

Hertzberg, R.C. and M.L. Dourson. 1993. Using categorical regression instead of a NOAEL to characterize a toxicologist's judgment in noncancer risk assessment. In: Proceedings, Second International Symposium on Uncertainty Modeling and Analysis, College Park, MD, B.M. Ayyub, ed. IEEE Computer Society Press, Los Alamitos, CA.: 254-261.

## **Using Categorical Regression Instead of a NOAEL to Characterize a Toxicologist's Judgment in Noncancer Risk Assessment**

**Richard C. Hertzberg, Ph.D. and Michael L. Dourson, Ph.D.**  
Environmental Criteria and Assessment Office  
U.S. EPA  
Cincinnati, OH 45268

### **Abstract**

*Noncancer health risk assessment involves the evaluation of multiple types of toxic effects. For regulatory recommendations, such as the Reference Dose (RfD), the U.S. EPA relies heavily on expert judgment. This toxicologic judgment mixes toxic impact with likelihood: what effects are adverse, which of these is "critical," and which dose is the highest reliable NOAEL (No-Observed-Adverse-Effect Level). Uncertainty is indicated by qualitative statements of confidence. Statistical regression using ordered categories of overall toxicity is proposed as a superior alternative: uncertainty and variability are represented by statistical models, all relevant data are used, not just the NOAEL for the critical effect, and health risk can be estimated at exposure levels above the RfD.*

### **1: Background**

Most health risk assessments by regulatory agencies have one of two goals: to estimate a regulatory exposure level where health risk is minimal or zero, or to estimate the health risk at existing or proposed exposure levels. For the first goal, the oral Reference Dose (RfD) has been the mainstay of noncancer risk assessment in the U.S. Environmental Protection Agency (EPA) for several years. The RfD is defined as [1]:

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 appreciable risk of deleterious effect during a lifetime.

The RfD has an inhalation counterpart, the Reference Concentration or RfC. The RfDs and RfCs are published in an electronic data base called IRIS [2]. For clarity without loss of generality, this paper will only discuss the RfD. The Reference Dose has official status in EPA as a scientifically derived exposure level that feeds into the standard-setting process that regulates oral exposures. The RfD has been applied most often to risk assessments for chronic or lifetime exposures.

For the second goal, no procedures for noncancer risk have been universally adopted by the U.S. EPA. Statistically based approaches have been generally limited to characterizations of the experimental data set; extrapolation to probabilistic human risk estimates, as is performed for cancer risk, has not been advocated because of high uncertainty and lack of empirical verification. When human exposure-response data are available, three procedures have been used or recommended for obtaining a human risk estimate: expert panel judgment of risk, Bayesian modeling of the critical effect, and regression of adversity categories for all effects combined. Recently, the EPA's Risk Assessment Council issued a memorandum requiring adequate discussion of uncertainties in all risk characterizations. No standard guidance yet exists for conducting and reporting quantitative uncertainties in noncancer risk assessment.

This paper presents some ideas on uncertainty that apply to noncancer risk assessments based on categorical regression of toxic adversity on exposure. The basic regression approach is contrasted with the Reference Dose approaches that use a NOAEL or use a benchmark dose. Suggestions for graphical presentation of results are also given.

### **1.1: Reference Dose**

The determination by EPA of the RfD for a given chemical involves several judgmental steps: (1) evaluate all pertinent research articles on its toxic effects to determine the critical effect, i.e., the toxic effect or its known precursor occurring at the lowest exposure level. (2) choose the most scientifically sound study with the most appropriate NOAEL (No-Observed-Adverse-Effect Level) of the critical effect. (3) decide the magnitude of one or more uncertainty factors, then divide the selected NOAEL by those factors to arrive at the RfD [1]. The five uncertainty factors in current use primarily address data gaps. Three of these *scale* the dose, defined here as the conversion of one dose to a toxicologically equivalent dose by multiplying the first dose by a constant. The scaling factors then convert the entire probabilistic dose-

response curve, usually by a shift to the left on the dose scale (dose reduction). These factors account for three extrapolations: from animal to human data, from subchronic to chronic exposures, and from an incomplete bioassay (e.g., no reproductive studies) to one addressing all major effects. The other two factors move the dose down the dose-response curve to reach a lower risk level. These factors shift the dose from the LOAEL (Lowest-Observed-Adverse-Effect Level) to the NOAEL, and move into the tail of the average response distribution to account for sensitive population subgroups. Additional factors are sometimes used to account for other concerns or weaknesses (see [3] for a more complete discussion). The quality of the RfD is indicated by ratings of confidence in three areas: general data base completeness and reliability, quality of the main study on the critical effect (the one containing the NOAEL), and overall quality of the resulting RfD itself. Originally these statements covered areas not well addressed by the uncertainty factors. Currently, the uncertainty factors are broader in scope, and the confidence statements are usually closely tied to the values selected for the uncertainty factors, i.e., they are somewhat redundant.

The RfD procedure has several good characteristics. As practiced by EPA, the RfD is a unanimous group judgment of all available toxicologic data on a given chemical. This judgment allows the consideration not only of multiple toxic effects in the same animal when deciding adversity, but also the relative tradeoffs between study quality and the biological significance of the observed effects. The definition of the RfD implies that an exposure at that level corresponds to a low, or zero, risk of adverse effects. It is not assumed to be a threshold level, but the better quality RfDs are believed to be near the upper end of the subthreshold exposure levels. As is shown later, this interpretation is easily converted into a probabilistic, mathematical definition.

## 1.2: Benchmark Dose

Recently, interest has been expressed in replacing the NOAEL-based RfD with one based on dose-response modeling. This concept is not new to the Agency, being first mentioned in an EPA workshop in 1982 [4]. This method involves fitting a mathematical model to quantal dose-incidence data on a single effect, estimating a dose corresponding to a preset benchmark risk, and dividing that benchmark dose estimate (BMD) by one or more uncertainty factors to obtain an RfD [5-8]. The central dose estimate as well as a lower confidence bound have been suggested for the benchmark dose. The data to be modeled are from a single study and usually represent the critical toxic effect. The benchmark risk level is chosen to

be low, e.g., 1-10%, and is often interpreted as the lowest incidence that can be statistically significant in common toxicology studies. For studies with dose groups of roughly 20 or fewer animals, the 1% response level may be considered consistent with the NOAEL, and the 10% response level with the LOAEL [4]. The uncertainty factors that are divided into the benchmark dose may be different from those used in the NOAEL-based procedure but should address many of the same extrapolation issues.

The benchmark procedure has positive attributes, primarily in describing the selected experimental data. The confidence bound is especially useful in reflecting not only the low risk region, but also statistical variation in the data. If the procedure includes screening of data sets by minimal criteria for model fit, then the model quality is also reflected. If the same benchmark risk (e.g., 1%) is employed across chemicals, then this dose precursor to the RfD is much more consistent than is the NOAEL and has a more precise interpretation.

## 2: Uncertainty in existing procedures

Several studies have looked at noncancer risk assessment and the RfD from a quantitative perspective. Nearly all focus on only one or two steps in the RfD procedure and as such, they enhance our understanding of some key sources of uncertainty. For example, some studies have focused on one or more of the uncertainty factors, considering distributional or probabilistic replacements [8-13], and proposing separate factors for cross-species scaling of delivered dose and for species differences in metabolism and pharmacodynamics [14]. Because these studies leave out important steps in the RfD process, they do not address the overall uncertainty of the RfD. These difficulties are now discussed in terms of the potential for assessing uncertainty in the RfD and its interpretation. The NOAEL and BMD approaches as well as the categorical regression procedure discussed below all require use of uncertainty factors. The comparison of these three approaches can then be performed separately; further analysis of the uncertainty factors will not be included in this paper.

### 2.1: Uncertainty in the NOAEL

Variability in the NOAEL has usually been investigated by modeling the likelihood that one of the experimental doses would be selected as "the NOAEL" for RfD calculation. Brown and Erdreich [15] used a modified weighted average method that allowed the optimal pseudo-NOAEL to be intermediate between experimental levels. One of the more complete statistical analyses of the NOAEL selection is Leisenring and Ryan [16]. In one of

their many investigations of NOAEL selection and interpretation, they evaluated the excess risk at the NOAEL for an artificial data set of four dose groups. The risks were based on a Weibull dose-response model with four parameter sets, selected to illustrate extremes in model shape. Particularly satisfying is that their conclusions are generally consistent with previous "common sense" opinions. For example, the smaller the dose group, the higher will be the selected NOAEL and higher will be the undetected risk at the NOAEL. Specifically, they concluded that the selected NOAEL could, for plausible group sizes of 10-20 animals, represent an excess risk as high as 30%. They also show sensitivity to the background response rate and to the shape of the true dose-response curve. Perhaps more important is their demonstration of the considerable *range* of likely underlying risks at the NOAEL, a convincing illustration of the potential for extreme inconsistency in interpretation of the NOAEL and its resulting RfD.

## 2.2: Uncertainty in the benchmark dose

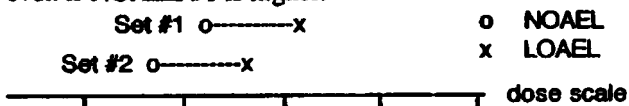
Quantitative studies on the benchmark procedure usually examine the variation in the benchmark dose for a given data set. Crump [5] (see also Kimmel [7] and references) shows the influence of dose group size and the choice of benchmark risk on the resulting benchmark dose, and demonstrates the general lack of sensitivity to the choice of mathematical model as long as model fit is adequate. Gaylor [17] focuses instead on the interpretation of the procedure, at least as far as dividing the benchmark dose by a factor. Gaylor's procedure differs slightly from the BMD method described above. Let the  $ED_{01}=5.0$  (in some arbitrary dose units) be the BMD, with an upper bound estimated risk of 4%. Let the desired "allowable" risk be  $10^{-4}$ . Gaylor proposes dividing the BMD by the ratio of the upper risk estimate to the desired risk, e.g.,  $0.04/0.0001=400$ . As long as the true dose-response curve is linear or convex (concave upwards), then the resulting dose,  $5/400=0.0125$ , is associated with a risk no greater than  $10^{-4}$ . Because this procedure gives an upper bound on true risk for the selected data set, and does not extrapolate across species or duration, investigations of uncertainty are easier. Such analyses have yet to be performed.

## 2.3: Inference about uncertainty in the RfD

None of the studies seems to have considered the overall uncertainty in the entire process, i.e., the uncertainty in the RfD and its interpretation. The variation and uncertainty in the numerical value of the NOAEL may not be the most significant or relevant aspect of the RfD procedure [18]. The decision of what effect is adverse is a process based on

skilled judgment. Any improvement or replacement of the NOAEL-based method will include at least that judgmental step, if not others. The uncertainty analysis of the improved procedure must also reflect that judgment.

The NOAEL evaluation by Leisenring and Ryan [16] does cast the uncertainty analysis in terms of judgment, i.e., the identification of one of the experimental doses as the preferred NOAEL. Their results, however, do not provide much insight into the uncertainty of the RfD. First, their analysis of the NOAEL seeks the highest response rate that is not statistically significant. In contrast, the RfD process seeks the highest NOAEL such that the risk of adverse effects is minimal (see the RfD definition given earlier). The EPA selection of NOAEL is then strongly influenced by the LOAELs, and less so by the properties of the NOAELs. For example, if two acceptable data sets each have a NOAEL and LOAEL, and set #2 has a lower LOAEL, as shown in the following schematic, then its NOAEL (#2) will often be selected even if NOAEL #1 is higher.



Second, their research does not consider the judgment of whether an effect is "adverse," i.e., the likelihood that the observed effects will be counted at all, so the NOAEL uncertainty's contribution to the RfD's uncertainty cannot be assessed. Further, the likely conservatism in the joint use of several uncertainty factors is not addressed, so any evaluation of the resulting human risk is incomplete. Consequently, their conclusions regarding likely human risk, e.g., that the NOAEL could "identify a dose level associated with unacceptably high risk" (see [16] p. 170), are not valid.

Similarly, the benchmark dose (BMD) method described above is also incomplete since it, too, fails to consider the judgment of adversity or the selection of uncertainty factors. This procedure has other flaws that render it difficult to use in uncertainty analysis. It is inconsistent with the current NOAEL-based procedure and its resulting RfD is difficult to interpret. The inconsistency arises because all effects are not included in the model. The RfD definition gives the goal of a dose level where the likelihood of *any* toxic effect is small. As mentioned above, the NOAEL selection process also seeks to ensure that the more serious effects have even smaller risks. In fact, if the only adverse effects in a study are severe (the dose level has been termed a Frank Effect Level), then that study is not recommended for NOAEL selection. In contrast, the benchmark method only guarantees small risk for the selected effect. Note that there are certain circumstances where only one effect

group need be considered, e.g., when public concern is high and where animal-to-human concordance is known or believed. Such a case is developmental toxicity, where any such effect is serious; EPA has proposed guidelines for using the BMD method in developmental risk assessment [19].

For general noncancer toxicity, focussing on one effect may ignore important information [4]. To demonstrate the unsatisfactory nature of this result, consider the artificial data set given in Table 1 with model results in Table 2. If the 10% benchmark risk were used, then effect 3 would be chosen since it has the lowest ED<sub>10</sub>. Using EPA's cancer risk procedures as a guide, serious toxic effects (frank effects and mortality) should be restricted to extremely low risk, perhaps in the range 10<sup>-3</sup> down to 10<sup>-6</sup>. If effect 1 were serious, however, the BMD would correspond to a 2% risk of that serious effect. In order to ensure that the risk of serious effects is small at the BMD, all serious effects would also have to be modeled, requiring some accounting for their statistical and physiological correlations, as well as for the combined impact of multiple effects on the same individual. This combination of quantitative and judgmental information is difficult to work into the benchmark procedure.

**Table 1. Artificial Data for Three Effects**

Dose	Effect1	Effect2	Effect3
0	0	0	2
1	0	1	1
2	1	2	4
3	4	2	8
4	5	4	10

**Table 2. Estimated Risks and Benchmark Levels for Three Effects\***

Dose	Estimated Risk		
	Effect 1	Effect 2	Effect 3
2.58	.10	.11	.28
2.31	.07	.10	.24
1.35	.02	.06	.10

\* The dose in each row is the BMD (10% response) for effects 1, 2, 3, respectively.

The upper bound approach of Gaylor [17] is by itself not useful for uncertainty analysis of the RfD. First, it only addresses one effect and thus has similar difficulties to the benchmark dose method. Second, it only addresses the experimental data set. Uncertainty factors to account for species differences and other extrapolations would also be needed, injecting other sources of uncertainty into the human risk assessment. Kimmel [7] expands Gaylor's

procedure to include the uncertainty factors: the lower bound on the ED<sub>10</sub> (denoted LED<sub>10</sub>) is used as the BMD and is divided by uncertainty factors in the same way Gaylor divides his dose by a risk reduction factor. But Kimmel then interprets this procedure as a way of bounding the human risk (p. 194): "...the true risk at the LED<sub>10</sub>/UF is equal to or less than 10%/UF, and may be below the threshold dose." The uncertainty factors account for differences in species pharmacokinetics and pharmacodynamics, along with other extrapolations that compensate for data weaknesses. In fact, if selected well, the "scale" uncertainty factors would convert dose perfectly. For example, if a NOAEL were used and the "sensitive subgroup" factor were 10, then:

$$(\text{animal LED}_{10})/UFs = (\text{human LED}_{10})/10$$

The estimated human risk may then have an upper bound as high as 1%.

### 3: Regression of Toxic Adversity

The two main obstacles to using the previous approaches for uncertainty analysis are their inability to use all data, and their exclusion of uncertainty in judgment of the extent of toxicity or adversity (sometimes also called toxic "severity"). One approach that includes the toxicologist's judgment into the dose-response model is generalized linear regression on toxicity categories [20-23]. This statistical procedure requires the toxicologist to assign the observed effects to one of a fixed number of categories according to overall toxic impact on the exposed individual. The basic approach is motivated by a probabilistic interpretation of the RfD definition: high confidence that for doses below the RfD, there is little risk of adverse effects. The translation in terms of confidence limits is: find the dose *d* such that

$$p(\text{risk} < 0.01 \mid \text{dose} < \text{RfD}) \geq 0.95$$

The values 0.01 and 0.95 are only used for illustration.

The regression model for adversity on dose is:

$$\log(\pi_i / (1 - \pi_i)) = \alpha_i + \beta_1 \cdot \log(\text{dose}) \quad (1)$$

where  $\pi_i$  is the cumulative frequency or probability of adversity category *i* or less,  $\alpha_i$  is a model location parameter specific to category *i*, and  $\beta_1$  is a common slope parameter applicable to all categories. If duration were also a covariate, there would be a  $\beta_2$  parameter.

If the first "adverse effect" category is 2, then Equation (1) can be expressed in terms of the chance of adverse or worse effects (toxicity score *S* is category 2 or higher):

$$p(S > 1) = \frac{1}{1 + \exp[\alpha_1 + \beta_1 \cdot \log(\text{dose})]} \quad (2)$$

Note that if the desired categories were 3 and higher, then  $\alpha_1$  would change to  $\alpha_2$ . This equation now gives the risk of adverse effects in terms of a toxicologist's judgment. If data are available on individual exposed animals, then *p*

represents actual risk. If data are only available at the dose group level, e.g., the dose is considered to cause adverse effects in the group and is reported as such, then  $p$  represents the probability that a toxicologist would assign the dose group to category 2 or higher.

### 3.1: Procedure description

Because most toxicity studies include only a few dose levels, the problems of data sparseness dictate that the number of adversity categories should be small. EPA has previously used four categories to characterize doses by their observed toxic effects [6]: no effects (NOEL), no adverse effects (NOAEL), adverse effects (AEL) and frank or lethal effects (FEL). We have adopted those general categories:

Score	Toxic Adversity	Corresponding Dose-Level Acronym
0	None	NOEL
1	Minor, non-adverse	NOAEL
2	Moderate	AEL
3	Severe, lethal	FEL

except in cases of minimal data, when the lower two (NOEL, NOAEL) were merged [24].

The data on noncancer toxicity is typically characterized by several types and extent of toxicity, and in different physiological organs or systems. An example of some of the data for the chemical, hexachlorobenzene, along with the adversity category codes, is given in Table 3.

Table 3. Oral Data for Hexachlorobenzene

Dose	Organ	Effects	Adversity
0.016	liver	no effects	1
0.08	liver	glycogen depletion in males	2
0.4	liver, heart	increase liver pathology, increase heart weight	3
2.0	liver, kidney, other sites	increase mortality in pups, incr. liver and kidney pathology, incr. adrenal pheochromocytomas in females	4

The advantage of categorical regression is that these numerical codes (1-4) are not directly used in the calculations; they only indicate the ordering of the categories, so codes 1,3,10,17 would work equally well. Uncertainty in the categories is given by the probability that the toxicologist would assign a different category, and is then directly interpreted in terms of judgment. For a given dose, the model says that the chance of being

assigned to one of the categories follows a 4-way multinomial distribution with a success parameter that depends on dose. For lower doses, the chances are best that the dose will be classified in category 1 (a NOEL); for the highest doses, the greatest chance is for category 4. These probability functions have the general shapes given in Figure 1. Note that the density functions need not be symmetric, e.g., curves 1 and 4 are not necessarily mirror images of each other.

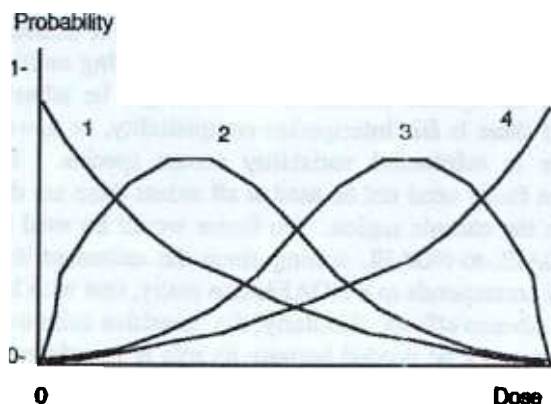


Figure 1. Probability density functions for the four severity categories.

The fitting procedure we have used is Proc Logistic, which is implemented on the Macintosh computer via the JMP package [25]. This procedure uses maximum likelihood and can produce goodness-of-fit statistics along with confidence limits [26]. For data at the dose group level, we have explored several ways to use the dose group size as a weighting factor in the regression. The easiest way is to ignore group size and treat the groups equally as data elements. One simple way to incorporate group size is to substitute group size for frequency. This method implies that all animals in the group respond identically. The resulting confidence bands narrow considerably compared with the unweighted method because of the much larger sample size. A better way is to view group size,  $n_i$ , as a quality indicator, and assign it to one of a small number of weighting factors, as follows:

Group size	Weighting factor
1-10	1 or .15
11-25	2 or .35
26->	3 or .50

This simple scheme prevents one study from dominating the results. Another scheme involves the total number of animals as well as the number of dose groups:

$$w_i = \frac{n_i}{\sum_{j=1}^N n_j} \cdot N \quad (3)$$

The regression results can be converted into an RfD in a similar fashion as was the BMD. A dose for chronic exposure duration is identified that corresponds to a low risk of adverse effects, and, if applicable, to an even lower risk of severe effects. The use of a lower confidence bound on the dose allows the interpretation of "high confidence" as is in the RfD definition. This dose is then divided by uncertainty factors to obtain the RfD. The magnitudes for the standard uncertainty factors have not yet been determined, but some should be smaller than those used either for the NOAEL or BMD approaches. For example, because multiple studies are combined, reflecting multiple species, the species sensitivity factor might be adjusted down if there is fair interspecies compatibility, or upward if there is substantial variability across species. The duration factor need not be used at all unless there are data gaps in the chronic region. No factor would be used for the LOAEL-to-NOAEL scaling since the estimated dose already corresponds to a NOAEL-like entity, that with low risk of adverse effects. Similarly, the "sensitive subgroup" factor may not be needed because its role is to reduce the risk to average humans to protect sensitive groups, a process already included if the target risk is set sufficiently low, e.g., 0.01.

### 3.2: Examples

A simple example uses the data for manganese (Figure 2). These data represent human toxicity from a variety of exposure levels and durations. The exposure conditions are the axes. The toxicity categories are represented by symbols: square=NOAEL, diamond=AEL, x=FEL. The incidence data include multiple toxicity categories for each exposure pair, so the durations have been jiggled ( $\pm 200$ ) to distinguish the overlapping points. The larger symbols denote incidences  $>0.4$ . Trends are difficult to detect in plots with multiple levels, but the general sense of increasing toxicity (lower dose) with increasing duration is indicated, both by AELs at lower doses and increasing incidence of FELs. The extent of overlap of different symbols indicates the variability in observed toxicity at a given exposure.

We have investigated several chemical data sets using this approach. The regression usually includes  $\log_{10}(\text{dose})$  and duration as covariates. One display is to present the data as in Figure 1, along with two lines representing the dose\*duration combinations that are associated with two benchmark risks: a 10% or 1% probability of observing adverse or severe toxic effects (see [22] and [24] for examples). This type of graph is useful not only to see where the "benchmark region" is on the dose\*duration plane, but also to identify data gaps where the model results are less secure. The use of symbols for adversity

categories allows the risk lines to be overlaid onto three-dimensional data, but given as a two-dimensional graph.

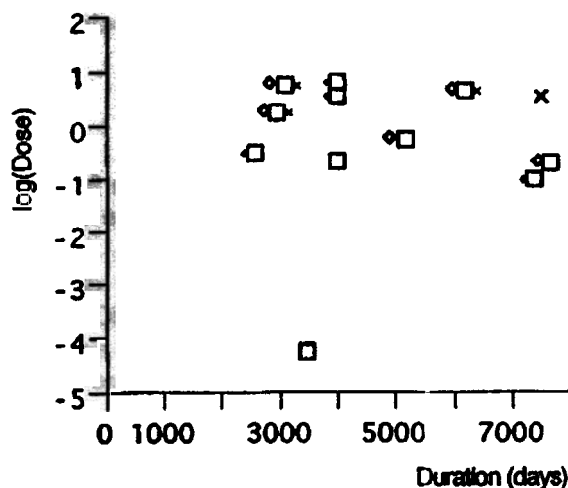


Figure 2. Human toxicity data for manganese.

The risk results can also be displayed as risk vs. dose, either for different durations (one curve for each duration), or for a fixed duration along with confidence bands. One of the more useful displays is to overlay the risk of the moderate effects with the risk of severe effects. This display allows the risk manager to balance the two risks in the regulatory strategy, an option not available with the NOAEL or benchmark approaches.

### 3.3: Inference about uncertainty in the RfD

The categorical regression procedure is highly consistent with the RfD definition and procedure. It can be applied to data at the dose group level (e.g., hexachlorobenzene) as well as data on individual animals (e.g., manganese). The exposure level in the experimental data set associated with an acceptably small risk can be bracketed by confidence bounds. The visual display can also alert the risk assessor to data gaps, especially if duration is a covariate, or to dose ranges with high overlap of adversity categories, indicating high uncertainty in the reported studies. These uncertainties directly transfer to the RfD. If the regression is applied at the dose group level, then the uncertainty in the estimated low-risk dose incorporates the uncertainties in the toxicologists' judgments of NOAELs, AELs and FELs. That type of uncertainty cannot be represented by the NOAEL or BMD methods. If the data are at the level of individual animals, then within-group variability can also be identified. The resulting RfD can then be interpreted in the context of actual experimental risk, as can the BMD.

The difference is that the risk here is of any toxic effect, not just the critical effect, and risks of both moderate and severe effects can be calculated.

#### 4: Discussion and Conclusions

The benchmark dose method and the adversity regression method must be treated carefully when assessing uncertainty. As is well described by Lu and Sielken [27], any modeling approach must be interpreted with caution. In their section, "Potential pitfalls in quantitative dose-response modeling," they note several concerns, many of which apply to the context of uncertainty analysis. Two of the more important concerns for risk estimation are:

1) goodness-of-fit is misleading. It does not confirm accuracy in extrapolation and may not have sufficient power to differentiate models or reject models. For decision making, several numerical criteria can be used in addition to statistical goodness-of-fit measures. Among these are the relative width of the confidence interval, statistical significance of model parameters, concordance measures (for ordinal regression), and performance near the low risk or threshold dose region.

2) confidence bands and model results can hide key assumptions and decisions. Particularly in the combining of data across species, care must be taken to ensure compatibility. For the Benchmark Dose method, one key decision is the choice of effect to model. As mentioned above, this issue is not critical for developmental effects. For general toxicity, the relation between the critical effect and other important effects must be evaluated. For categorical regression of dose group data, one decision is the weighting function. For both methods, the choice of confidence level is important.

The use of regression on ordered categories of toxic adversity is a useful approach for incorporating all relevant studies and toxicological judgment into the calculation and uncertainty analysis of the Reference Dose. The clarity of interpretation that is also consistent with the RfD definition allows for a shift in focus to the uncertainty factors. Once these factors can be modeled or described in a distributional context, they can be combined with the categorical regression to fully characterize the uncertainty in the RfD. Work has begun in that direction [28]. A more ambitious goal would be to replace each factor with a model-based scaling factor (or function) so the animal adversity and incidence can be directly translated into terms of human risk. Until that happens, the Reference Dose concept will continue as the best approach to regulating chemicals based on the exposure-toxicity relationship.

#### 5: References

- [1] Barnes, D.G. and M.L. Dourson: "Reference Dose (RfD): description and use in health risk assessment" *Reg. Toxicol. Pharmacol.* 8:471-486 (1988)
- [2] U.S. EPA: "Integrated Risk Information System (IRIS)" Online. Office of Health and Environmental Assessment, Environmental Criteria and Assessment Office, Cincinnati, OH. (1993)
- [3] Lu, F.C. and M.L. Dourson: "Safety/risk assessment of chemicals: principles, procedures, and examples" *J. Occup. Med. Toxicol.* 1:321-335 (1992)
- [4] Dourson, M.L.: "Toxicity-based methodology, thresholds and possible approaches, and uncertainty factors" with critique and discussion, in *Approaches to risk assessment for multiple chemical exposures.* EPA-600/9-84-008, pp.1-42 (1984)
- [5] Crump, K.S.: "A new method for determining allowable daily intakes" *Fund. Appl. Toxicol.* 6:41-93 (1984)
- [6] Dourson, M.L., R. C. Hertzberg, R. Hartung, and K. Blackburn: "Novel methods for the estimation of acceptable daily intake" with discussion, *J. Toxicol. Ind. Health* 1:23-41 (1985)
- [7] Kimmel, C.: "Quantitative approaches to human risk assessment for noncancer health effects" *Neurotoxicol.* 11:189-198 (1990)
- [8] Dourson, M.L. and J.F. Stara: "Regulatory history and experimental support of uncertainty (safety) factors" *Reg. Toxicol. Pharmacol.* 3:224-238 (1983)
- [9] Calabrese, E.J. "Uncertainty factors and interindividual variation" *Reg. Toxicol. Pharmacol.* 5:190-196 (1985)
- [10] Hattis, D., L. Erdreich and M. Ballew: "Human variability in susceptibility to toxic chemicals- a preliminary analysis of pharmacokinetic data from normal volunteers" *Risk Analysis* 7:415-426 (1987)
- [11] Travis, C.C. and R.K. White: "Interspecific scaling of toxicity data" *Risk Analysis* 8:119-126 (1988)
- [12] D. Hattis and S. Lewis: "Reducing uncertainty with adjustment factors" *The Toxicologist* 12:23 (1992)
- [13] Dourson, M.L., L. Knauf and J. Swartout: "On reference dose (RfD) and its underlying toxicity data base" *Toxicol. Ind. Health* 8:171-189 (1992)
- [14] Calabrese, E.J., B.D. Beck and W.R. Chappell: "Does the animal to human uncertainty factor incorporate interspecies differences in surface area?" *Reg. Toxicol. Pharmacol.* 15:172-179 (1992)
- [15] Brown, K.G. and L.S. Erdreich: "Statistical uncertainty in the no-observed-adverse-effect level" *Fund. Appl. Toxicol.* 13:235-244 (1989)

- [16] Leisenring, W. and L. Ryan: "Statistical properties of the NOAEL" *Reg. Toxicol. Pharmacol.* 15:161-171 (1992)
- [17] Gaylor, D.W.: "The use of safety factors for controlling risk" *J. Toxicol. Environ. Health* 11:329-336 (1983)
- [18] Dourson, M.L., R.C. Hertzberg and J.F. Stara: "Letter to the editor: comments on a new method for determining allowable daily intakes" *Fund. Appl. Toxicol.* 6: 182-83 (1986)
- [19] U.S. EPA: "Guidelines for developmental toxicity risk assessment; Notice" *Fed. Reg.* 56(234):63798-63826.
- [20] Hertzberg, R. and M. Miller: "A statistical model for species extrapolation using categorical response data" *Toxicol. Ind. Health* 1(4): 43-57 (1985)
- [21] Hertzberg, R. C.: "Fitting a model to categorical response data with application to species extrapolation" *Health Physics* 57, Supplement 1:405-409 (1989)
- [22] Hertzberg, R.C. and L. Wymer: "Modeling the severity of toxic effects" #91.173.4. Presented and published at the annual meeting of the Air & Waste Management Association, Vancouver, B.C. (1991)
- [23] Hertzberg, R. C.: "Quantitative extrapolation of toxicological findings" in *Statistics in Toxicology*. D. Krewski and C. Franklin, eds. Gordon and Breach, New York: 629-652 (1991)
- [24] Guth, D. J., A. M. Jarabek, L. Wymer and R. C. Hertzberg: "Evaluation of risk assessment methods for short-term inhalation exposure" #91.173.2. Presented and published at the annual meeting of the Air & Waste Management Association, Vancouver, B.C. (1991)
- [25] SAS: "JMP User's Guide" SAS Institute Inc., Cary, NC (1989)
- [26] McCullagh, P. and J.A. Nelder: "Generalized Linear Models" Chapman and Hall, Ltd., New York, NY (1983)
- [27] Lu, F.C. and R.L. Sielken Jr.: "Assessment of safety/risk of chemicals: inception and evolution of the ADI and dose-response modeling procedures" *Toxicol. Letters* 59:5-40 (1991)
- [28] Swartout, J.C. "Defining the 'reference' in the Reference Dose" In press (1993)