FUNDAMENTAL AND APPLIED TOXICOLOGY 20, 1–14 (1993)

SYMPOSIUM OVERVIEW

Improvements in Quantitative Noncancer Risk Assessment[1]

Sponsored by the Risk Assessment Specialty Section of the Society of Toxicology

BARBARA D. BECK,* RORY B. CONOLLY,† MICHAEL L. DOURSON,‡ DANIEL GUTH,§ DALE HATTIS,¶ CAROLE KIMMEL,‖ AND STEVEN C. LEWIS⊥

*Gradient Corporation, 44 Brattle Street, Cambridge, Massachusetts 02138; †CIIT, Inhalation Toxicology and Biomathematical Modeling, 6 Davis Drive, Research Triangle Park, North Carolina 27709; ‡Environmental Criteria & Assessment Office, U.S. EPA, 26 Martin Luther King Drive (MS-114), Cincinnati, Ohio 45268; §U.S. EPA, Office of Research and Development, MD-52, Research Triangle Park, North Carolina 27711; ¶Clark University, CENTED, 950 Main Street, Worcester, Massachusetts 01610; ‖U.S. EPA, Office of Research and Development (RD-689), 401 M Street, S.W., Washington, DC 20460; and ⊥Exxon Biomedical Sciences, Inc., Toxicology Division, Mettlers Road (CN-2350), East Millstone, New Jersey 08875-2350

Received July 20, 1992; accepted July 22, 1992

Improvements in Quantitative Noncancer Risk Assessment. BECK, B. D., CONOLLY, R. B., DOURSON, M. L., GUTH, D., HATTIS, D., KIMMEL, C., AND LEWIS, S. C. (1993). *Fundam. Appl. Toxicol.* 20, 1–14. © 1993 Society of Toxicology.

Overview (Barbara D. Beck)

At the 1992 Society of Toxicology meeting, the Risk Assessment Specialty Section held a symposium titled "Improvements in Quantitative Noncancer Risk Assessment," chaired by Michael Dourson and Barbara Beck. The purpose of the symposium was severalfold:

- To identify and quantify the major uncertainties with the standard RfD (reference dose) or ADI (acceptable daily intake) method;
- To describe alternatives to the standard methodologies —alternatives which make better use of different types of data, including dose–response and severity information;
- To present a methodology which uses biologically based modeling to predict risk.

Overall, there was a general conclusion that the standard approach of applying default uncertainty factors to a no observed adverse effect level (NOAEL) to yield a "permissible" level (i.e., the RfD) for chronic human exposure yields limited information. For example, as described by Lewis in the next section, while a 10-fold uncertainty factor to adjust from subchronic to chronic exposures represents the typical default value, a value of 3.5 or less would be appropriate for most chemicals analyzed.

On the other hand, as discussed by Dale Hattis (see subsequent section), interindividual variability in susceptibility to some compounds can be fairly large. For postnatal neurological damage resulting from fetal exposure to methyl mercury, the range in susceptibility (i.e., the individual-specific threshold dose) was estimated to be approximately 10,000-fold. In this case, application of the standard 10-fold default uncertainty factor for interindividual variability could leave a significant fraction of individuals at risk.

Thus, improvements to the RfD approach might be accomplished through in-depth evaluation of the toxicological database for a particular chemical and selection of variable chemical-specific values.

Even with the use of judgment-based uncertainty factors, the RfD approach is still limited in that it does not consider the shape of the dose–response curve and is based on a NOAEL whose value is dependent upon the conditions of a particular experiment. For example, as described by Kimmel and Zenick (see subsequent section), a smaller study may result in a higher NOAEL (and thus a higher RfD) than a larger, better quality study. Kimmel and Zenick describe an alternative to selection of the NOAEL—the benchmark dose. The benchmark dose is "the lower confidence limit on dose corresponding to an increase in the incidence of an effect at a particular risk level," which might for example be a 10% response level for a predetermined adverse effect. This approach has been applied to developmental toxicity studies and appears to be quite promising. The benchmark method is likely to provide

[1] Symposium held at the Society of Toxicology meeting, February 27, 1992, in Seattle, WA.

0272-0590/93 $5.00
Copyright © 1993 by the Society of Toxicology.
All rights of reproduction in any form reserved.

greater consistency among RfD's as well as rewarding studies of higher quality. For calculating the benchmark dose, judgments must be made regarding the selection of a particular risk level, which might vary according to the severity of the effect.

Risk assessment for acute inhalation exposures involves important considerations regarding exposure duration. For example, as described by Guth, Hertzberg, and Jarabek (see subsequent section), an adverse effect level for nervous system effects for a 1-hr exposure to dichloromethane can be 10-fold or more higher than for a 4-hr exposure. Thus, the interrelationship between exposure concentration, time and severity of effect is complex and not amenable to application of standard default factors. Guth and co-workers describe a modeling technique that makes use of a range of endpoints and provides an estimate of the probability that, at a defined concentration and exposure duration, the severity will be less than or equal to the defined severity category. This type of approach could, in theory, lead to the development of exposure/concentration nomograms for noncancer risk assessment.

Unlike cancer risk assessment, noncancer risk assessment has made little use of information on the biological basis of toxicity, at a cellular or molecular level, despite the wealth of information on mechanisms for some compounds, such as chloroform. As described by Conolly (see subsequent section), mechanisms of chloroform toxicity have been relatively well evaluated and appear to involve the metabolism of chloroform to produce phosgene, which can form adducts with macromolecules, followed by cytolethality. Cytolethality would, in turn, result in liver cell proliferation which may be important in hepatocarcinogenesis. Conolly provides an elegant demonstration of how this particular model yields results consistent with the differential toxicity of chloroform according to the route of exposure. The model also provides an explanation of this phenomenon—i.e., differences in integrated dose over time of active metabolite. Biologically based modeling could, for well-studied compounds, be used in the same manner as cancer risk assessment models—to predict the likelihood of an adverse effect in humans over defined conditions of exposure and dose, as well as to estimate the variability in responses.

Reducing Uncertainty with Adjustment Factors (Steven C. Lewis, Exxon Biomedical Sciences, Inc.)

Background. A wide variety of laws and regulations place limits on human exposures to potentially toxic chemicals. Allowable levels of chemicals in ambient or workplace air, in water, and in foods, drugs, cosmetics, and other consumer products are specified in virtually every developed nation in the world. Deriving acceptably safe exposure guidelines is one of the practicing toxicologist's most challenging tasks.

Deriving human exposure guidelines usually involves downwardly adjusting the results of tests on laboratory animals. The downward adjustments are to account for differences between laboratory animals and humans, for differences between experimental conditions and the "real world," and for uncertainty in the safety assessment process. This paper addresses one of the more frequently encountered adjustments: the adjustment which is intended to account for a less-than-lifetime exposure in the animal experiment (i.e., to accomplish "subchronic-to-chronic" extrapolation). Although a 10-fold adjustment is most commonly used, research suggests that the use of alternate values is scientifically plausible. Weil and McCollister (1963) and McNamara (1976) found that the ratios of subchronic-to-chronic NOAELs from a large number of well-designed studies were almost always 5 or less, with 2 being the approximate average.

Design and Methods. A follow-up study was conducted, based on data from more recently conducted studies than had been available at the time of Weil's and McNamara's publications. In the present study, candidate data sets came primarily from two sources, from peer-reviewed journals or from the U.S. National Toxicology Program reports. Criteria for including a data set in the $NOAEL_{subchr}/NOAEL_{chr}$ ratio[2] calculations were

only noncancer endpoints were considered,

the "critical" effect must have been the same in both subchronic and chronic studies,

original investigators' opinions regarding NOAELs, LOAELs, and FELs were accepted.

Only high quality data were included, i.e., from studies which scored 5 or greater on the DARS (data adequacy rating scale; see Table 1). That is, to be admitted to the study database, the results must have included *both* a dose at which no adverse effect was observed, plus at least one higher dose at which signs of significant toxicity were observed. Fifty three percent (18 of the first 34 evaluated data sets) met or exceeded the criterion for inclusion in the study database.

Results. Figure 1 displays the $NOAEL_{subchr}/NOAEL_{chr}$ ratio determinations for the first 18 data sets. The most frequently observed $NOAEL_{subchr}/NOAEL_{chr}$ ratios fell between 1.6 and 2.5. In 4 of 18 cases (22%), the $NOAEL_{subchr}/NOAEL_{chr}$ ratio ≈ 1. Seventy eight percent of the ratios (14 of the 18 studies) were 3.5 or less. Only 1 (of 18) was greater than 10. Arithmetic average was 3.3 (excluding the "outlier"); including the outlier raised the average value to 6.4.

[2] All acronyms are defined in the Appendix.

TABLE 1
Data Adequacy Rating Scale

Score	When study results provide
1	FEL only
2	LOAEL only
3	NOAEL only
4	LOAEL + FEL
5	NOAEL + FEL
6	NOAEL + LOAEL
7	NOAEL + LOAEL + FEL

Discussion. Inflexible and arbitrary data adjustment factors result in limited utilization of existing knowledge. Their use essentially assures that some (often much) information, particularly regarding the slope and shape of the dose–response relationship, is excluded from the safety assessment process. Inflexible adjustment factors further serve to discourage the practice of advanced toxicology; i.e., if a single adjustment factor is to be employed (regardless of the numbers of animals studied), there is little incentive to expand the study to provide greater reliability. Finally, the practice often results in unnecessary and expensive overregulation. Lewis *et al.* (1990) recently proposed an alternative approach (to the arbitrary use of factors-of-10), employing flexible adjustments, based on expert consensus.

The null hypothesis was that the findings from the present study would be indistinguishable from the results of the studies of Weil and McCollister (1963) and McNamara (1976). Indeed, the subchronic-to-chronic NOAEL ratios from this study confirmed the earlier findings of Weil and McNamara almost exactly (i.e., that Weil and McNamara were correct in their observations that for downwardly adjusting a $NOAEL_{subchr}$, a value of 3 is nearly always sufficient to include the $NOAEL_{chr}$).

In a single instance, a ratio greater than 10 was observed. The finding, however, may be misleading and deserves further attention. The data were taken from a study in which the doses were very widely separated. More refined results and more-detailed characterizations of the $NOAEL_{subchr}$/$NOAEL_{chr}$ ratios are expected as additional studies are added to the database.

FIG. 1. Subchronic/chronic NOAEL ratios.

With the development of an even larger and more representative database, it should be possible to achieve a consensus on what the most appropriate alternative to the arbitrary use of factor-of-10 data adjustments is. The most promising alternative seems to be ranges of values for each factor, which reflect the full richness of current toxicologic understanding.

Appendix

Adverse (effect)	Functional impairment or pathological lesion which may affect the performance of the organism or which reduces its ability to respond to additional challenge
Critical (adverse effect)	Specific adverse effect on which the risk assessment is based
LOAEL	Lowest observed adverse effect level; the lowest dose which produces *either statistically or biologically significant increases in either frequency or severity of adverse effects* between the exposed population and its appropriate control
NOAEL	No observed adverse effect level; the highest dose of a chemical at which there are *neither statistically nor biologically significant increases in the frequency or the severity of adverse effects* between the exposed population and its appropriate control
FEL	Frank effect level; less rigorously defined; distinguished from the LOAEL by a greater frequency and/or severity of the adverse effect
$NOAEL_{subchr}$	NOAEL from a subchronic study (usually 90 days)
$NOAEL_{chr}$	NOAEL from a chronic study (usually 2 years)

Definitions are rephrased slightly from Dourson and Stara (1983).

Acknowledgments

Special thanks are extended to Carrol Weil, Scientific Advisor, Kathe Stauber and John Adgate, Report Analysts, and Larry Andrews (ARCO

Chemical), Matt Bogdanffy (DuPont), Vivian Carlson (Exxon Biomedical), Rick Davis (American-Cyanamid), Carolyn Leep (Chemical Manufacturers Association), Hon-Wing Leung (Union Carbide), and Craig Nessel (Exxon Biomedical).

References

Dourson, M. L., and Stara, J. F. (1983). Regulatory history and experimental support of uncertainty (safety) factors. *Regul. Toxicol. Pharmacol.* **3**, 224–238.

Lewis, S. C., Lynch, J. R., and Nikiforov, A. I. (1990). *Regul. Toxicol. Pharmacol.* **11**, 314–330.

McNamara, B. P. (1976). Concepts in health evaluation of commercial and industrial chemicals. In *New Concepts in Safety Evaluation.* Hemisphere, Washington, DC. [Cited in Dourson and Stara, 1983]

Weil, C. S., and McCollister, D. D. (1963). Relationship between short- and long-term feeding studies in designing an effective toxicity test. *Agric. Food Chem.* **11**, 486–491.

Going Beyond Uncertainty Factors—Opportunities for Quantitative Toxicology (Dale Hattis, Clark University)

The primary theme of this discussion is that we need not accept the status quo NOEL/uncertainty factor procedure as the end of discussion, research, and analysis of noncancer risks. There is a rich field for study here that has the capability to yield insights of both scientific and social policy interest. Many measurements already exist in the literature that are relevant to assessment of the degree of protection offered by the 10-fold rules of thumb and which might guide us to adapt the 10-fold factors to more accurately gauge risks in specific types of cases (e.g., different categories of chemicals, processed or absorbed in different ways, producing different kinds of noncancer effects). Yet these data are unfortunately not at present used, even to check and calibrate the overall procedure, because there has as yet been no regulatory/decision-making "market" for either novel analyses of existing data or projects to make new and better measurements of the degree of human interindividual variability in pharmacokinetic or other response factors.

First, for clarity, I need to distinguish between two different concepts/phenomena that are both describable as probability distributions. A description of "uncertainty" is a description of the *imperfection in knowledge* about something that is conceived of as having some "true" single value in some inaccessible reality. "Heterogeneity" or "interindividual variability" is the distribution of true values of a parameter that would be found in a population by perfectly accurate measurement techniques. For example, if I were to weigh a set of individuals with an excellent scale, I would not get the same results for different people. People really do differ in their weights (and other characteristics that affect individual risks) and no degree of improvement in my measuring instrument will make them all the same. By contrast, an increasingly accurate scale would be expected to reduce my uncertainty about the true weight of any individual or about the average weight of the individuals in the room.

My focus here will be on interindividual variability. For noncancer effects, the degree of human interindividual variability plays a central role in determining the risk that can be expected at low exposures (Hattis and Silver, 1990).

For purposes of measurement and analysis, it is useful to divide the overall interindividual variability in susceptibility into three components:

Uptake—Individual differences in the environmental concentration needed to produce a given intake of toxicant into the body, e.g., due to differences in breathing rates, dietary habits, etc.,

Pharmacokinetic—Individual differences in the amount of uptake needed to produce a particular concentration × time of active agent in the blood or at the site of action, e.g., due to differences in metabolic activation or clearance,

Response—Individual differences in the dose at the active site that produces a similar risk of response.

As an example of the first category, Fig. 2 shows a plot of the correspondence of breathing rate data of Jones *et al.* (1981) to a normal distribution (the data in this case are better described as a normal, rather than a lognormal distribution—for a lognormal distribution, one would simply plot the logarithms of the values vs the same "Z-score").[3]

The second category—pharmacokinetic parameters—can be usefully divided into three subcategories:

- *Elimination half-lives* ($t_{\frac{1}{2}}$) are defined as the time required for a twofold reduction in the concentration of the substance in some compartment (usually blood) after absorption is complete. The longer the elimination half-life, the higher the concentrations that will be attained in the body if exposure is continuous or repeated at a frequency that is short relative to the half-life. Therefore, elimination half-lives are likely to be key determinants of susceptibility to toxicity from chronic continuous exposures.

[3] To create this type of plot, the measurements are first arranged in order and given ranks i (1 through N). Then one calculates a "percentage score" for each ordered value as $100^{*}(i-0.5)/N$. (This is simply the percentage of an infinite sample that would be expected to be less than or equal to the observed value. It differs from the usual definition of a "percentile" in which the highest observation is assigned a score of 100.) Finally, from tables of probits in or areas under a cumulative normal distribution, one calculates the number of standard deviations above or below the median of a normal distribution that would be expected to be associated with each percentage score, if the distribution of values were in fact normal (Gaussian). In the regression line calculated from this type of plot, the intercept ($Z = 0$) represents the expected median, and the slope represents the standard deviation. One advantage of this type of data analysis is that it does not require making arbitrary assumptions about the values of the plotted parameter below a specific detection limit. One simply includes the "nondetect" values in computing N and fits the line to the points in the region where measurements are deemed to be meaningful.

• *Area under the curve* (AUC) is a plot of the plasma concentration versus time after exposure. The AUC integrates variability in the efficiency of absorption as well as metabolism and elimination. It is likely to be a good predictor of pharmacokinetic variability in susceptibility for effects that are linearly related to the amount of slowly accumulating products of reaction between the toxicant and resident macromolecules.

• *Peak concentration* (C_{max}) in blood represents an effective dose in cases of acute toxicity from an isolated single exposure to a substance. Like AUC, it also integrates information about absorption: in this case, both the efficiency and dynamics of absorption.

As a rule, lognormal distributions are more frequently better descriptions of interindividual variability in these parameters than normal distributions. Figure 3 shows a lognormal plot of the distribution of methyl mercury half-lives. In earlier work, I have compiled the results of about 100 such data sets on the interindividual variability of different systemic pharmacokinetic parameters (Hattis *et al.*, 1987).

Effect parameters can either represent susceptibility to end effects in themselves or be "functional intermediate parameters" that mediate, or are good proxies that can be used to make estimates of the risk of an end effect of concern (such as infant mortality—see Ballew and Hattis, 1989). The value of continuous functional intermediate parameters like birth weight or sperm counts or FEV_1 is that they are generally much more amenable to measurement in large populations in relation to toxicant exposures than quantal end-effect parameters such as infant mortality, impaired male fertility, and chronic obstructive lung disease cases.

Occasionally, however, reasonable quantal dose–response data on end effects in humans are forthcoming, and

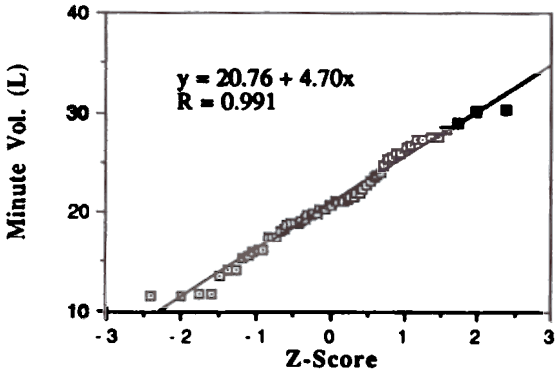

FIG. 2. Fit of a normal distribution to the minute volumes observed in 62 British coal miners (data of Jones *et al.*, 1981). The straight line represents the expected distribution of the points for a normal distribution with a mean of 20.76 liters/minute and a standard deviation of 4.70 liters/minute.

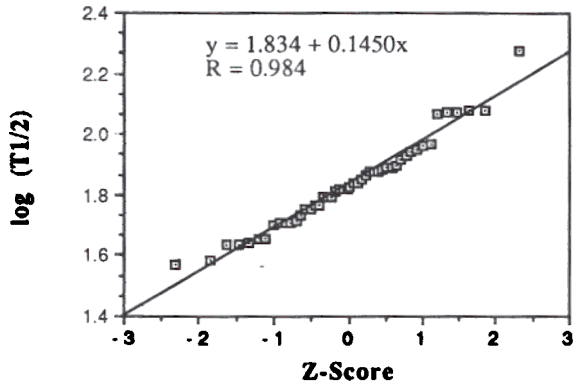

FIG. 3. Lognormal plot of the distribution of methyl mercury biological half-lives (data of Al-Shahristani and Shihab, 1974).

when that occurs we have the opportunity to directly evaluate how protective our standard 10-fold "uncertainty" factors are in different circumstances. For example Marsh *et al.* (1987) have published detailed information on the incidence of a variety of fetal methyl mercury effects in relation to the maximal levels of mercury found in the hair of the mothers during gestation. (The observations come from an Iraqi mass poisoning incident which resulted from the distribution of methyl mercury-treated Green Revolution seed grain.) Maximum mercury concentrations were assessed by a series of sequential measurements along the hair shafts during fetal development. Log probit dose–response fits to these data (e.g., Fig. 4) indicate very large amounts of interindividual variability in response—indicated by lower slopes than are observed for effects in adults (e.g., Fig. 5, from data of Clarkson *et al.*, 1976, and Bakir *et al.*, 1973).[4]

An important caveat to this comparison is that in addition to statistical uncertainties in the determination of these slopes from limited data, there are questions of biological interpretation. A conclusion that the relationships represented in Fig. 4 represent true interindividual variability depends on an assumption that the biomarker of exposure used in this case—the maximum hair mercury found at any time during gestation—is the most appropriate direct causal predictor of response that can be developed. Other

[4] In such plots, a probit slope of about 1 (corresponding to a geometric standard deviation for the individual thresholds of 10) would imply that 95% of the population would have thresholds for effect spread out over a span of about 10,000-fold in dosage—from 100-fold lower to 100-fold higher than the dose that would cause the effect in people of median susceptibility in an exposed population. A probit slope of 2 would suggest less, but still appreciable variability—with the thresholds of 95% of the population spread over a 100-fold range in dosage from 10-fold lower to 10-fold higher than the threshold for the median person. Such a large amount of interindividual variation could suggest appreciable risks at the much lower dosages that are present in the diets of people who consume relatively large amounts of fish with relatively large methyl mercury concentrations.

possibilities might well include the concentration of mercury at a specific sensitive time during gestation or a weighted sum of concentrations × duration over a specific set of sensitive periods. Accurate assessment of the degree of interindividual variability in susceptibility in humans, and consequent low dose risks, might well depend on quantitative measurement and modeling of the causal processes involved in this case and reanalysis of the data according to the most likely causally predictive summary measure of delivered dose. Nevertheless, if we take these results at face value, the suggestion is that a 10-fold uncertainty factor is likely to be very much less protective for some fetal/developmental effects than for effects in adults. The quantitative implications of these data for potential human risk have been extensively analyzed in a recent NAS/NRC report (IOM, 1991).

The example above illustrates the substantial opportunity to gather and analyze interindividual variability information quantitatively. In the long run, making use of the data we can gather of these types can allow us to provide quantitative estimates of potential human risks for selected noncancer effects and associated uncertainties. In the process, we will be challenged to develop innovative models of both the pharmacokinetics and the pharmacodynamics of the underlying causal processes that will be of fundamental scientific interest.

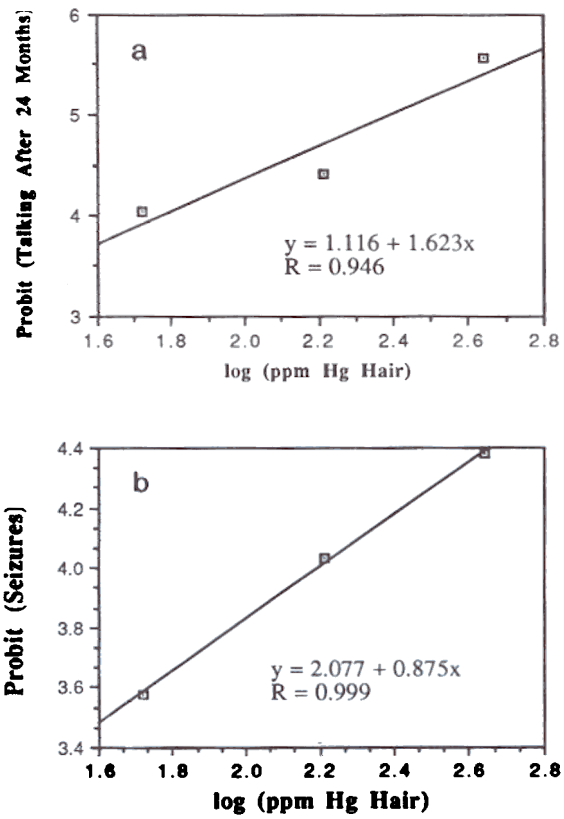

FIG. 4. Apparent log probit dose–response relationships for some fetal effects of methyl mercury—Data of Marsh *et al.* (1987). (a) Late talking. (b) Seizures.

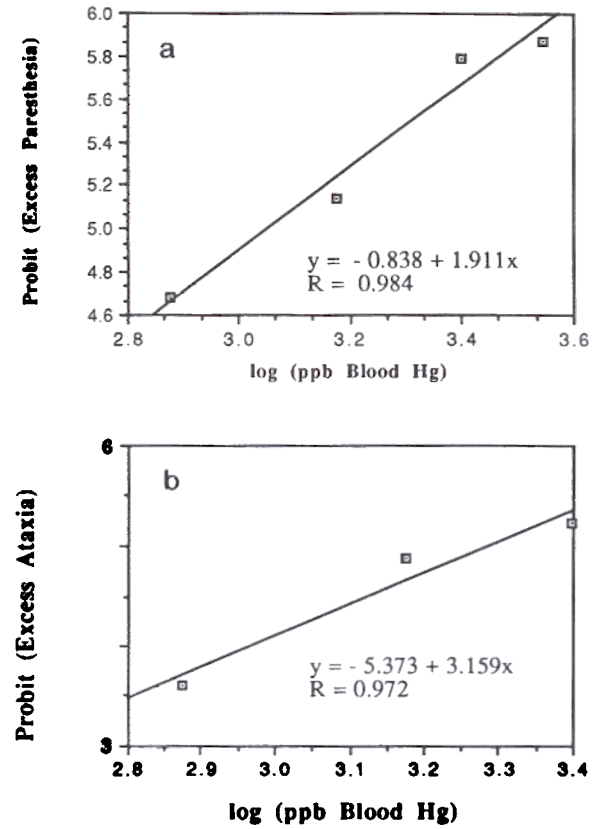

FIG. 5. Apparent log probit dose–response relationships for some effects of methyl mercury in adults—Data of Marsh *et al.* (1987). (a) Excess paresthesia over background. (b) Excess seizures over background.

References

Al-Shahristani, H., and Shihab, K. (1974). Variation of biological half-life of methylmercury in man. *Arch. Environ. Health* **28**, 342–344.

Bakir, F., Damluji, S. F., Amin-Zaki, L., Murtadha, M., Khalidi, A., al-Rawi, N. Y., Tikriti, S., Dahahir, H. I., Clarkson, T. W., Smith, J. C., and Doherty, R. A. (1973). Methylmercury poisoning in Iraq. *Science* **181**, 230–241.

Ballew, M. and Hattis, D. (1989) *Reproductive Effects of Glycol Ethers in Females—A Quantitative Analysis.* MIT Center for Technology, Policy, and Industrial Development. Report #CTPID, 89-7, July.

Clarkson, T. W., Amin-Zaki, L., and Al-Tikriti, S. K. (1976). An outbreak of methylmercury poisoning due to consumption of contaminated grain. *Fed. Proc.* **35**, 2395–2399.

Hattis, D., Erdreich, L., and Ballew, M. (1987). Human variability in susceptibility to toxic chemicals—A preliminary analysis of pharmacokinetic data from normal volunteers. *Risk Anal.* **7**(4), 415–426.

Hattis, D., and Silver, K. (1990). *Human Interindividual Variability—A Major Source of Uncertainty in Assessing Risks for Non-Cancer Health*

Effects. Presented at the Air and Waste Management Association Annual Meeting, Pittsburgh, PA, June.

Institute of Medicine (1991). *Seafood Safety*. Nat. Acad. Press, Washington, DC.

Jones, C. O., Gauld, S., Hurley, J. F., and Rickmann, A. M. (1981). *Personal Differences in the Breathing Patterns and Volumes and Dust Intakes of Working Miners*. Report to the Commission of the European Communities, Report No. TM/81/11, Environmental Branch, Institute of Occupational Medicine, Roxburgh Place, Edinburgh, Scotland.

Marsh, D. O., Clarkson, T. W., Cox, C., Myers, G. J., Amin-Zaki, L., and Al-Tikriti, S. (1987). Fetal methylmercury poisoning: Relationship between concentration in single strands of maternal hair and child effects. *Arch. Neurol.* **44**, 1017–1022.

Alternatives to the NOAEL/Uncertainty Factors (UF) Approach for Quantitative Noncancer Risk Assessment

(Carole A. Kimmel and Hal Zenick† U.S. Environmental Protection Agency, *Office of Health and Environmental Assessment, 401 M Street, SW, Washington, DC 20460, and †Health Effects Research Laboratory, Research Triangle Park, North Carolina 27711)*

The current approach to risk assessment for noncancer health effects involves the determination of a NOAEL from standard studies often with a limited data set (usually three dose levels and a control) and the application of an uncertainty factor to derive a RfD. This approach has several limitations, including: (1) the NOAEL/UF approach focuses only on the dose that is the NOAEL, which must be one of the experimental doses; (2) it ignores the shape of the dose–response curve; (3) small studies result in higher NOAELs; and (4) it does not provide an estimate of risk at a given dose, especially above the RfD.

A number of alternative approaches utilizing dose–response modeling have been proposed in the literature. The most immediately applicable approach appears to be the use of dose–response modeling and estimation of a "benchmark dose" as proposed by Crump (1984) for derivation of the RfD. The application of this approach and others is being investigated in various projects sponsored by EPA.

In the EPA's recent Guidelines for Developmental Toxicity Risk Assessment (EPA, 1991), the first of several guidelines for noncancer risk assessment developed by the Agency, the benchmark dose approach was proposed for use. The guidelines indicated that EPA is evaluating the use of the benchmark dose as an additional approach for more quantitative dose–response evaluation when sufficient data are available. The guidelines also indicate that benchmark doses will begin to be calculated for comparison with the dose–response data from studies, but that there is insufficient experience with the application of this approach at present to use it for calculating the RfD. Once sufficient information is gained from ongoing studies, EPA's intent is to write guidance on the application of the benchmark dose approach for calculation of the RfD, as an amendment to current risk assessment guidelines.

The benchmark dose approach was originally proposed in 1984 by Crump as a simple, but important, improvement in the estimation of the RfD. The benchmark dose is the lower confidence limit on a dose corresponding to an increase in the incidence of an effect at a particular risk level, e.g., the LED_{10} is the lower confidence limit on a dose that is effective in producing a 10% increase in response (Fig. 6). Uncertainty factors may then be applied to the LED_{10} to calculate the RfD. Since the NOAEL theoretically can fall anywhere between zero and an incidence just below that detectable as an increase above control levels (usually in the range of 7–10% for quantal data), the benchmark dose would provide a common starting point for applying uncertainty factors and would result in RfDs with more comparable levels of protection than when NOAELs are used. Which benchmark dose to use has been discussed in the literature. For example, Crump (1984) and Kimmel and Gaylor (1988) discussed the use of the LED_{10} since it usually falls within the experimental range. If enough data are available at the lower end of the dose–response range, it is also possible to calculate an LED_{05} or an LED_{01} (as discussed by Gaylor, 1989, and Chen and Kodell, 1989), values that would be closer to a true NOAEL and would require application of fewer uncertainty factors than the LED_{10}.

Various mathematical models have been proposed for

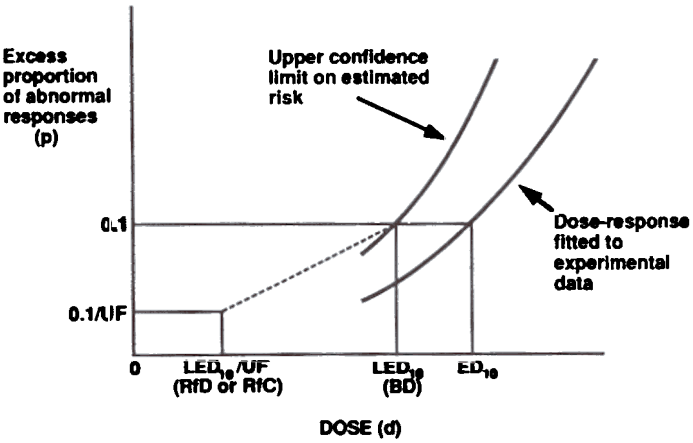

FIG. 6. This graphical illustration of the benchmark dose approach is based on Crump (1984) and Kimmel and Gaylor (1988). The benchmark dose (BD) is derived by modeling the data in the observed range, selecting an incidence level within or near the observed range (e.g., the effective dose to produce a 10% increased incidence of response, the ED_{10}), and determining the upper confidence limit on the model. The upper confidence value corresponding to, for example, a 10% excess in response is used to derive the BD which is the lower confidence limit on dose for that level of excess response, in this case, the LED_{10}. The RfD or RfC may be estimated by applying uncertainty factors (UF) to the BD.

use in the benchmark dose approach and, theoretically, the choice of the model should not be critical, as long as it fits the data well, since estimation is within the observed dose range for most quantal endpoints. Thus, the assumption of a threshold would not be of concern since risk would not be extrapolated to low levels of exposure. If, however, there are biological reasons to incorporate particular factors in the model (e.g., intralitter correlations), these should be included to account, as much as possible, for variability in the data.

The major study underway at this time is being conducted in collaboration with Dr. Elaine Faustman (University of Washington) and Bruce Allen (ICF Clement), to apply several available models to actual data sets. For this purpose, over 250 data sets from standard developmental toxicity studies were obtained from EPA labs, the National Toxicology Program, Argus Research Labs, WIL Research Labs, Mobil Research Labs, and the International Research and Development Corp. Two generic models (Crump, 1984) and three developmental toxicity-specific models (Rai and Van Ryzin, 1985; Kupper *et al.*, 1986; Kodell *et al.*, 1991) are being applied to the data sets available. The objectives of the study are

- to evaluate the fidelity of outputs and consistency of the BMD estimates,
- to contrast these estimates with the NOAELs,
- to evaluate the impact of the threshold parameter and litter size adjustments on the estimates, and
- to assess the impact of the modeling approaches on various aspects of study design, by using data simulations with different group sizes, number of dose levels, etc.

Data from each study were broken down into several subsets (e.g., dead and resorbed/litter, fetuses with external, visceral, or skeletal malformations, total affected (dead + resorbed + malformed), and other combinations). For each endpoint in the analysis, data were expressed either on a litter basis (i.e., affected litter based on one or more fetuses affected in the litter), or on an individual fetus basis (i.e., proportion of fetuses within the litter affected). The former expression is a quantal measure, while the latter is considered a quasicontinuous expression of the data. NOAELs were calculated for all of the endpoints on the basis of both quantal and continuous measures of the data, and modeling was done on the same basis.

Preliminary data have been presented (Allen *et al.*, 1991, 1992) which show that a comparison of the benchmark dose for a 5% response level with the NOAEL is similar for the quasicontinuous expression of the data, but the benchmark dose tends to be lower than the NOAEL for the quantal expression of the data. For the developmental toxicity-specific models, the BMD_5 and the NOAELs were similar (58–62% within a 2-fold range), but there were some cases in which the two were as much as 10-fold different. Further analyses of these data is continuing.

Another issue that must be considered for application of the benchmark dose is its use for continuous data such as weight, biochemical measures, etc. One possibility is to categorize data as to whether it falls above or below a certain level defined as adverse. Another approach has been suggested by Gaylor and Slikker (1990), in which the continuous data are modeled, a value is chosen as an adverse response level, and then assuming a certain distribution around the mean, the proportion of the populations outside the "normal" range can be estimated. This approach utilizes the power of continuous data, while providing a way to estimate the probability of an effect within a population.

One other aspect of modeling for developmental toxicity involves the interrelationship of multiple outcomes in developmental toxicity (and other noncancer) studies. This project is being conducted in conjunction with Dr. Louise Ryan at Harvard School of Public Health. In a recent paper (Ryan *et al.*, 1991), the process of development was described as a continuum of events, leading to resorption or death, or to a viable fetus at term, and for that fetus, its malformation status and weight. These manifestations are routinely assessed in standard developmental toxicity studies, but the results are almost always handled as independent experimental outcomes. Assessments of multiple outcomes are complicated by the presence of competing risks (e.g., implants that die during organogenesis cannot go on to express malformations at term). However, it is clear that there are correlations between some of these possible outcomes and these can be incorporated into models that better characterize the nature of the dose–response relationship.

TABLE 2
Comparison of the NOAEL and Benchmark Dose Approaches

NOAEL	Benchmark dose
Advantages	
Simple	Uses all of the data
Identification of the critical effect based on the lowest LOAEL	Takes into account the variability in the data
	Provides a consistent basis for calculating the RfD
Limitations	
Focuses only on the dose that is the NOAEL	More laborious
Ignores the shape of the dose–response curve	Only single or combined endpoints can be modeled
Smaller studies result in higher NOAELs	Not all data sets will be amenable to modeling

In summary, the advantages and disadvantages of the benchmark approach and the NOAEL approach are shown in Table 2. Although it is difficult to provide proof of one approach being more correct than the other, the benchmark approach provides some inherently obvious advantages over the NOAEL. There are still a number of decisions to be made in the final application of the benchmark dose approach, some of which have been pointed out here, but efforts toward more quantitative methods are being strongly encouraged by the EPA, its Science Advisory Board, and others.

References

Allen, B. C., Van Landingham, C., Howe, R. B., Kavlock, R. J., Kimmel, C. A., and Faustman, E. M. (1991). *Comparison of Approaches to Risk Assessment for Developmental Toxicants.* Presented at the Society for Risk Analysis, Baltimore, December 8.

Allen, B. C., Van Landingham, C., Howe, R. B., Kavlock, R. J., Kimmel, C. A., and Faustman, E. M. (1992). Dose-response modelling for developmental toxicity. *The Toxicologist* **12**, 300.

Chen, J. J., and Kodell, R. L. (1989). Quantitative risk assessment for teratological effects. *J. Am. Stat. Assoc.* **84**, 966–971.

Crump, K. S. (1984). A new method for determining allowable daily intakes. *Fundam. Appl. Toxicol.* **4**, 854–871.

Gaylor, D. W. (1989). Quantitative risk analysis for quantal reproductive and developmental effects. *Environ. Health Perspect.* **79**, 243–246.

Gaylor, D. W., and Slikker, W., Jr. (1990). Risk assessment for neurotoxic effects. *Neurotoxicology* **11**, 211–218.

Kimmel, C. A., and Gaylor, D. W. (1988). Issues in qualitative and quantitative risk analysis for developmental toxicology. *Risk Anal.* **8**, 15–20.

Kodell, R. L., Howe, R. B., Chen, J. J., and Gaylor, D. W. (1991). Mathematical modelling of reproductive and developmental toxic effects for quantitative risk assessment. *Risk Analysis* **11**, 583–590.

Kupper, L. L., Portier, C., Hogan, M. D., and Yamamoto, E. (1986). The impact of litter effects on dose-response modeling in teratology. *Biometrics* **42**, 85–98.

Rai, K., and Van Ryzin, J. (1985). A dose response model for teratological experiments involving quantal responses. *Biometrics* **41**, 1–10.

Ryan, L. M., Catalano, P. J., Kimmel, C. A., and Kimmel, G. L. (1991). Relationship between fetal weight and malformation in developmental toxicity studies. *Teratology* **44**, 215–223.

Exposure-Response Analysis: Modeling Severity against Concentration and Duration (Daniel J. Guth, Richard C. Hertzberg, and Annie M. Jarabek, U.S. Environmental Protection Agency, Environmental Criteria and Assessment Office, Research Triangle Park, North Carolina, and *Cincinnati, Ohio)*

Current approaches to risk assessment for noncancer endpoints have relied on the determination of an appropriate experimental NOAEL and the consistent application of UF to derive an oral RfD or inhalation reference concentration (RfC) to serve as an estimated subthreshold exposure for humans. Although all available data are considered, the focus of the quantitative analysis is on a single endpoint and exposure level. Concentration–response modeling, recently referred to as the benchmark approach, has been proposed as a logical improvement of the NOAEL/UF approach (Crump, 1984). The benchmark approach as defined in this presentation is the use of a specific mathematical model (e.g., Weibull, polynomial) to determine a concentration (or applied dose), and its lower confidence bound, associated with a predefined effect measure (e.g., 10% response) as the benchmark. The benchmark concentration has the advantages that it utilizes more information from the dose–response curve, is less influenced by experimental design (e.g., dose spacing), and is sensitive to the influence of sample size. Application of this approach has been demonstrated for developmental endpoints (Allen *et al.*, 1992; Kimmel and Zenick, 1992), but it has yet to be applied widely to other noncancer outcomes.

The benchmark procedure provides an estimate of an exposure that would produce a fixed incidence or magnitude of response. The definition is thus critically dependent on the choice of an effect measure, and this choice must be based on a judgement of the response severity. The benchmark concentration method has not previously been adapted to include terms for both concentration and duration of exposure and is not useful for quantitative evaluation of the role of exposure duration. An additional shortcoming of the benchmark concentration approach is its requirement for effect measures expressed as dichotomous or continuous variables. Categorical effect measures, in contrast, are frequently encountered in the toxicological literature, as in descriptions of histopathology. Some studies are not explicitly designed to examine the concentration–response relationship, such as single concentration time course or mechanistic studies. Such data are often not amenable to quantitative concentration–response analysis.

Data which are not amenable to concentration–response analysis can be expressed categorically and analyzed as such. This paper demonstrates the application of a categorical analysis method which allows the combination of categorical data and models the relationship between the severity of the effect and the concentration and duration of inhalation exposure. A thorough literature search was performed and studies involving acute (<24 hr) inhalation exposures to methyl isocyanate (MIC) or dichloromethane (DCM) were reviewed. These chemicals were selected in order to include both a direct acting respiratory toxicant and a chemical which acts at a site remote from the portal of entry. The criteria for a minimum data base for the benchmark approach adopted for the purpose of this study were a requirement for at least five animals per group and at least two exposure levels with nonzero response levels for dichotomous or continuous endpoints other than lethality. Most of the studies were not useful for benchmark concentration analysis based on these criteria. Of the 23 studies of MIC

containing information on acute effects, only 2 studies contained data amenable to the benchmark approach. Similarly, a total of 30 studies of DCM were reviewed, and only 2 were amenable to the benchmark approach, using the same criteria. Since such a limited sample of the available data could be used for benchmark concentration analysis, a more useful analytical approach is needed.

To address the difficulties mentioned above, a collapsed categorization scheme was used in quantitative exposure–severity analysis, with severity category as the dependent variable and with concentration and exposure duration as independent variables (Guth *et al.*, 1991). The severity scheme consists of three categories representing no observed adverse effect levels, adverse effect levels (AEL), and lethality. The response data from all studies were categorized according to the reported effects. More complicated severity ranking schemes that can be applied become contentious due to the difficulty in equating severity of effects measures across target organs, endpoints, and species.

Regression analysis is performed using the logistic model for ordinal data with both exposure concentration and duration as independent variables (SAS Institute, 1985). The form of the model is

$$\ln(p/1 \quad p) = A_i + B_1 \ln (\text{concentration} + B_2(\text{duration}),$$

where p is the probability that, at a given concentration and duration of exposure, severity will be less than or equal to the severity category with rank = i, and A and B are estimated model parameters. The model is solved for $P = 1 - p$, or the probability that, at a given concentration and duration, the severity will be greater than the severity category with rank = i. As will be shown, this model output is readily interpreted in the context of risk assessment. The regression analysis assumes constant slope parameters, hence the values of B_1 and B_2 are constant across severity categories. The order or rank of the categories are used, rather than the numerical values. The logistic model was selected for computational convenience, and ongoing work will evaluate the assumptions of linearity and model form.

The form of the input data depends on the available data. When the data are such that it can be determined how many animals from each experimental group (exposed at a particular concentration and duration) fell in each severity category, the individual animal data can be used in the model. For the examples of MIC and DCM, this information is not available for most studies. In this case, all data are entered by assigning a severity category to the experimental group and using information at the group level as the input data. Each data point then represents a unique experimental group, as shown in Fig. 7, and in the regression analysis the data points are weighted by the number of animals in the group. The latter approach is similar to assuming that all animals in the group responded identically. Although not an optimal estimate, some weighting based on group size is a necessary assumption when the data are presented as such, or when a continuous variable is presented as a group mean value.

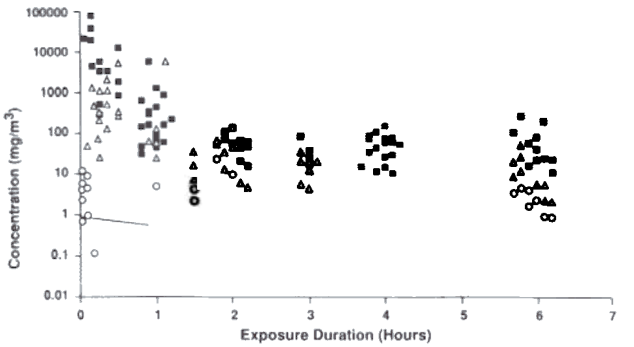

FIG. 7. Categorical data from published results on methyl isocyanate for exposures of less than 8 hr in duration and shown as NOAEL (circles), AEL (triangles), or lethality (squares). The maximum likelihood model fit is shown by the line representing the model prediction of $p = 0.1$ that severity is greater than the NOAEL category at the corresponding exposure concentration and duration.

The regression analysis can then be performed on the entire data set for the chemical and duration of interest, or on specified subsets of the data which may be identified based on a common target organ or endpoint, or based on study quality criteria. The data and model output for MIC are shown for all data from exposures less than 8 hr in duration (Fig. 7). Each point on the graph represents a unique experimental group, and data from 23 studies are represented. Exposure concentrations from animal studies are calculated as human equivalent concentrations as described in U.S. EPA (1990). Inclusion of all data in the analysis leads to complications due to combination of data from different studies, species, and endpoints. In the MIC data base, almost all data are on respiratory system effects, somewhat reducing the confounding due to combination of endpoints. On a log-linear axis, the model prediction for severity category = NOAEL is linear and is shown for MIC in Fig. 7. The line represents the combination of concentrations and durations that result in a probability of 0.1 that the NOAEL category is exceeded (i.e., the predicted effect would be in the "adverse" range or higher). An increase in either exposure concentration or duration would result in a higher probability of exceeding the NOAEL category. Figure 8 shows the results of this analysis when the data showing lethality are not included in the analysis and only respiratory effects are considered. The fit to the data is visually improved, but comparison with Fig. 7 shows that the resulting model prediction is within a factor of 3 in concentration from lethal effects at some exposure durations. This em-

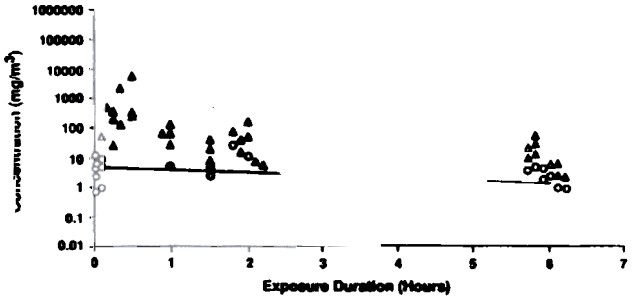

FIG. 8. Categorical data from published results on methyl isocyanate-induced sublethal respiratory effects for exposures of less than 8 hr in duration and shown as NOAEL (circles) or AEL (triangles). The maximum likelihood model fit is shown by the line representing the model prediction of $p = 0.1$ that severity is greater than the NOAEL category at the corresponding exposure concentration and duration.

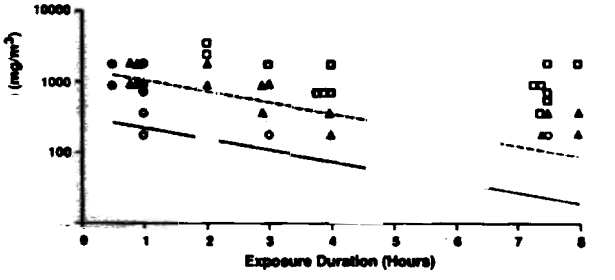

FIG. 10. Categorical data from published results on carboxyhemoglobin levels in exposed human volunteers for exposures of less than or equal to 8 hr in duration and shown as <2% COHb (circles), 2–5% COHb (triangles), and >5% COHb (squares). The maximum likelihood model fit is shown by the lines representing the model prediction of $p = 0.1$ that COHb concentration is greater than 2% (solid line) and for COHb concentration greater than 5% (dashed line) at the corresponding exposure concentration and duration.

phasizes the need to examine all available data when interpreting the result of the analysis.

Data from the regression analysis of dichloromethane are shown in Figs. 9 and 10. Figure 9 shows the data and model results for DCM-induced central nervous system effects in animals and humans. The line shown is the model estimate of the probability of 0.1 that the severity will exceed the NOAEL category. Analogous information is presented in Fig. 10 for increased carboxyhemoglobin (COHb) after DCM exposure. In the case of Fig. 10, the data analyzed are COHb measurements from exposed human volunteers. The analysis still combines data from different studies, but the data are quite similar otherwise. In this example, the categories were defined based on the COHb levels, with category 1 being <2% COHb, category 2 being 2–5% COHb, and category 3 being >5% COHb. Specific definition of the categories in terms of the effect allows more specific interpretation of the model output. Thus the model output for the NOAEL category is the probability of 0.1 that the COHb level will exceed 2% at the corresponding concentration and duration.

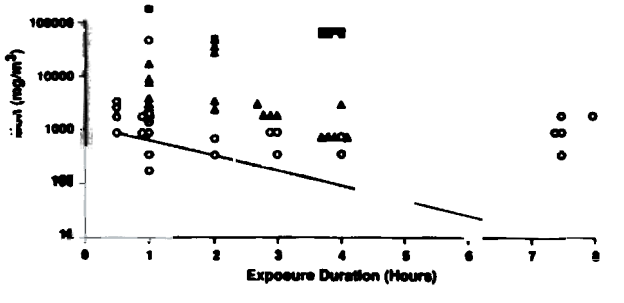

FIG. 9. Categorical data from published results on central nervous system effects of dichloromethane for exposures of less than or equal to 8 hr in duration and shown as NOAEL (circles), AEL (triangles), or lethality (squares). The maximum likelihood model fit is shown by the line representing the model prediction of $p = 0.1$ that severity is greater than the NOAEL category at the corresponding exposure concentration and duration.

Alternatives to the NOAEL/UF approach such as the benchmark procedure and categorical analysis are under development with the objective of improving the use of available data in risk assessment. These are promising approaches for improvement of the quantitative basis of concentration–response analysis but require further definition in areas such as minimum data base, model selection, criteria for model fit, and identification of the model output to be used. The approach to integration of these modeling analyses across multiple endpoints for application to risk assessment also remains to be defined. The application of the results of these analytical techniques to risk assessment would presumably include application of uncertainty factors. Guidelines for application of uncertainty factors with these approaches have yet to be determined. Based on development in these areas, the choice of the quantitative approach that will be applicable to a particular risk assessment need will be determined largely by the amount and type of toxicological data available. The categorical analysis approach demonstrated here offers the advantages of allowing the use of data which is not otherwise amenable to quantitative concentration–response analysis, such as categorical data and data from single-dose studies, and of incorporating both concentration and duration of exposure as explanatory variables. Various types of data (dichotomous, continuous, categorical) can be considered simultaneously by converting each to a categorical descriptor. This approach will be particularly useful when categorical data at the level of the experimental group makes up a substantial part of the available data, as in the case of MIC, and when the nature of the risk assessment requires that a range of durations be considered.

References

Allen, B. C., Van Landingham, C., Howe, R. B., Kavlock, R. J., Kimmel, C. A., and Faustman, E. M. (1992). Dose-response modelling for developmental toxicity. *Toxicologist* **12**, 300.

Crump, K. (1984). A new method for determining allowable daily intakes. *Fundam. Appl. Toxicol.* **4**, 854–871.

Guth, D. J., Jarabek, A. M., Wymer, L., and Hertzberg, R. C. (1991). *Evaluation of Risk Assessment Methods for Short-term Inhalation Exposure.* Proceedings of the 84th Annual Meeting of the Air and Waste Management Association, Vancouver, British Columbia, June 16–21. Paper No. 91-173.2.

Kimmel, C. A., and Zenick, H. (1992). Alternatives to the NOAEL/UF approach for quantitative noncancer risk assessment. *Toxicologist* **12**, 23.

SAS Institute Inc. (1985). *SAS Procedure Guide,* Version 6, 4th. ed., SAS Institute Inc., Cary, NC.

U.S. Environmental Protection Agency (1990). *Interim Methods for Development of Inhalation Reference Concentrations.* Office of Research and Development, Washington, DC. EPA/600/8-90/066A.

Biologically Based Risk Assessment for Chloroform Hepatotoxicity (R. B. Conolly)

A major goal of quantitative risk assessment is to accurately predict the shape of the exposure–response curve for people, thereby allowing direct translation from exposure to risk of toxic effect. Accuracy reflects the degree to which the predicted exposure–response curve reproduces the actual curve for the target population. When the prediction is accurate, it is possible to identify exposure levels causing no risk, or an acceptable level of risk, to public health and also avoid unnecessarily stringent regulation and associated adverse economic impacts. Perfectly accurate risk assessment, i.e., identification of true risk, may be an unattainable goal but it is possible to identify research strategies to increase the accuracy of risk assessment. The following describes a generic strategy for improving the accuracy of risk estimates for noncarcinogenic endpoints using chloroform hepatotoxicity as an example.

Every toxic response has an exposure–response curve. The shape of this curve is determined by:

(1) The relationship between exposure and dose in the target tissue,

(2) The relationship between the parent toxicant in the target tissue and its biologically active form, e.g., a metabolite, which initiates the toxic effect, and

(3) The sequence of events triggered by the biologically active form which culminates in the actual toxic effect.

The accuracy of a predicted exposure–response curve will be related to the degree to which it is based on relevant data describing each of these three processes. Because such data are typically acquired in laboratory animals, we also encounter the problem of how the exposure–response description based on this information scales from animal to people.

Corley *et al.* (1990) described a physiologically based pharmacokinetic (PBPK) model for chloroform which draws on quantitative, physiological, and chemical-specific information to depict the processes controlling the dose of chloroform to liver and kidney, the main target organs for the acute toxic effects of chloroform. For the work reported here, this PBPK model was extended to describe cytolethality based on a modification of the approach of Reitz *et al.* (1990). These authors evaluated both total amount of chloroform metabolized and rate of metabolism as alternative measures of the biologically active form of chloroform, i.e., as the expression for target tissue dose that best predicted observed cell killing. Rate of metabolism was a much better predictor of both quantitative histopathology and chloroform-stimulated regenerative cellular replication than amount metabolized (Reitz *et al.*, 1990).

The relationship of chloroform cytolethality to rate of chloroform metabolism is biologically plausible: (1) The main metabolites of chloroform are phosgene and HCl (Pohl, 1979), both of which have very short biological half-lives, and (2) phosgene should form primarily labile macromolecular adducts. Thus, the major metabolite and macromolecular adducts resulting from chloroform/phosgene exposure are expected to be transient. In the previous cytolethality model (Reitz *et al.*, 1990), hepatocytes in the liver have a normal distribution of sensitivities to killing by chloroform—some cells will be killed by smaller amounts of the dose surrogate than required to kill others. The parameters describing the normal distribution were optimized to fit the cytolethality data. In the current work, the normal distribution of cell sensitivities was not used. Instead, an empirical function relating dose surrogate to cell killing was constructed (Fig. 11). This avoids assumptions about the distribution of cell sensitivities and allows a more accurate simulation of the cell killing data (Fig. 12).

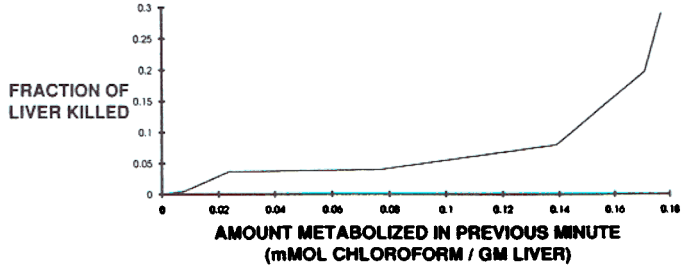

FIG. 11. Empirical relationship between chloroform metabolism and hepatotoxicity. Function, used in the chloroform exposure–toxic response model, describing relationship between target tissue dose surrogate and cell killing. During a simulated chloroform exposure, the value of the dose surrogate is calculated at each time point t and the function specifies the corresponding amount of cytolethality. Function was obtained by adjusting its shape to optimize the fit of the simulated cell killing to the quantitative histopathology and cell replication data described by Reitz *et al.* (1990).

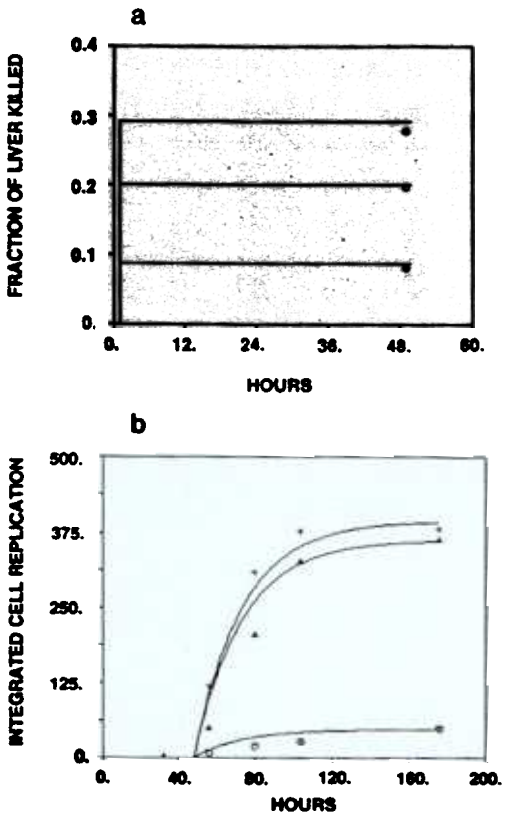

FIG. 12. (a) Actual (circles) and simulated (solid lines) cell killing in male B6C3F1 mice after single corn oil gavage exposure to indicated doses of chloroform. Dead cell data are from Reitz *et al.* (1990). Simulations show death occurring quickly after gavage, even though histopathological changes are not apparent for about 24 hr. In the simulation model, cell killing is linked to rate of chloroform metabolism, which is maximal shortly after gavage. (b) Actual (O, ▲, +) and simulated (solid lines) cell replication after male B6C3F1 mice were exposed to chloroform by inhalation at indicated concentrations once for 6 hr. Data (Reitz *et al.*, 1990) were originally reported as the ratio of cell replication of exposed to control groups and were integrated over time for representation shown here.

The dose surrogate used in the empirical linking function (Fig. 11) was the amount of chloroform metabolized in the previous minute. As noted above, available data indicate that chloroform-mediated cell killing is closely related to rate of metabolism. Although not described here, equally good fits to the cytolethality and cell replication data of Reitz *et al.* (1990) were obtained when intervals of up to 15 min were used for the dose surrogate.

One advantage of PBPK models is their ability to account for pharmacokinetic behavior by different routes of administration (e.g., Ramsey and Andersen, 1984). The cytolethality model described here allows these kinds of comparisons for cell killing (Fig. 13). The greatest toxicity is predicted for corn oil gavage, with drinking water the least toxic and inhalation in between. A major determinant of this behavior of the model is linkage of cell killing to a dose surrogate reflecting rate of chloroform metabolism. Oral gavage dosing briefly achieves high rates of metabolism as the whole dose is given as a single bolus. Rate of metabolism for drinking water exposure, described in the PBPK model as a continuous infusion to the liver over time, will be lower than gavage for an equivalent total exposure. Inhalation exposures extend over some period of time. The maximum rate of metabolism achieved by inhalation will also be less than for an equivalent total exposure by gavage.

The simulations depicted here (Figs. 12 and 13) are for B6C3F1 mice. Human risk assessment requires extrapolation of the exposure–toxic response description to people. A considerable amount of work has been done on interspecies scaling of PBPK models (Rowland, 1985; Krishnan and Andersen, 1991), but less is known about scaling of response, i.e., in this case, the relationship between target tissue dose surrogate and cell killing. Recently, the *in vivo* kinetics of furan metabolism in the F344 rat were accurately predicted using freshly isolated hepatocytes (Kedderis *et al.*, 1992). The method is applicable to other chemicals and may be extensible to *in vitro* assessment of cytotoxicity. *In vitro* methods like this, carefully calibrated against parallel *in vivo* studies, hold promise for obtaining meaningful estimates of parameter values for human PBPK and tissue response models.

One possible scaling approach for the exposure–toxic response model for chloroform is to change the PBPK component as described by Corley *et al.* (1990) but retain the tissue response component as developed for the mouse. This hybrid model for human risk assessment then incorporates mechanistic information, as delineated above, while at the same time minimizing the use of response assumptions that cannot readily be tested against data.

In summary, a strategy for increasing the accuracy of noncancer human risk assessment is described using chloroform as an example. The starting point is a physiologically based pharmacokinetic model for chloroform that is extended to include linkage between chloroform dose to the

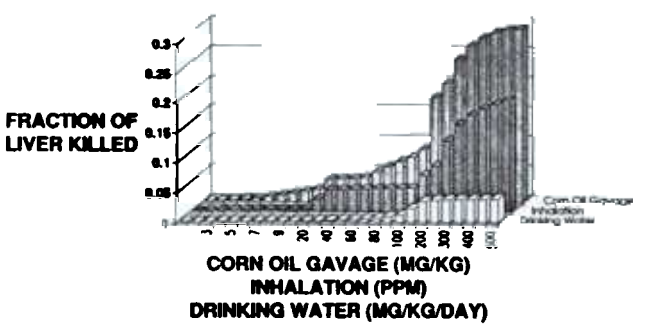

FIG. 13. Simulation of expected hepatic cytolethality in male B6C3F1 mice for drinking water, inhalation, and corn oil gavage exposures to chloroform. Simulated exposure conditions were a 24-hr drinking water exposure, a single 6-hr inhalation, or a single oral gavage.

liver and cytolethality. The resulting exposure–toxic response model can be used to predict equivalently toxic exposures by different exposure routes in mice. Finally, a strategy for scaling the model from mice to humans to allow simulation of expected hepatotoxicity in people exposed to chloroform is described. This scaled description could be used to predict human risks of hepatotoxicity from chloroform exposure. A safety factor might still be retained to account for uncertainties related to model structure and heterogeneity of the target population, etc. This exposure–toxic response model for chloroform is not complete in a mechanistic sense, particularly with respect to the mechanism by which chloroform kills cells. It does, however, use a great deal of the available mechanistic data and, at the same time, identifies data gaps, thereby guiding the design of future research.

References

Corley, R. A., Mendrala, A. L., Smith, F. A., Staats, D. A., Gargas, M. L., Conolly, R. B., Andersen, M. E., and Reitz, R. H. (1990). Development of a physiologically based pharmacokinetic model for chloroform. *Toxicol. Appl. Pharmacol.* **103,** 512–527.

Kedderis, G. L., Murphy, J. E., Batra, R., Held, S. A., Carfagna, M. A., and Gargas, M. L. (1992). *In vivo* and *in vitro* kinetic analysis of furan biotransformation by F-344 rats. *The Toxicologist* **12,** 348. [Abstract 1360]

Krishnan, K., and Andersen, M. E. (1991). Interspecies scaling in pharmacokinetics. In *New Trends in Pharmacokinetics* (A. Rescigno and A. K. Thakur, Eds.), pp. 203–226. Plenum, New York.

Polh, L. (1979). Biochemical toxicology of chloroform. In *Reviews in Biochemical Toxicology* (Hodgson, Bend, and Philpot, Eds.), Vol. 1, pp. 79–107. Elsevier, New York.

Ramsey, J., and Andersen, M. E. (1984). A physiologically based description of the inhalation pharmacokinetics of styrene in rats and humans. *Toxicol. Appl. Pharmacol.* **73,** 159–175.

Reitz, R. H., Mendrala, A. L., Corley, R. A., Quast, J. F., Gargas, M. L., Andersen, M. E., Staats, D. A., and Conolly, R. B. (1990). Estimating the risk of liver cancer associated with human exposures to chloroform using physiologically based pharmacokinetic modeling. *Toxicol. Appl. Pharmacol.* **105,** 443–459.

Rowland, M. (1985). Physiologic pharmacokinetic models and interanimal species scaling. *Pharmacol. Ther.* **29,** 49–68.