Comparative Environmental Risk Assessment

Edited by
C. Richard Cothern

1992

LEWIS PUBLISHERS
Boca Raton Ann Arbor London Tokyo

CHAPTER 7

Noncancer Health Endpoints: Approaches to Quantitative Risk Assessment*

William Farland and Michael Dourson

ABSTRACT

Traditional approaches to health risk assessment for cancer and noncancer endpoints represent a dichotomy. Carcinogens have been treated as nonthreshold, intrinsically toxic compounds and incremental risk estimates have been calculated. Other health effects are generally thought to exhibit thresholds of response and substances producing these effects are thought to be situationally toxic. This perspective has led to the development of procedures for defining exposures (doses) for these substances at or below which no adverse effect is likely to occur. This so-called reference dose (RfD) is often used to indicate an exposure that is considered to be safe, i.e., without adverse effects, and therefore without risk. On the other hand, exposures above this level may represent a toxicological hazard with or without significant risk. Quantitative rankings of environmental problems would be improved by application of procedures to estimate risks for noncancer endpoints at exposures above the reference dose. No well-accepted approaches to this problem are currently available. Research to evaluate a number of statistically and/or biologically based methods for estimating risk under this situation will be described and implications for improved ranking of environmental toxicants will be discussed.

INTRODUCTION

Based in part on the seminal publication on risk assessment by the National Academy of Sciences (NAS, 1983), and in part on previous work in this area of science by the World Health Organization (WHO) and the U.S. Food and Drug Administration (FDA), the Environmental Protection Agency (EPA) has

* The views in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

been developing methods to clarify certain aspects of its risk assessments and to ensure consistent application of risk assessment principles across its programs. While hazard identification and risk characterization for both cancer and noncancer endpoints involve a similar weight of the evidence approach dose-response assessments have been handled differently. In the area of noncancer health risk, for example, the EPA has published reports on the reference dose (RfD) and reference concentration (RfC) as alternatives to the acceptable daily intake (ADI) (U.S. EPA, 1990a; Barnes and Dourson, 1988; Jarabek et al., 1989) and formed an RfD, RfC Work Group to review its dose-response assessments for noncancer effects. This risk information has been summarized and placed on the Agency's Integrated Risk Information System (IRIS) (U.S. EPA, 1990b). In addition, the Agency has developed several other quantitative approaches for assessing toxicological data, such as the benchmark dose (Crump, 1984; Dourson et al., 1985; Dourson, 1986; Kimmel and Gaylor, 1988; Kimmel 1990) and categorical regression (Hertzberg and Miller, 1985; Hertzberg, 1989).

The purpose of this manuscript is to describe some of the EPA's quantitative procedures in formulating judgments concerning the nature and magnitude of the noncancer hazard and risk, and to highlight new knowledge and new assessment methods. As new quantitative approaches are developed for noncancer endpoints, we also raise the question as to the usefulness of maintaining the dichotomy between the quantitative assessment of cancer and noncancer endpoints.

THE CONCEPT OF THRESHOLD

The common approach to assessing the risks associated with noncancer and nonmutagenic toxicity is generally different from that used to assess the potential risks associated with carcinogenicity or mutagenicity. It is most often assumed that a small number of molecular events can evoke carcinogenic and/or mutagenic changes in a single cell, which can lead to self-replicating damage. Traditionally, this approach has led to these endpoints being considered nonthreshold effects, since there is presumably no level of exposure that does not pose a small, but finite, probability of generating a response. It is most often assumed that noncarcinogenic and/or nonmutagenic effects have a threshold, that is, a dose level below which a response is implausible, because homeostatic, compensating, and adaptive mechanisms in the cell protect against toxic effects.

This threshold concept is important in many regulatory contexts. The individual threshold hypothesis holds that some exposures can be tolerated by an organism with essentially no chance for expression of a toxic effect. Further, risk management decisions frequently focus on protecting the more sensitive members of a population. In these cases efforts are made to keep

exposures below the more sensitive subpopulation threshold, although it is recognized that hypersensitivity and chemical idiosyncrasy (i.e., genetically determined abnormal reactivity to a chemical) may exist at yet lower doses. This concept and focus have resulted in the development of quantitative risk assessment models (or approaches) that estimate subthreshold doses, such as ADI, RfD, and RfC. The scientific community has identified certain limits on some of these approaches: these models cannot make fuller use of available toxicity data in order to estimate possible health risks at doses higher than the RfD, for example, because they generally focus only on the no observed adverse effect level (NOAEL) of the critical effect. As a result, several quantitative procedures have been developed by EPA scientists and others, two of which are described in this paper. These procedures do not obviate the threshold concept, but rather make fuller use of the available toxicity data in estimating low adverse effect levels. In this capacity, these models might be useful to estimate the potential health risk for chemicals that cause cancer by overcoming homeostatic control on biological processes that leads to irreversible events in the multiple process which leads to tumor formation (see, for example, Hill et al., 1989). In addition, these methods also offer the flexibility to estimate the noncancer health risk for chemicals that may not exhibit a threshold. In both capacities, the dichotomy in the quantitative procedures for cancer and noncancer endpoints lessens.

HAZARD IDENTIFICATION

Hazard identification is a necessary first step in the risk assessment of a chemical. Hazard identification involves an evaluation of (1) the appropriateness, nature, quality, and relevance of scientific data on the specific chemical; (2) the characteristics and relevance of the experimental routes of exposure; and (3) the nature and significance to human health of the observed effects. The Agency has developed guidelines that explain the process of hazard identification for developmental (U.S. EPA, 1986, 1991) and reproductive (U.S. EPA, 1988a, b) health effects.

Much of the Agency's hazard identification process for noncancer toxicity depends in part on professional judgment as to whether or not an effect or collection of effects observed at any given dose of a chemical constitutes an adverse response. Such judgment may not be easily rendered, and requires experts trained in the area.

In addition, a chemical often elicits more than one toxic effect, even in a single species, or in tests of the same or different duration or dose. After assessing the quality of each study, identifying the biological and statistical significance of observed effects, and distinguishing between reversible and irreversible endpoints, the Agency often identifies the critical effects as another part of its hazard identification. The critical effects are the first adverse

effects or their known precursors (such as mottling of teeth by fluoride as a precursor to pitting and tooth loss) that occur as dose rate increases.

Current dose-response methods described in this text and elsewhere often use the distinction of adverse/nonadverse effects and the choice of critical effects as a basis for the dose-response assessment. The judgment of whether an effect is adverse or critical may change among toxicity studies of different durations, may be influenced by toxicity in other organs or pharmacokinetics, and may differ depending on the availability of data or the shape of the dose-response curve. For example, increased liver weight due to a proliferation of smooth endoplasmic reticulum (SER) in response to an oral administration of a chemical, might be judged as *not* adverse if the parent chemical is the toxic moiety and such an increase in SER is likely to quicken its metabolism. The same increase in SER, however, might be judged as an adverse effect if a metabolite is the toxic moiety.

DOSE-RESPONSE ASSESSMENT

Empirical observation reveals that as the dosage of a chemical is increased, the toxic response generally increases. This increase occurs in both severity of the response (e.g., with new and more serious effects being observed) and intensity of the response (i.e., the magnification of an effect of given severity), and in the percent of the population affected. Such dose-response relationships are well founded in the theory and practice of toxicology and pharmacology.

Dose-response assessment follows hazard identification in the risk assessment process as defined by the NAS (1983). Dose-response assessment involves the quantitative evaluation of toxicity data to determine the likely incidence of the associated effects in humans. The information available for dose-response assessment ranges from well-conducted and controlled studies on human exposures, epidemiology studies with large numbers of subjects and well-characterized exposures, and supportive studies in several animal species, to a lack of human and animal toxicity data with only structure-activity relationships to guide the evaluation.

In any case, the Agency considers all pertinent studies in this process. However, only data of sufficient quality are used in the dose-response assessment of a chemical. Because we often must model animal data, the models themselves may not be good predictors of *human* risk. Assumptions can be made *a priori* or after modeling to enable the extrapolation to human risk from animal data. This aspect of dose-response modeling, however, will be problematic for the foreseeable future.

THE CHRONIC REFERENCE DOSE (RfD), REFERENCE CONCENTRATION (RfC), AND DEVELOPMENTAL TOXICITY RfD

Given at least a moderate amount of toxicity data, one risk assessment goal is to determine a level of daily exposure that is likely to be without an appreciable risk of deleterious effects during a lifetime. The Agency's oral RfD and inhalation RfC approaches strive to include scientific considerations in their determination. The history and rationale for the change in terminology and practice from ADI to RfD are described in detail in documentation supporting the Agency's Integrated Risk Information System (IRIS) (U.S. EPA 1990b; Barnes and Dourson, 1988).

The Agency defines the oral RfD as an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime (Barnes and Dourson, 1988). The oral RfD is determined by use of the following equation:

$$\text{RfD} = \text{NOAEL or LOAEL}/(\text{UF} \times \text{MF})$$

where

NOAEL = No observed adverse effect level. An exposure level at which there are no statistically or biologically significant increases in the frequency or severity of adverse effects between the exposed population and its appropriate control; *some effects may be produced at this level, but they are not considered as adverse, nor precursors to specific adverse effects.* In an experiment with several NOAELs, the regulatory focus is primarily on the NOAEL seen at the highest dose. This leads to the common usage of the term NOAEL to mean the highest exposure without adverse effect.

LOAEL = Lowest observed adverse effect level. The lowest exposure level at which there are statistically or biologically significant increases in frequency or severity of adverse effects between the exposed population and its appropriate control group.

UF = Uncertainty factor. One of several, generally tenfold, factors used in operationally deriving the RfD from experimental data.

MF = Modifying factor. An uncertainty factor to account for scientific uncertainties not covered with the traditional factors. The value of the MF is greater than zero and less than or equal to ten; the default value for the MF is one.

The Agency is also using this model for inhalation exposures and similarly defines an inhalation RfC. Interim methods for development of inhalation

reference concentrations (U.S. EPA, 1990a; Jarabek et al., 1989) require the conversion of experimental exposure NOAEL to human equivalent concentrations (e.g., to a NOAEL [HEC]) before the critical study and species are identified. The conversion is specific both to the type of inhaled agent (particle or gas) and to the observed effect (respiratory or systemic) and adjusts for the dosimetric differences between various experimental species and humans. Once the NOAEL(HEC) is identified, the same equation is used to estimate the inhalation RfC with the application of similar, although not identical, uncertainty factors.

A reference dose for developmental toxicity (RfD_{DT}) has also been proposed (U.S. EPA, 1991). The RfD_{DT} is based upon NOAEL from short durations of exposure as are typically used in developmental toxicity studies. The term RfD_{DT} is used to distinguish from the oral RfD and inhalation RfC which refer to chronic exposure situations. Uncertainty factors for developmental toxicity generally include a tenfold factor for interspecies variation and a tenfold factor for intraspecies variation; in general an uncertainty factor is not applied to account for duration of exposure. In some cases, additional factors may be applied due to a variety of uncertainties that exist in the data base. For example, the standard study design for a developmental toxicity study calls for a low dose that demonstrates a NOAEL, but there may be circumstances where a risk assessment must be based on the results of a study in which a NOAEL for developmental toxicity was not identified. Rather, the lowest dose administered caused significant effect(s) and was identified as the LOAEL. In circumstances where only a LOAEL is available, questions relative to the sensitivity of endpoints reported, adequacy of dose levels tested, or confidence in the LOAEL reported may require the use of an additional uncertainty factor of 10. The total uncertainty factor selected is then divided into the NOAEL or LOAEL for the most sensitive endpoint from the most appropriate and/or sensitive mammalian species to determine the RfD_{DT}.

The RfD, RfC, and RfD_{DT} are useful as reference points for gauging the potential effects of other doses. Doses at the RfD or less are not likely to be associated with any health risks, and are therefore assumed likely to be of little regulatory concern. In contrast, as the amount and frequency of exposures exceeding the RfD increase, the probability that adverse effects may be observed in a human population also increases. However, the conclusion that all doses below the RfD are acceptable and that all doses in excess of the RfD are unacceptable cannot be categorically stated because these models cannot effectively predict the likelihood of particular effects above the RfD, RfC, or RfD_{DT}. This is because the assumption — that the effect observed in the experimental animal species is relevant to human — may not be correct. Moreover, the precision of these estimates depends in part on the overall magnitude of the composite uncertainty and modifying factors used in their calculation. The precision at best is probably one significant figure and more generally an order of magnitude. As the magnitude of the composite uncertainty factor increases, these estimates become even less precise.

Another risk assessment goal is to determine or estimate the likely human response to various exposure levels above the RfD/RfC of a particular contaminant. In most cases a dose-response model is appropriate if sufficient data exist. In the context of risk assessment, dose-response models for noncancer endpoints are just now starting to be used. This paper highlights two new procedures and gives examples: benchmark dose and categorical regression.

The Benchmark Dose

Several limitations in the use of the RfD model have been described (Gaylor, 1983; Crump, 1984; Dourson et al., 1985; Dourson, 1986; Kimmel and Gaylor, 1988; Kimmel, 1990; Brown and Erdreich, 1989); for example, use of the NOAEL focuses only on the dose that is the NOAEL and does not incorporate information on the slope of the dose-response curve at higher doses where effects are observed, or the variability in the data. Since data variability is not taken into account, the NOAEL from a study with few animals will likely be higher than the NOAEL from a similar study, but with more animals in the same species. Additionally, the NOAEL is usually one of the experimental doses, and the number and spacing of doses in a study can influence the dose that is chosen for the NOAEL. Since the NOAEL is defined as a dose that does not produce an observable change in adverse resoonses from control levels and is dependent on the power of the study, theoretically, the risk associated with it may fall anywhere between zero and an incidence just below that detectable from control levels (usually in the range of 7 to 10% for quantal data). Crump (1984) and Gaylor (1989) have estimated the 95% upper confidence limit on developmental risk at the NOAEL for several data sets to be 2 to 6%.

The benchmark dose is a model-derived estimate of the lower confidence limit on the effective dose that produces a certain increase in incidence above control levels, such as 1, 5, or 10%. Most toxicity studies are designed to detect changes from control on the order of 10%. Some studies are so well designed they can detect yet smaller derivations from control, on the order of 5 or even 1%. Thus, in theory the use of a benchmark dose associated with a 10, 5, or 1% increased incidence over controls is congruent with the expected toxicity data. Because the application of the model is to derive an estimate of dose for a given incidence that is likely to fall within the experimental dose range and does not require extrapolation to estimates far below the experimental dose range, the uncertainty in the estimates is lessened.

The benchmark dose is derived by modeling the data in the observed range, calculating the upper confidence limit on the dose-response curve, and selecting the point on the upper confidence curve corresponding to a 10% increase in incidence of an effect, for example. Using this approach, a benchmark dose is calculated for each response that has adequate data. In some

cases, the data may be adequate to also estimate benchmark doses at lower incidences, incidence levels which may be closer to a true no effect dose. As for the choice of the appropriate model to use in deriving a benchmark dose, various mathematical approaches have been proposed for modeling toxicity data (e.g., Crump, 1984; Kimmel and Gaylor, 1988; Rai and Van Ryzin, 1985; Faustman et al., 1989). Such models may be used to calculate the benchmark dose, and choice of the model may not be critical since estimation is within the observed dose range. Thus, any model which fits the empirical data well is likely to provide a similar estimate of the benchmark dose, although if there is some biological reason to incorporate particular factors in the model (e.g., intralitter correlation for developmental toxicity data), these should be included to account as much as possible for variability in the data. The Agency is currently supporting studies to evaluate the application of several models to data sets for calculating the benchmark dose.

To illustrate and compare the benchmark dose approach with the standard RfD/RfC method, benchmark doses were estimated for five chemicals listed on EPA's IRIS (U.S. EPA, 1990b) using a linearized multistage model. These benchmark doses were compared to the NOAEL and LOAEL of the critical effect on IRIS, as shown in Table 1. For each chemical, the oral RfD on IRIS was developed from a chronic animal study using an uncertainty factor of 100, and benchmark doses were calculated for all effects that were characterized by adequate dose responses in the critical study and other studies. Benchmark doses in this evaluation were considered as the dose associated with the 95% upper confidence limit on the response at 10%. Other values are possible. The choice of 10% here is used only as an example.

Ratios of the benchmark dose (95% upper confidence limit on the response at 10%) to the NOAEL for alachlor, hexachlorobenzene, and paraquat are nearly all within an order of magnitude of the NOAEL — where a ratio of 1 indicates identical values and an order of magnitude range is defined on a log 10 basis (i.e., 3 to 0.3). Ratios of the benchmark dose to the NOAEL for pentachlorophenol all fall within two orders of magnitude of the NOAEL (i.e., between 30 and 0.03) with many values falling within one order of magnitude. Ratios of the benchmark dose to the NOAEL for dichlorvos are uniformly two or three orders of magnitude higher than the ratio of 1 (i.e., 30 to 300) with one value yet higher. The benchmark doses estimated for dichlorvos are based on a study not used to estimate the RfD because the critical study for dichlorvos did not have sufficient quantitative data in order to estimate a benchmark dose. Thus, the benchmark dose to NOAEL ratios for dichlorvos may be misleading.

Ratios of the benchmark dose to the LOAEL were also calculated. Ratios of the benchmark doses to the LOAEL for alachlor and hexachlorobenzene are all within an order of magnitude of the LOAEL. Ratios of the benchmark dose to the LOAEL for paraquat and pentachlorophenol nearly all fall within

two orders of magnitude of the NOAEL with many values falling within one order of magnitude. Ratios of the benchmark dose to the NOAEL for dichlorvos are uniformly one to two orders of magnitude higher than the ratio of 1 with one value yet higher.

Categorical Regression

The evaluation of toxicity data by categorical regression is a rather new application of an established statistical method. As described more extensively in Hertzberg (1989, 1991), "The underlying assumption for regression on ordered categories is that the response variable is distributed across the response categories according to a multinomial distribution, with different distributional parameters for each dose (McCullagh and Nelder, 1983)." For risk assessment, the response categories are progressive toxic effects, and the explanatory variables are usually dose (or exposure level) and duration of exposure. The probabilistic output of such a model is to estimate the risk of an effect worse than a given category, given a particular dose and duration. Several publications describe results with this model (Guth et al., 1991; Hertzberg and Wymer, 1991; Knauf and Hertzberg, 1990). In this example, the logistic model is used to estimate the probabilities. Other models could be used as well.

In this method toxicological responses are translated into four ordered categories of progressive effects: no effects, nonadverse effects, mild-to-moderate adverse effects, and severe or lethal effects. This set of categories resembles the dose categories used in setting the RfC (Jarabek et al., 1989) or RfD (Barnes and Dourson, 1988), namely, the no observed effect level (NOEL), NOAEL, LOAEL, and frank effect level (FEL), respectively. However, since all of the data are used in the regression, there is no need to specify a LOAEL, i.e., the lowest dose showing "mild-to-moderate" adverse effects. Thus, the more general term adverse effect level (AEL) is introduced.

To illustrate categorical regression, human toxicity data for arsenic (Tseng et al., 1968; Tseng, 1977) were summarized into categories as above (Table 2). These categories were regressed on the logarithms of exposure and duration; both parameters were statistically significant. The mathematical model used in this example is nonthreshold, but the extrapolation has been truncated at the oral RfD for arsenic of 3×10^{-4} mg/kg/d, since this value is considered by many to be a subthreshold dose for keratosis, the critical effect as defined by EPA (1990b).

Some difficulty exists with the interpretation of the Tseng (1977) and Tseng et al. (1968) toxicity data. For example, Tseng et al. (1968) did not show all necessary data for categorical regression; thus, some data have been estimated. Likewise, these authors maintained that blackfoot disease was attributed to arsenic exposure, but a recent publication disputes this (Lu, 1990). However, this data set is only used to illustrate the concept of categorical regression.

Table 1
Comparison of Estimated Benchmark Doses with Current IRIS No Observed Adverse Effect Level (NOAEL)

Chemical (CAS No.)	IRIS NOAEL and LOAEL[a] (mg/kg/d) and effect(s)	Study, species, sex and modeled effect[b]	Benchmark dose[c] (mg/kg/d)	Ratio of benchmark dose/NOAEL	Ratio of benchmark dose/LOAEL
Alachlor (15972-60-8)	1.0 and 3.0; hemosiderosis, hemolytic anemia (Monsanto Company, 1984)	Monsanto Co., 1984; D; M; kidney-hemosiderosis	0.86	0.9	0.3
		Monsanto Co., 1984; D; M; liver-hemosiderosis	1.9	2	0.6
		Monsanto Co., 1984; D; M; spleen-hemosiderosis	1.1	1	0.3
Dichlorvos (62-73-7)	0.08 and 0.8; increased liver weight and enlarged liver cells (Shell Chemical Co., 1967)	NCI, 1977; R; M; bile duct hyperplasia	2.2	30	3
		NCI, 1977; R; F; bile duct hyperplasia	2.3	30	3
		NCI, 1977; R; M; prostrate supporative inflammation	5.8	70	7
		NCI, 1977; R; L; liver-fatty metamorphosis	5.4	70	7
		NCI, 1977; R; F; bile duct inflammation	8.9	100	10
		NCI, 1977; R; F; pancreas-fibrosis	11	100	10
		NCI, 1977; R; M; pancreas-acinal atrophy	8.4	100	10
		NCI, 1977; R; F; pancreas-acinal atrophy	4.5	60	6
		NCI, 1977; M; M; endometrial cystic hyperplasia	11	100	10
		NCI, 1977; M; M; kidney-chronic inflammation	12	200	20
		NCI, 1977; M; F; kidney-chronic inflammation	20	300	30
		NCI, 1977; M; F; liver-chronic focal inflammation	49	600	60
		NCI, 1977; M; F; endometrial cystic hyperplasia	7.1	90	90
Hexachlorobenzene (118-74-1)	0.08 and 0.29; liver effects (Arnold et al., 1985)	Arnold et al., 1985; R; M; liver-centrilobular basophilic chromogenesis	0.11	1	0.4
		Arnold et al., 1985; R; F; liver-centrilobular basophilic chromogenesis	0.08	1	0.3
		Arnold et al., 1985; R; M; kidney-chronic nephrosis	0.16	2	0.6
Paraquat (1910-42-5)	0.45 and 0.93; chronic pneumonitis (Chevron Chemical Company, 1983)	Chevron Chem. Co., 1983; D; M; lung-chronic pneumonitis	0.16	0.4	0.2
		Chevron Chem. Co., D; F; lung-chronic pneumonitis	0.36	0.8	0.4
		Chevron Chem. Co., D; M; lung-hyperpnea	0.37	0.8	0.4
		Chevron Chem. Co., 1983; D; F; lung-hyperpnea	0.14	0.3	0.2
		Chevron Chem. Co., 1983; D; F; lung-increased vesicular sound	0.11	0.2	0.1
Pentachlorophenol (87-86-5)	3 and 10; liver and kidney pathology (Schwetz et al., 1978)	Schwetz et al., 1978; R; M; liver-brown pigment, etc.	28	9	3
		Schwetz et al., 1978; R; F; liver-brown pigment, etc.	3.0	1	0.3
		Schwetz et al., 1978; R; F; kidney-brown pigment, etc.	3.7	1	0.4
		NTP, 1989; M; M; liver-clear cell focus	11	4	1
		NTP, 1989; M; F; liver-clear cell focus	14	5	
		NTP, 1989; M; F; liver-basophilic focus	23	8	2

Table 1 (continued)
Comparison of Estimated Benchmark Doses with Current IRIS No Observed Adverse Effect Level (NOAEL)

Chemical (CAS No.)	IRIS NOAEL and LOAEL[a] (mg/kg/d) and effect(s)	Study, species, sex and modeled effect[b]	Benchmark dose[c] (mg/kg/d)	Ratio of benchmark dose/NOAEL	Ratio of benchmark dose/LOAEL
		NTP, 1989; M; M; liver-basophilic focus	24	8	2
		NTP, 1989; M; M; liver-hematopoietic cell proliferation	2.5	0.8	0.3
		NTP, 1989; M; M; liver-chronic active diffuse inflammation	0.55	0.2	0.06
		NTP, 1989; M; F; liver-chronic active diffuse inflammation	0.99	0.3	0.1
		NTP, 1989; M; M; liver-multifocal pigmentation	0.46	0.2	0.05
		NTP, 1989; M; F; liver-multifocal pigmentation	0.78	0.3	0.08
		NTP, 1989; M; F; liver-hepatocytomegaly, diffuse	0.22	0.07	0.02
		NTP, 1989; M; M; liver-bile duct, multifocal hyperplasia	1.6	0.5	0.2
		NTP, 1989; M; F; liver-bile duct, multifocal hyperplasia	26	9	3
		NTP, 1989; M; M; liver-hepatocyte necrosis	0.63	0.2	0.6
		NTP, 1989; M; F; liver-hepatocyte necrosis	0.60	0.2	0.06
		NTP, 1989; M; M; adrenal gland-medulla, hyperplasia	6.6	2	0.7
		NTP, 1989; M; M; seminal vesicle dilation	1.0	0.3	0.1
		NTP, 1989; M; F; mammary gland-cystic hyperplasia	3.5	1	0.4

[a] As on the Integrated Risk Information System (IRIS) as of 07/20/91 for the verified RfD. Note values since then may have changed.
[b] Effects modeled are from the critical study as on IRIS, and/or other studies on IRIS. When information on both sexes (M = male, F = female) and different species (R = rat, M = mouse, D = dog) are available, separate benchmark doses have been estimated.
[c] The dose associated with the 95% upper confidence limit on the response at 10%.

Table 2
As an Example of Categorical Regression, This Table Shows Arsenic Exposure and Resulting Toxicity from the Tseng et al. (1968) and Tseng (1977) Studies. Some Data Have Been Estimated. See Text for Discussion

Age (years)	Number of people	Dose[a] (μg/kg/d)	Effect and its severity and prevalence[b] per 1000: Hyperpigmentation (NOAEL)	Keratosis (AEL)	Blackfoot disease (FEL)
20–39	1561	0.8	0	0	0
		13	90	40	4.5
	9499	37	250	120	13
		49	270	120	14
40–59	955	0.8	0	0	0
		13	140	80	11
	6104	37	420	230	32
		49	600	340	47
60+	278	0.8	0	0	0
		13	250	140	20
	2021	37	390	220	32
		49	600[c]	340[c]	61

[a] Estimations of dose are based on Abernathy et al. (1989). Assumptions include the consumption of 4.5 l of water per day, arsenic from sweet potatoes and rice (which do not significantly contribute to the estimated dose), and a body weight of 55 kg. Arithmetic averages of well arsenic concentrations are used as a starting point of these estimated doses.

[b] All prevalence values have been rounded to two digits. Neither Tseng et al. (1968) nor Tseng (1977) give dose-specific prevalence rates for hyperpigmentation or keratosis. Thus, the prevalence rates for hyperpigmentation and keratosis with dose have been estimated based first on the averages across dose for these endpoints at each age given by Tseng et al. (1968) (Figures 5 and 6). These averages were modified first by assuming that the rate of blackfoot disease by dose and age (given in Tseng, 1977, page 113) could be used to build dose-specific rates for hyperpigmentation and keratosis. The resulting values were modified again to avoid the double-counting inherent in the Tseng et al. (1968) study as shown in their Table 2. We are attempting to get the raw data from Tseng et al. (1968).

[c] In this categorical regression method the prevalence per category is exclusive. Thus, the prevalence rates per 1000 of hyperpigmentation, plus keratosis, plus blackfoot disease cannot exceed 1000 (rounded). These values were reduced from 750 and 430, respectively.

Definitive conclusions on the toxicity of arsenic at levels above the RfD should not be drawn from this example.

Figure 1 shows the observed proportion of AEL or FEL (e.g., keratosis and/or blackfoot disease) in each of the dose and age groups listed in Table 2. This proportion is given by both age in years and dose in micrograms of arsenic per kilogram of body weight per day. This proportion has not been adjusted by mathematical manipulation. In this example, an increasing and sometimes erratic trend in a higher proportion of adverse and frank effects is seen with dose and duration.

Figure 2 shows the estimated probability of keratosis after 70 years of exposure using data from Table 2. The model shows an increasing probability of adverse effects with dose at doses at excess of the verified RfD. In this

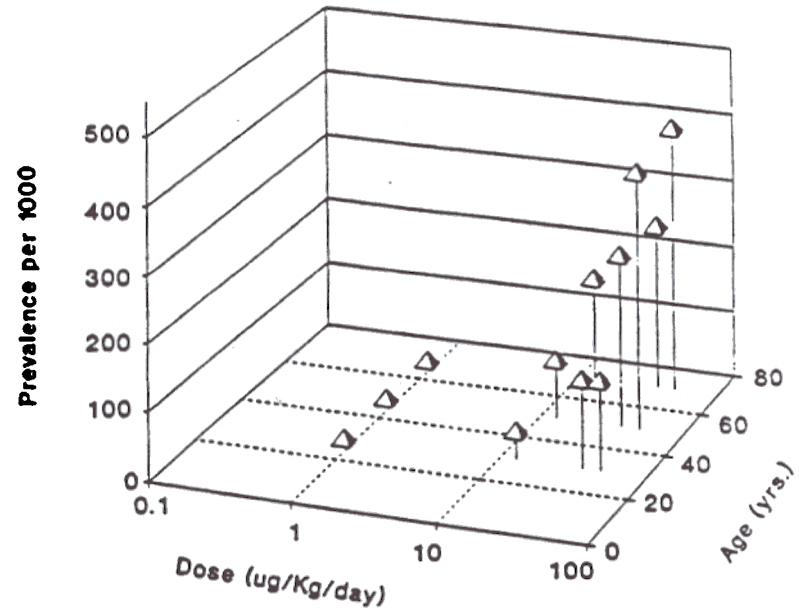

FIGURE 1. The prevalence per 1000 people of keratosis and/or blackfoot disease with both dose (μg/kg/d) and age (years). Data are from Tseng et al. (1968) and Tseng (1977). This analysis is only used as an example of categorical regression. Some values have been estimated; see text for discussion and caveats.

example, at three- and tenfold of the RfD (that is at 0.9 and 3 μg/kg/d, respectively), the risk of adverse effects (i.e., keratosis) is estimated to be 7 and 28 per 1000, respectively. Both of these estimations lie between the NOAEL of 0.8 μg/kg/d and LOAEL of 13 μg/kg/d shown in Table 2.

DISCUSSION

EPA, FDA, WHO, and others have developed useful approaches for defining exposures that are likely to be without appreciable risks of deleterious effects during a lifetime (e.g., for EPA the oral RfD, inhalation RfC, and developmental RfD). These approaches involve a consensus process. Risk assessment information generated in EPA through this process is made available through an electronic data base (i.e., IRIS). However, unlike the traditional approach for carcinogens, estimates of incremental risk for the noncancer endpoints of concern (or disease states) are not generally possible with these approaches.

New methods for the analysis of noncancer data are being developed and tested on existing data. As shown in this paper, these methods have advantages

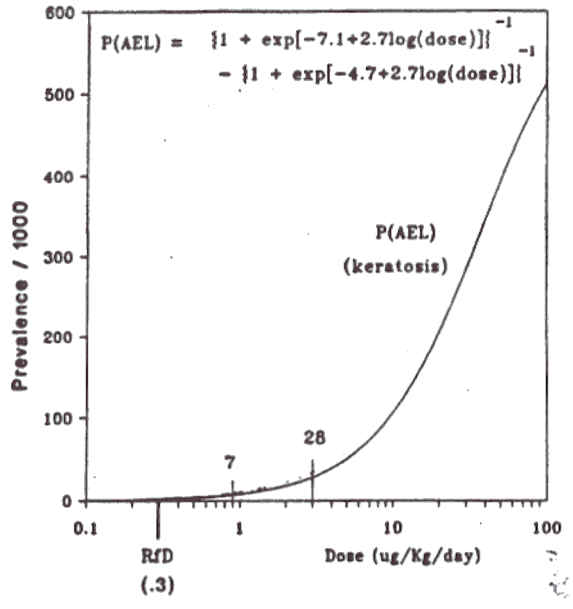

FIGURE 2. The prevalence per 1000 people of keratosis with both dose (μg/kg/d) at 70 years of age. Dashed line indicates upper 95% confidence limit. These values are estimated from data in Table 1 which are themselves estimated in part. This analysis is used only as an example of categorical regression. See text for caveats.

over existing approaches, but generally have requirements for larger and better data bases. Such requirements have not yet been worked out in detail, but in general would require incidence or continuous response data over several doses for the benchmark dose, and a sufficient number of dose groups in each of several categories of effect for the categorical regression approach. These models use more of the toxicity data, and thus, more information can be gleaned from their use — for example, risk above the oral RfD is possible. These methods might also be useful for evaluating carcinogenicity data where the mechanism of action appears to involve a threshold, e.g., homeostatic control.

In the example of benchmark dose in his paper, Table 1 is a comparison between the current NOAEL and LOAEL (as found on EPA's IRIS) and values obtained by the benchmark method. It is difficult to draw concrete conclusions from these comparisons. The current NOAEL and LOAEL are subject to a number of shortcomings as described earlier. As a result, the degree to which a benchmark dose does or does not agree is not a particularly good indicator of the performance of the new method. It could, in fact, be argued that the new method need not be totally consistent with the old one. If it were, then there would be no need for the new method. In addition, data

sets were modeled for benchmark more on the basis of availability rather than from a close study of relevance to the particular critical effect as described on IRIS. While one might argue that this is a useful application of the model in that it accounts for adverse effects other than the critical effect, this has the result of increased variability in any resulting comparisons.

It is, however, useful to make the comparison in terms of orders of magnitude. For the data presented here, the estimation of benchmark dose reasonably predicts (i.e., within an order of magnitude) the NOAEL dose as found on IRIS in a majority of cases. From this limited analysis it also appears that ratios based on the LOAEL rather than the NOAEL lead to somewhat better predictions when the benchmark is restricted to a 10% response.

One limiting factor in this analysis is the small number of chemicals for which the toxicity data are dichotomous. IRIS was searched randomly for 40 chemicals and very few were found to have sufficient data on which to estimate benchmark doses. The fact that many toxicity data involve a continuous response variable (e.g., decrease in body weight) presents an obstacle that must be overcome before general application of the benchmark dose approach is feasible. Crump (1984) indicated that there has been little experience with using continuous response data with dose-response models, but that it would be possible to transform continuous data to dichotomous form by considering any animal with a value of the continuous response variable more extreme than some arbitrarily defined cut-off value to be a responder. His proposed alternative is based on a continuous linear regression model.

In the example of categorical regression in this paper, the RfD is assumed to be the threshold. However, the conclusion that all doses above the RfD are associated with risk or that all doses at or below these values are without risk cannot be unequivocally stated. As discussed in Barnes and Dourson (1988), as the frequency of exposures exceeding the RfD increases, the probability that adverse effects may be observed in the human population increases. Such an increase is not a certainty. It follows then, that the categorical model may overstate risk at doses near the RfD. For example, it will give estimates of risk at a dose just above the RfD, even though health scientists believe that an RfD is a subthreshold dose.

This model is quite useful when the risks above the RfD for multiple chemicals are compared for different regulatory actions because the slope of the dose-response curve is shown. For example, it can readily distinguish the chemical representing the more serious potential health risk among several, when the data for these chemicals are normalized on a scale, such as 10-fold, 100-fold, of the oral RfD. This model's usefulness for estimating a single chemical risk such as arsenic at exposures near the RfD is more limited, however, due in part to its assumption that the RfD is the threshold. Moreover, the data on arsenic that have been used to generate this analysis are themselves problematic. (Definitive conclusions regarding the likely health risk above the arsenic RfD should not be drawn from this work. Rather, this example is used to illustrate the concept of categorical regression.)

The most important principle in quantitative risk assessment consists of the knowledge and judgments of scientific assessors in the interpretation and integration of available information. Not all models described here apply at all times. In addition, statistical issues must be considered such as model significance, model fit as compared with the experimental data, and the testing of underlying mathematical assumptions for the model. For experts, these models should serve as an adjunct to organizing factors and scientific interpretations with which they are familiar. For others, these models should aid in the preparation or review of a dose-response assessment and of the issues to be discussed with experts. The available information when combined and used with one or more quantitative procedures generally represents a substantial, but not perfect data set. Thus, scientific judgment by experts is critical to a defensible quantitative risk evaluation.

EPA is in the process of developing guidance in the quantitative assessment of noncancer health effects. Several features of these guidelines, such as the oral RfD, inhalation RfC, or developmental RfD, have been published elsewhere. However, these guidelines also encourage the use and discussion of several of the newer approaches to this area of science.

ACKNOWLEDGMENTS

The authors acknowledge the comments of Linda Knauf in the preparation of this text and her guidance on several statistical issues. The authors are also thankful to Bette Zwayer for the typing of this manuscript.

BIBLIOGRAPHY

Abernathy, C. O., W. Marcus, C. Chen, H. Gibb, and P. White. 1989. Office of Drinking Water, Office of Research and Development. EPA memorandum to P. Cook, Office of Drinking Water and P. Preuss, Office of Regulatory Support and Scientific Management. Report on Arsenic (As) Work Group Meetings. February 23.

Anatra, M. 1985. A method for the estimation of incidence from continuous response data: proposed modification in health score evaluation. Prepared for American Management Systems, Inc. Federal Consulting Group, Arlington, VA.

Arnold, D. L., C. A. Moodie, S. M. Charbonneau et al. 1985. Long-term toxicity of hexachlorobenzene in the rat and the effect of dietary vitamin A. *Food Chem. Toxicol.* 23(9):779–793.

Barnes, D. G. and M. L. Dourson. 1988. Reference dose (RfD): description and use in health risk assessments. *Reg. Toxicol. Pharmacol.* 8:471–486.

Brown, K. G. and L. S. Erdreich. 1989. Statistical uncertainty in the no-observed-adverse-effect level. *Fund. Appl. Toxicol.* 13(2):235–244.

Chevron Chemical Co. 1983. MRID No. 00132474. Available from Environmental Protection Agency. Write to Freedom of Information, EPA, Washington, D.C. 20460.

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

Dourson, M. L. 1986. New approaches in the derivation of acceptable daily intake (ADI). *Comments Toxicol.* 1(1):35–48.

Dourson, M. L., R. C. Hertzberg, R. Hartung, and K. Blackburn. 1985. Novel approaches for the estimation of acceptable daily intake. *Toxicol. Ind. Health.* 1(4):23–41.

Faustman, E. M., D. G. Wellington, W. P. Smith, and C. A. Kimmel. 1989. Characterization of a developmental toxicity dose-response model. *Environ. Health Perspect.* 79:229–241.

Gaylor, D. W. 1983. The use of safety factors for controlling risk. *J. Toxicol. Environ. Heath.* 11:329–336.

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

Guth, D. J., A. M. Jarabek, L. Symer, and R. C. Hertzberg. 1991. Evaluation of risk assessment methods for short-term inhalation exposure. Presented and published at the Annual Meeting of the Air & Waste Management Association, #91.173.2. June 1991, Vancouver, B.C.

Hertzberg, R. C. 1989. Fitting a model to categorical response data with application to species extrapolation of toxicity. *Health Phys.* 57:405–409.

Hertzberg, R. C. 1991. Quantitative extrapolation of toxicologic findings. in *Statistical in Toxicology*, D. Krewski and C. Franklin, Eds. Gordon and Breach, New York. 629–652.

Hertzberg, R. C. and M. E. Miller. 1985. A statistical model for species extrapolation using categorical response data. *Toxicol. Ind. Health.* 1(4):43–63.

Hertzberg, R. C. and L. Wymer. 1991. Modeling the severity of toxic effects. Presented and published at the Annual Meeting of the Air & Waste Management Association, #91.173.4. June 1991. Vancouver, B.C.

Hill, R. N., L. S. Erdreich, O. E. Paynter et al. 1989. Thyroid follicular cell carcinogenesis. *Fund. Appl. Toxicol.* 12:629–697.

Jarabek, A. M., M. G. Menach, J. H. Overton, M. L. Dourson, and F. J. Miller. 1989. Inhalation reference dose (RfD/i?): an application of interspecies dosimetry modeling for risk assessment of insoluble particles. *Health Phys.* 57:177–183.

Kimmel, C. A. 1990. Quantitative approaches to human risk assessment for noncancer health effects. *Neurotoxicology.* 11:189–198.

Kimmel, C. A. and D. W. Gaylor. 1988. Issues in qualitative and quantitative risk analysis for developmental toxicity. *Risk Anal.* 8(1):15–20.

Knauf, L. and R. C. Hertzberg. 1990. Statistical methods for estimating risk for exposure above the reference dose. EPA/600/8-90/065. NTIS/PB-90261504XSP. 104 p.

McCullagh, P. and J. Nelder. 1983. *Generalized Linear Models.* Chapman and Hill, New York.

Monsanto Co. 1984. MRID No. 00148923. Available from Environmental Protection Agency. Write to Freedom of Information, EPA, Washington, D.C. 20460.

NAS (National Academy of Sciences). 1983. *Risk Assessment in the Federal Government: Managing the Process*. National Academy Press, Washington, D.C.

NCI (National Cancer Institute). 1977. Bioassay of dichlorvos for possible carcinogenicity. NCI Tech. Rep. Ser. 10.

NTP (National Toxicology Program). 1989. Technical report on the toxicology and carcinogenesis studies of pentachlorophenol in B6C3F1 mice (feed studies). NTP Tech. Rep. No. 349. NIH Publ. No. 89-2804.

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

Schwetz, B. A., J. F. Quast, P. A. Keeler et al. 1978. Results of 2-year toxicity and reproductive studies on pentachlorophenol in rats. in *Pentachlorophenol: Chemistry, Pharmacology, and Environmental Toxicology*. K. R. Rao, Ed. Plenum Press, New York. p. 301.

SRC (Syracuse Research Corporation). 1989. RfDs from dose response models. Contract No. 68-C8-0004. Prepared for the Environmental Criteria and Assessment Office, Cincinnati, OH. Project Officer: Carol Haynes. September 13.

Tseng, W. P. 1977. Effects and dose-response relationships of skin cancer and blackfoot disease with arsenic. *Environ. Health Perspect.* 19:109–119.

Tseng, W. P., H. M. Chu, S. W. How, J. M. Fong, C. S. Lin, and S. Yeh. 1968. Prevalence of skin cancer in an endemic area of chronic arsenicism in Taiwan. *J. Natl. Cancer Inst.* 40:453–463.

U.S. EPA. 1986. Guidelines for the health assessment of suspect developmental toxicants. *Fed. Regist.* 51:34028–34040. September 24.

U.S. EPA. 1988a. Proposed guidelines for assessing male reproductive risk and request for comments. *Fed. Regist.* 53:24850–24869. June 30.

U.S. EPA. 1988b. Proposed guidelines for assessing female reproductive risk and request for comments; notice. *Fed. Regist.* 53:24834–24847. June 30.

U.S. EPA. 1990a. Interim methods for development of inhalation reference doses. Office of Health and Environmental Assessment, Washington, D.C. EPA/600/8-90/066A. August.

U.S. EPA 1990b. The Integrated Risk Information System (IRIS). Online. Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, Cincinnati, OH.

U.S. EPA. 1991. Guidelines for the developmental toxicity risk assessment; final. *Fed. Regist.* 56:63798–63826. December 5.