease process than would modeling changes in level of performance. If the features of a within-subjects individual growth curve model vary among subjects, a between-subjects model can be fitted with the features of the within-subjects model as dependent variables and subject characteristics (e.g., hormonal status, sex, age) as independent variables (Fletcher et al., 1991). In the framework of biologic markers, these are potential markers of susceptibility.

These statistical techniques have not been widely used in neurotoxicological research (Watertaux et al., 1989). In one study, the relationship between dentin lead level and children's scores on a word reading test between 8 and 12 years of age was better described by a constant decrement model (growth curves have parallel slopes but different intercepts) than by either a catch-up or a deterioration model (Fergusson and Horwood, 1993). In a study in which early developmental deficits associated with high cord blood lead levels appeared to attenuate by age 5 years, several correlates of improved performance were identified: sex (female > male), social class (higher > lower), and concurrent blood lead level (lower > higher) (Bellinger et al., 1990).

**Quantitative Risk Assessment of Neurotoxicants with the Use of Continuous Data (W. Slikker)**

Most health risk estimations have focused on quantal data where human or animal model data are categorized as a positive or negative biological effect (e.g., birth defect or tumor). Different procedures are necessary for nonquantal or continuous data. Examples of continuous data include hematomal measurements and body or organ weight as well as many neurotoxicological end points (e.g., behavioral, neurophysiological, and neurochemical data) (Slikker and Gaylor, 1995b). Although Crump (1984) utilized several dose–response models to estimate the doses corresponding to various changes (e.g., 1.5, or 10%) from the control means, attempts to estimate the risk from continuous data have only recently been presented (Gaylor and Slikker, 1990; Kodell and West, 1993; Crump, 1995).

In an approach to assess risk utilizing continuous data developed by Gaylor and Slikker (1990), a mathematical relationship is first established between the average biological effect and the dose of a given chemical. A second step determines the distribution (variability) of individual measurements of biological effects about the dose–response curve. The third step statistically defines an adverse or "abnormal" level of biological effect in an untreated population. The fourth step estimates the probability of an adverse or abnormal level as a function of dose, utilizing the information from the first three steps. The advantages of these dose–response models are that they encourage the generation and use of data needed to define a complete dose–response curve and provide an estimate of risk and/or changes in the average response as a function of dose.

**Biologically based risk assessment.** The development of quantitative risk assessment approaches depends, in part, on the availability of information on the mechanism of action and pharmacokinetics of the agent in question. In the development of a biologically based, dose–response model for the psychoactive agent methylenedioxymethamphetamine (MDMA), Slikker and Gaylor (1990) considered several factors, including the pharmacokinetics of the parent chemical, the target tissue concentrations of the parent chemical or its bioactivated proximate toxicant, the uptake kinetics of the parent chemical or metabolite into the target cell and membrane interactions, and the interactions of the chemical or metabolite with presumed receptor site(s). Because these theoretical factors contain a saturable step due to limited amounts of required enzyme, reuptake, or receptor site(s), of a nonlinear, saturable dose–response curve was predicted. In this case of neurochemical effects of MDMA in the rodent, saturation mechanisms were hypothesized and indeed...
saturation curves provided relatively good fits to the experimental results. Some of the advantages of the biologically based, quantitative approaches over the currently used RfD risk assessment procedures include the ability to (1) utilize continuous data, (2) utilize all of the dose–response curve data, (3) incorporate biological information into the dose–response model, and (4) provide an actual risk (e.g., 1 in 10,000) to a given dose (Slikker and Gaylor, 1995a). The conclusion was that use of dose–response models based on plausible biological mechanisms provides more validity to prediction than purely empirical models.

This quantitative risk assessment procedure originally illustrated with a serotonergic neurotoxicant, MDMA (Slikker and Gaylor, 1990), has been applied to another neurotoxicant, 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), that primarily affects the dopaminergic neurotransmitter system (Tipton and Singer, 1993). MPTP was selected as the agent for assessment because its neurotoxicity is well described in the literature and is composed of behavioral, neurohistological, and neurochemical effects. The results of this assessment are systematically compared to those obtained with the currently used safety factor approach with and without the use of a benchmark dose adjustment (Slikker and Gaylor, 1995a).

The discovery that MPTP, a contaminant of a synthetic meperidine analog sold as a street drug, produced symptoms resembling Parkinson's disease stimulated a large number of studies on the mechanisms involved with the neurotoxic effects of MPTP and their possible relationship to idiopathic Parkinson's disease (Tipton and Singer, 1993). The neurotoxicant MPTP causes selective destruction of nigrostriatal dopaminergic neurons in rodents, nonhuman primates, and humans (Kopin, 1986).

The results from a host of studies reported since the 1980s have been used to formulate a mechanism of action for MPTP (McCrodden et al., 1990; Johannessen, 1991; Tipton and Singer, 1993). In brief, MPTP is a lipophilic molecule that enters the brain rapidly after administration. A two-step reaction requires the enzyme monoamine oxidase (MAO)-B to convert MPTP to the pyridinium metabolite 1-methyl-4-phenylpyridinium (MPP⁺). After this conversion, which occurs primarily in astrocytes, MPP⁺ is sequestered within monoaminergic neurons by energy-dependent and selective monoaminergic transporters. MPP⁺ is a potent inhibitor of mitochondrial oxidation of NAD⁺-linked substrates. The blockade of respiration leads to a depletion of ATP and consequently the loss of membrane potential. The mechanisms underlying the subsequent destruction of nigrostriatal nerve cells may involve altered calcium homeostasis and perhaps free radical formation (Adams et al., 1992).

Specific differences in susceptibility to MPTP (monkey > dog > mouse > rat) may be explained, at least in part, by the length of time the MAO-B-generated MPP⁺ is retained in nigrostriatal tissue. Primates retain MPP⁺ in the striatum with a half-life of about 10 days, whereas MPP⁺ concentrations in rat brain are negligible within hours after MPTP administration (Johannessen, 1991).

The mechanism of action of MPTP provides important information concerning the development of a biologically based, dose–response risk assessment model for this neurotoxicant, as was described previously for MDMA (Slikker and Gaylor, 1990). Several steps in the pathway for MPTP to produce neurotoxicity require either energy-dependent uptake of MPP⁺ by monoaminergic transporters and/or stereospecificity (MAO-B metabolism). These proposed mechanisms or alternatives (e.g., oxidative stress) rely on pathways that are specific and saturable. Therefore, a risk assessment model that incorporates saturation kinetics appears appropriate.

Quantitative risk assessment. To evaluate our quantitative biologically based, dose–response model, MPTP data sets from two mouse studies and one monkey study are utilized as previously reported for MDMA (Slikker and Gaylor, 1990). Alterations in neurotransmitter concentrations are used as end points of toxicity. It must be emphasized that behavioral (decreased motor activity) and neurohistological damage (increased degenerating nigrostriatal cells) also accompany MPTP administration (Ricaurte et al., 1987; Johannessen et al., 1989). Dopamine concentrations from several studies (Ricaurte et al., 1987; Sonsalla et al., 1987; Slikker et al., 1994, unpublished observation) were used to develop the risk analysis (Slikker and Gaylor, 1995a).

The four-step, quantitative risk procedures were applied to these data sets (Gaylor and Slikker, 1990). For continuous data a distinct value that separates normal levels from those associated with toxicity is seldom obvious. In the present example with MPTP, low levels of the neurotransmitter dopamine are associated with impaired motor function, but there is not an identifiable value below which brain/motor function is always impaired and above which brain/motor function is always normal. In addition, even compromised neurotransmitter levels may be sufficient for subjects to perform most behavioral tasks but if they are "challenged," the same neurotransmitter levels may be too low to allow adaptation and normal functioning. Instead of a distinct threshold situation, a continuum exists with the probability of impaired brain/motor function increasing with decreasing levels of dopamine. In the absence of a dopamine level that is always indicative of impaired brain/motor function, it is nevertheless possible to statistically define an abnormal level of dopamine based on the distribution characteristics in a study population. Dopamine concentrations can be measured in the striatal areas of the brain in control (nontreated) ani-
mals and a normal range of dopamine concentrations can be established. Concentrations outside this normal range would be considered abnormal. Then, for animals administered a potentially toxic substance like MPTP, the proportion (risk of animals with abnormal levels of dopamine) can be observed.

Abnormal range. If measurements of a particular neurotransmitter level or other biological end point are available for a large number of untreated control animals, the percentiles of the distribution can be observed and an abnormal range can be selected. For a group of 20 animals, values below the smallest and above the largest measurements can be considered abnormal, and this would estimate the approximate 5th and 95th percentiles of the distribution of values. If the effects are described by a Gaussian (normal) distribution, then it is possible to calculate or look up the value corresponding to any percentile from the estimates of the mean ($\mu$) and standard deviation ($\sigma$): $\mu - 2.33\sigma$ and $\mu + 2.33\sigma$ estimate the 1st and 99th percentiles for a normal distribution, respectively.

As with many biological end points, neurotransmitter concentrations have only positive values that are relatively large. Such data tend to be described by a log-normal distribution (i.e., the logarithms of the values are normally distributed). Thus, converting the data to logarithms and then calculating the mean and standard deviation is an approach that can be used to estimate percentiles.

For such data (assuming a normal distribution) the mean ($\mu$) and standard deviation ($\sigma$) provide estimates of the percentiles. Each experimental observation ($x$) can be standardized to a normal deviate, $Z = (x - \mu)/\sigma$. The probability that $Z$ exceeds a certain value (risk) can be calculated or looked up in normal tables (Beyer, 1968). For example, the probabilities that $Z$ exceeds 1.28, 1.64, and 2.33 are 0.10, 0.05, and 0.01, respectively. In this case we decided to define abnormal values for striatal dopamine concentrations as any value below the 99th percentile of those in the control animals. By definition, then 1% of the control animals are regarded as abnormal. The mean value for the control is $\mu_0$, and the abnormal level is $(\mu_0 - 2.33\sigma)$.

Dose–response model. It is usually necessary to estimate the risk for obtaining an abnormal value for doses other than those that were selected for the bioassay (e.g., lower doses). This requires the fitting of a dose–response curve to the experimental data in order to estimate the average response, $Y$, as a function of dose, $f(d)$. Because the mechanism of action of MPTP is at least partially understood, it is possible to select a particular mathematical form, $f(d)$, and estimate the unknown parameters. Because the biological effect of MPTP is reported to require an enzyme-mediated activation (MAO-B conversion of MPTP to MPP$^+$) and because MPP$^+$ is selectively taken up into dopaminergic cells via an energy-dependent transporter, saturation of MPTP's effects is anticipated. Therefore, a Michaelis–Menten-type equation, $Y = (K + Vd)/(K + d)$, appears appropriate. The minimum values, $V$, and $K$ are estimated from the experimental data. The dose–response models for dopamine concentration versus total dose of MPTP are shown in Table 1.

### Comparison of the three risk assessment approaches.

The same three data sets concerning MPTP exposure were evaluated by the current RfD approach, the BMD approach, and our quantitative, biological model approach and the results are compared in Table 2. The traditional approach generated RfD levels of 0.1 and 0.04 mg/kg of MPTP from the two mouse studies and 0.008 mg/kg from the monkey data set. The safety factors used for the Ricaurte mouse study were 10 X 10 = 100 (10 for extrapolation from animal to human and 10 for heterogeneity among humans). The Sonsalla mouse study and the monkey study require an additional factor of 10 because the LOAEL had to be used instead of the NOAEL. The monkey study used a combined safety factor of 300 (10 X 10 X 3 = 300) because of the generic assumption that phylogenetic similarity of the species allows for a safety factor of 3 instead of 10 for extrapolation from nonhuman primates to humans.

Using a quantitative model, the benchmark dose was set at the lower 95% confidence limit on the dose estimated to have a 5% risk of abnormal animals (LED$_{0.5}$). For the BMD approach, the estimated "safe" dose of MPTP is four to
eight times lower than that obtained with the RfD approach at 0.018 and 0.005 mg/kg for the mouse studies and 0.002 mg/kg for the monkey study.

The quantitative approach was then used to provide an estimate of the dose that would result in a risk equal to or less than 1 in 10,000 subjects (i.e., only 1 in 10,000 subjects would suffer abnormally low levels of dopamine). This dose is two times lower than the BMD for the mouse studies and approximately seven times lower than the BMD for the monkey study.

This study of the neurotoxicant MPTP provides an example of the quantitative risk assessment process and the opportunity to compare the results obtained using this method to those obtained from the currently used safety factor method with and without a benchmark adjustment. By using three data sets from two different species, the degree of interspecies variability may also be assessed.

The MPTP dose determined acceptable or safe by the currently used qualitative risk assessment (RfD) procedure is considerably higher than the dose determined to produce a risk of 1 in 10,000 as defined by the quantitative method. The quantitative method applied to the mouse data resulted in a dose 20–20 times more conservative and the monkey comparison resulted in a dose 25 times more conservative than the currently used method. The benchmark adjustment of the current method resulted in intermediate values. In all three cases, the traditional approach of using the NOAEL or LOAEL for the calculation of a reference dose would have been based on MPTP doses resulting in large proportions of animals with abnormally low dopamine concentrations.

Shortcomings of the currently used qualitative or safety factor risk assessment method include (1) the use of only one data point (NOAEL or LOAEL), (2) the need for the assumption that a threshold of effect exists, (3) the need for an assumption that factors of 10 are appropriate to account for interspecies and intraspecies variability, and (4) the inability to adjust the assessment to achieve a desired risk (e.g., 1 in 10,000). The advantages of a quantitative approach demonstrated here over the currently used RfD risk assessment procedures include the ability to (1) utilize continuous data, (2) utilize the whole of the dose–response curve data, (3) incorporate biological information into the dose–response model, and (4) provide an actual risk of exposure to a given dose.

In general, the described quantitative approach attempts to replace some of the assumptions of the currently used risk assessment process with biological and/or mechanistic information and utilizes continuous dose–response data. Further studies to compare these quantitative approaches with current procedures and to validate these novel approaches are needed. A single risk assessment model may not be adequate for all conditions of exposure, for all end points, or for all agents. Risk assessment models of the future may well include biomarkers of both effect and exposure, as well as biologically based pharmacokinetic and mechanistic considerations derived from both epidemiologic and experimental test system data.

TABLE 2
Acceptable Total Dose (mg/kg)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Current: NOAEL/100 or LOAEL/1000*</td>
<td>10/100 = 0.100 mg/kg</td>
<td>40/1000 = 0.040 mg/kg</td>
<td>2.4/3000* = 0.008 mg/kg</td>
</tr>
<tr>
<td>Benchmark dose</td>
<td>1.8/100 = 0.018 mg/kg</td>
<td>0.5/100 = 0.005 mg/kg</td>
<td>0.066/30 = 0.002 mg/kg</td>
</tr>
<tr>
<td>Quantitative approach:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimate of dose with risk ≥0.0001</td>
<td>0.009 mg/kg</td>
<td>0.002 mg/kg</td>
<td>0.0003 mg/kg</td>
</tr>
</tbody>
</table>

*Safety [uncertainty factor of 10 for extrapolation from animal to human, 10 for heterogeneity among humans, and 10 if no-observed-effect level (NOAEL) not obtained and lowest observed effect (LOAEL) are used].

*For primates a safety factor of 3 instead of 10 was used for extrapolation from animals to humans.

*Lower 95% confidence limit on the dose estimated to have a 5% risk of abnormal animals (LED5).

REFERENCES


