Human and Ecological Risk Assessment: Vol. 2, No. 1, pp. 79-102, 1996

Noncancer Risk Assessment: A Probabilistic Alternative to Current Practice

Sandra J. S. Baird,[1,] Joshua T. Cohen,[2] John D. Graham,[2] Alexander I. Shlyakhter,[1] and John S. Evans[1,2]**

[1] Program in Environmental Health and Public Policy, Harvard School of Public Health, Boston, MA*

[2] Harvard Center for Risk Analysis, Harvard School of Public Health, Boston, MA**

ABSTRACT

Based on imperfect data and theory, agencies such as the United States Environmental Protection Agency (USEPA) currently derive "reference doses" (RfDs) to guide risk managers charged with ensuring that human exposures to chemicals are below population thresholds. The RfD for a chemical is typically reported as a single number, even though it is widely acknowledged that there are significant uncertainties inherent in the derivation of this number.

In this article, the authors propose a probabilistic alternative to the EPA's method that expresses the human population threshold as a probability distribution of values (rather than a single RfD value), taking into account the major sources of scientific uncertainty in such estimates. The approach is illustrated using much of the same data that USEPA uses to justify their current RfD procedure.

Like the EPA's approach, our approach recognizes the four key extrapolations that are necessary to define the human population threshold based on animal data: animal to human, human heterogeneity, LOAEL to NOAEL, and subchronic to chronic. Rather than using available data to define point estimates of "uncertainty factors" for these extrapolations, the proposed approach uses available data to define a probability distribution of adjustment factors. These initial characterizations of uncertainty can then be refined when more robust or specific data become available for a particular chemical or class of chemicals.

Quantitative characterization of uncertainty in noncancer risk assessment will be useful to risk managers who face complex trade-offs between control costs and protection of public health. The new approach can help decision-makers understand

* Program in Environmental Health and Public Policy, Department of Environmental Health, Harvard School of Public Health, 665 Huntington Ave., Boston, MA 02115; Tel: (617) 432-1259; Fax: (617) 432-1226

** Harvard Center for Risk Analysis, 718 Huntington Ave., Boston, MA 02115

*** Author to whom correspondence should be addressed.

how much extra control cost must be expended to achieve a specified increase in confidence that the human population threshold is not being exceeded.

Key Words: reference dose, uncertainty, human population threshold, noncancer risk assessment

INTRODUCTION

The USEPA's current approach to noncancer risk assessment is coming under increasing scrutiny. The approach of applying default uncertainty factors to a no-observed-adverse-effect-level (NOAEL) does not satisfy the information needs of decision-makers (Beck *et al.*, 1993; Farland and Dourson, 1993; Habicht memo, 1992). In response, quantitative methods have been developed for improving or replacing the NOAEL, such as benchmark dose (Crump, 1984; Kimmel and Gaylor, 1988; Barnes *et al.*, 1995) and categorical regression (Hertzberg, 1989; Hertzberg and Dourson, 1993; Farland and Dourson, 1993). Further, risk managers have recognized the need for improved quantitative characterization of uncertainties (SAB letter to Reilly, 1990). Nonetheless, relatively little attention has been focused on the development of quantitative methods for characterizing the uncertainty in estimates of reference doses or acceptable daily intake levels.

Recent work on uncertainty factors has focused primarily on: (i) refining the point estimates of the adjustments needed to account for differences in sensitivity between the tested animals and sensitive humans by better accounting for toxicokinetics and dynamics (Renwick, 1993), or (ii) separating risk assessment from risk management by using "most plausible" rather than "conservative" point estimates of each adjustment (Lewis, Lynch and Nikiforov, 1990). Calabrese (1985), Hattis, Erdreich, and Ballew (1987) and Hattis and Silver (1994) have critically assessed and quantitatively described human heterogeneity in response to toxic agents. Despite these efforts at analytical improvement, a lingering concern is that the critical number supplied to risk managers, the reference dose (RfD), is expressed as a point estimate, without any quantitative indication of how much confidence should be placed in this number.

In this paper, we propose an alternative approach to noncancer risk assessment that explicitly quantifies uncertainty about the human population threshold. The approach builds on our related studies of the uncertainty in estimates of cancer potency (Evans *et al.*, 1994a; Evans *et al.*, 1994b). The approach is illustrated using roughly the same database that EPA relies on to inform its current approach for noncancer risk assessment.

We begin by recognizing that any realistic strategy for risk assessment will be based on "imperfect" data. Any attempt to estimate the population threshold for humans is fraught with uncertainty (because of poorly understood differences in pharmacokinetics and pharmacodynamics between the test species and humans, and because of heterogeneity in these processes among human populations). Thus, estimates of the human population threshold will always involve uncertainty; and the degree of uncertainty may vary from chemical to chemical in a manner that is dependent on the properties of the chemical and the sources of evidence used.

We seek to develop a framework that characterizes this uncertainty about the human population threshold, that can be easily updated to reflect new sources of data and that encourages a separation of risk assessment and risk management.

Noncancer risk assessment is currently based on the assumption that a biologic threshold dose must be exceeded before exposure to a chemical causes effects. Under this view, if a person is exposed to a dose below their threshold of response, no effect is experienced and thus no risk is involved. However, there is variability in the sensitivity of individuals to chemicals and thus in order to assess population risks we must account for the responses of sensitive individuals. Unfortunately, we rarely have human epidemiologic data and, even when we do, the data may be inadequate to determine with any precision the threshold for particularly sensitive individuals. Hence, the human threshold dose is typically extrapolated from animal data. Animal test data provide an estimate of the subthreshold dose, or NOAEL. Because we do not know the true relationship between this animal subthreshold dose and the human threshold, uncertainty is inherent in the extrapolation.

The method we propose for characterization of uncertainty builds upon the current approach to noncancer risk assessment employed by the USEPA. In that approach the human subthreshold dose is estimated by dividing the NOAEL (of the most sensitive species tested) by a series of uncertainty factors (Barnes and Dourson, 1988). The resulting Reference Dose (RfD), or Reference Concentration (RfC) for inhalation exposures,[1] is defined by Barnes and Dourson (1988) 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 appreciable risk of deleterious effects during a lifetime.

The RfD is derived using the following equation:

$$\mathrm{RfD} = \frac{\mathrm{NOAEL}}{\mathrm{UF_A} * \mathrm{UF_H} * \mathrm{UF_S} * \mathrm{UF_L} * \mathrm{D} * \mathrm{MF}} \qquad (1)$$

Where the extrapolation from an animal NOAEL to an average human NOAEL is accounted for by the uncertainty factor UF_A; extrapolation from average to sensitive humans is accounted for by UF_H; UF_S is applied when a chronic NOAEL must be estimated from a critical study of less than chronic duration; UF_L adjusts a LOAEL to a NOAEL in cases where a NOAEL was not observed in the critical study; D, the data completeness factor, accounts for uncertainty that occurs when a NOAEL must be based on an incomplete database; and MF, the modifying factor accounts for any residual uncertainty (Barnes and Dourson, 1988). The uncertainty factors usually take on values of 10 and are intended to incorporate both an "adjustment" and a "margin of safety" (Dourson and Stara, 1983). By using large uncertainty factors, the EPA intends to guard against the possibility that the true threshold for the human population is below the computed RfD.

While this approach has a long history of use (see, for example, Lehman and Fitzhugh, 1954), the use of default safety factors has several potential weaknesses—the amount of protection offered by any RfD is unknown; the level of protection may differ from chemical to chemical; and risk assessment and risk management are inappropriately combined.

APPROACH

The fundamental modification of the EPA approach that we propose is simple, and relies on probabilistic characterization of the uncertainty in each step of the extrapolation from the animal NOAEL to the estimated human threshold. The overall uncertainty in the resulting estimate of the human threshold dose is assessed using standard approaches for the analysis of propagation of uncertainty. As Figure 1 illustrates, the overall uncertainty in the final estimate of the human population threshold reflects the combined influence of uncertainties inherent in each of the fundamental elements of the analysis.

Thus, a probabilistic characterization of the uncertainty in the human population threshold (PT) would be obtained by evaluating the propagation of uncertainty in the relationship:

$$PT = \frac{NOAEL}{AF_A * AF_H * AF_S * AF_L * AF_D * MF} \quad (2)$$

where the NOAEL, a surrogate for the experimental threshold dose, is modified by a series of adjustment factors (AF), to yield an estimate of the human population threshold, PT. We use the term adjustment factors, as opposed to "uncertainty factors" or "safety factors," to emphasize that such factors inherently involve probabilistic adjustments necessary to the interpretation of animal data. Any attempt to provide point estimates of the required adjustments involves both science and policy. By characterizing these adjustments probabilistically and using them to derive a probabilistic characterization of the human population threshold, we seek to encourage the separation of science and policy.

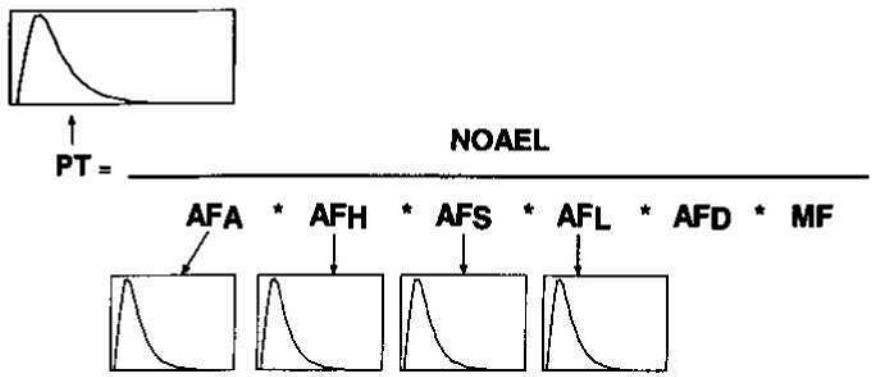

Figure 1. Conceptual illustration of approach.

The key to implementation of the approach is, of course, characterization of the uncertainty in each of the adjustment factors. In the absence of better information, it would seem to make sense to use historical information to characterize these uncertainties. Below, we briefly summarize the available data relevant to each extrapolation and use them to derive initial characterizations of the attendant uncertainties.

INPUTS

To illustrate the approach, we have relied primarily on evidence cited by Dourson and Stara (1983) in their justification of the point estimates of the uncertainty factors currently used by the EPA. However, rather than using the evidence to specify point estimates of the adjustment factors, we rely on it as a source of information about the uncertainty inherent in each term. The sections that follow summarize our characterizations of uncertainty in the four major adjustment factors.

Extrapolation from Animal to Human - AF_A

The animal-to-human adjustment factor, by definition, is the value by which the NOAEL observed in an animal study must be adjusted to yield the corresponding human NOAEL. Since the adjustment factors appear in the denominator of Equation 2, the appropriate adjustment factor is:

$$AF_A = d_a/d_h \tag{3}$$

where d_a is the dose in animals which yielded the NOAEL and d_h is the corresponding dose in humans.

It is not uncommon to assume that interspecies differences in sensitivity can be accounted for by normalizing the administered dose to body surface area, a proxy for basal metabolism (Freireich *et al.*, 1966; Crouch, 1983; Davidson, Parker, and Beliles, 1986). This approach, which was discussed as early as 1949 by Adolph (1949), is the basis for the interspecies extrapolation procedure recommended by Dourson and Stara (1983).

Under the surface area scaling assumption, the risk in one species exposed to an administered dose, d_1, is the same as the risk in another species exposed to an administered dose, d_2, as long as:

$$d_2 = d_1 * (sa_2/sa_1) \tag{4}$$

where d is the amount of contaminant (mg) administered per day and sa is body surface area (m^2). Recognizing that body surface area is roughly proportional to the 2/3 power of body weight, this relationship is commonly expressed as:

$$d_2 = d_1 * (bw_2^{2/3}/bw_1^{2/3}) \tag{5}$$

In many cases NOAELs reported from animal experiments are normalized to body weight (bw), *d'* (mg/kg/day). Similarly, regardless of the way they have been derived, estimates of human threshold doses are commonly expressed in terms of body weight. The appropriate adjustment factor in this case is:

$$d_2' = d_1' * (bw_1^{1/3}/bw_2^{1/3}) \tag{6}$$

Thus, under the assumption that risks in animals and humans are equivalent when doses are normalized to body surface area, the animal to human adjustment factor is:

$$AF_A = bw_2^{1/3}/bw_1^{1/3} \quad (7)$$

Values of AF_A for a number of common test species are summarized in Table 1. These were computed using a nominal weight of 70 kg for humans and typical weights of representative laboratory animals for each species listed.

While these point estimates may adjust for typical differences in the size and metabolism of various species, they are imperfect. Clearly no simple scaling factor can account perfectly for the complex, and perhaps chemical- and species-dependent interactions of pharmacokinetics and pharmacodynamics that are responsible for interspecies differences in sensitivity.

In theory, the uncertainty introduced by this extrapolation could be assessed by comparing estimates of human NOAELs derived in this way with observations of the true human NOAELs. In practice, we must rely on data from studies in which NOAELs have been derived experimentally for the same compounds in two or more animal species. An indication of the degree of uncertainty in the extrapolation may be obtained by examining the differences between, or ratios of, observed and predicted NOAELs:

$$r_{ij} = d_i/d_{ij} \quad (8)$$

where r_{ij} is the ratio of the observed NOAEL in species i, d_i, to the estimated NOAEL for species i based on extrapolation from species j, d_{ij}. Data from Dourson, Knauf, and Swartout (1992) for 69 pesticides tested in mice, rats and dogs were analyzed in this way. Estimated NOAELs (mg/kg/day) were derived using surface area scaling (*i.e.*, Equation 7) and nominal body weights for each species, which are summarized in Table 1. Both the bias and the imprecision in the extrapolation can be evaluated by examining the distribution of ratios, r_{ij}, obtained across the entire set of chemicals. Table 2 gives the medians and the geometric standard deviations of the ratios obtained from analysis of three different extrapolations—*i.e.*, mouse to rat, rat to dog, and mouse to dog.

If surface area scaling were perfect (and if NOAELs were estimated without error), both the median and the geometric standard deviation of the distribution of ratios would equal one. In fact, these data suggest that there are considerable uncertainties in estimates of NOAELs in one species derived by rescaling NOAELs observed in other species. The actual uncertainty in extrapolating from mice or rats to humans is likely to be at least as large as the uncertainty in extrapolating among mice, rats, and dogs. As a first approximation of the uncertainty in scaling from animals to humans, we have used the estimate derived from the mouse to dog data because this extrapolation involves the largest range of body weights (0.03 kg to 16 kg).

In summary, the animal to human adjustment factor, AF_A, may be characterized as approximately lognormal with a species dependent median, given in Table 1, and geometric standard deviation of about 5.

Table 1. Animal to Human Adjustment Factors: Based on Surface Area Extrapolation

Species	Body Weight bw (kg)	Adjustment Factor AF_A (dimensionless)
Mouse	0.03[a]	13.3
Hamster	0.125[a]	8.2
Rat	0.35[a]	5.8
Guinea Pig	0.8[b]	4.4
Hen	1.6[b]	3.5
Rabbit	4.0[b]	2.6
Monkey	7.0[b]	2.2
Dog	16[a]	1.6
Pig	48[b]	1.1
Human	70[c]	1

[a] From Gold *et al.* (1984).
[b] From Altman and Dittmer (1962).
[c] From EPA Exposure Factors Handbook (USEPA, 1989).

Average Human to Sensitive Human - AF_H

It is commonly assumed that the human population is more heterogeneous, with regard to sensitivity to chemicals, than are the inbred populations of laboratory animals used to study chemical toxicity (Dourson and Stara, 1983; Calabrese, 1985; and Hattis and Silver, 1994). This would be of little concern if our goal was to protect "typical" individuals. However, as illustrated below, differences in heterogeneity are critical in efforts to protect more sensitive individuals.

The significance of differences in sensitivity can perhaps best be illustrated graphically. Figure 2 demonstrates that as a population becomes more heterogeneous the ratio of the ED_{50}, the dose below which 50% of the population responds, to the ED_{01}, the dose below which 1% of the population responds, increases. If, for example, individual thresholds of response are lognormally distributed, then the ratio of the ED_{50} to the ED_{01} is given by:

$$ED_{50}/ED_{01} = \sigma_g^{2.33} \quad (9)$$

where σ_g is the geometric standard deviation of the distribution of individual thresholds (*i.e.*, the dose-response curve) and 2.33 is the z score corresponding to the lower 1% of the lognormal distribution. More generally, under this assumption, the ratio of the ED_{50} to the ED_f, the dose below which only f% of the population responds, is given by:

$$ED_{50}/ED_f = \sigma_g^{zf} \quad (10)$$

Table 2. Estimates of Uncertainty in Interspecies Extrapolation: Based on Surface Area

Pair of Species	Number of Pairs	Distribution of Ratios (d_i/d_{ij}) Median[a]	GSD
Mouse/Rat	37	0.6[b]	4.3
Rat/Dog	77	2.5[c]	4.1
Mouse/Dog	35	1.4[d]	4.9

[a] Although two of these three medians are significantly different from 1 (as determined by t-tests with p = 0.05), no adjustment in the median animal to human scaling factor was made.
[b] $t = 2.17$; $p < 0.05$.
[c] $t = 5.68$; $p < 0.001$.
[d] $t = 1.16$; not significant.

where all terms retain their previous definitions and z_f is the z score corresponding to the lower f% of the lognormal distribution. Table 3 gives values of z corresponding to several fractiles of potential interest.

Differences in heterogeneity could easily be accounted for if the distributions of individual thresholds in both animals and humans were known. For example, if individual thresholds were lognormally distributed in both the test species and in humans with geometric standard deviations σ_a and σ_h, respectively, then the adjustment required to allow for the extra heterogeneity among humans would be simply:

$$ED_{f,h}/ED_{f,a} = \frac{ED_{50,h}/\sigma_h^{z_f}}{ED_{50,a}/\sigma_a^{z_f}} = \frac{ED_{50,h}}{ED_{50,a}} * \frac{\sigma_a^{z_f}}{\sigma_h^{z_f}} \quad (11)$$

This expression consists of two terms. The first, the ratio of ED_{50}s in animals and humans, reflects interspecies extrapolation and in theory has been accounted for by the animal to human adjustment factor, AF_A. The second, which involves the ratio of σs, reflects any differences in heterogeneity of response between laboratory animals and humans.

Unfortunately, relatively little is known about either the distribution of individual human thresholds or the relative degree of heterogeneity of human and laboratory animal populations. Thus, all efforts to account for heterogeneity in the distribution of human sensitivity to chemical toxicity are inherently uncertain.

For certain laboratory animals, there are data which provide some insight about heterogeneity of response. For example, Weil (1972) reports data on the slopes of dose-response curves from nearly 500 experiments involving laboratory rats. More specifically, Weil (1972) gives the slope of the best fitting log-probit dose response function from each LD_{50} experiment. These values were relied on by Dourson and Stara (1983) in their derivation of the EPA's 10-fold "uncertainty factor."

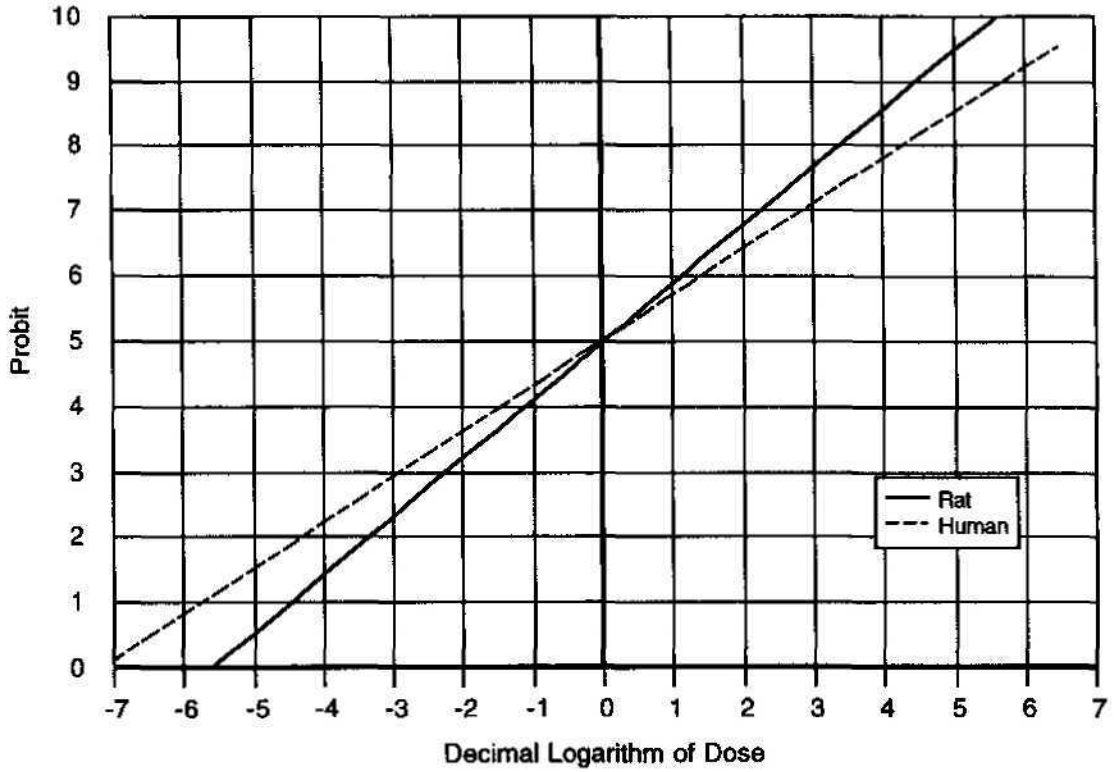

Figure 2. Accounting for heterogeneity in sensitivity.

Table 3. Z-scores Corresponding to Certain Fractiles of the Lognormal Distribution

Fractile	Z-value
1 %	2.33
0.1 %	3.09
0.01 %	3.72

As indicated below, these values can be directly converted to estimates of the geometric standard deviation of the distribution of individual thresholds among laboratory rats. Typically, the slope, b, of a log-probit dose response function is defined as:

$$b = \frac{p[F(d_2)] - p[F(d_1)]}{\log_{10}(d_2) - \log_{10}(d_1)} \tag{12}$$

where p[ ] is the probit, or z-score plus 5, corresponding to the value of the cumulative density function F (the population risk) evaluated at dose d and $\log_{10}(d)$ is the decimal logarithm of the dose d. Using this definition, the geometric standard deviation of the underlying distribution of individual thresholds is:

$$\sigma_g = 10^{(1/b)} \tag{13}$$

Given the limited information available about either the heterogeneity of human sensitivity or the relative heterogeneity of laboratory animals and humans, several approaches seem possible. The simplest perhaps is to directly use the information about heterogeneity of sensitivity among laboratory animals as a surrogate for the heterogeneity of humans. Note that under this approach, no adjustment is necessary for the differential heterogeneity of animals and humans because no difference in heterogeneity is explicitly postulated. A second alternative is to simply make an assumption about the relative heterogeneity of laboratory animals and humans, *i.e.*, σ_h/σ_a, and to use this ratio in the evaluation of Equation 11. Table 4 illustrates the impact of alternative assumptions about relative heterogeneity on the ratio of $ED_{f,h}/ED_{f,a}$.

In reality the average human to sensitive human adjustment factor, AF_H, is intended to adjust for two quite distinct concerns—(i) any genuine differences in heterogeneity between laboratory animals and humans, and (ii) the limitations in the statistical power of laboratory bioassays involving relatively small numbers of animals.

Typically animal bioassays involve between 10 and 50 animals in each of 2 or 3 dose groups, and a corresponding control. With studies this small, the risk at the NOAEL could easily be in the range of 1 to 10% (Gaylor, 1992; Leisenring and Ryan, 1992). If the average human to sensitive human adjustment factor, AF_H, accounted only for differences in relative heterogeneity, then the resulting estimate of the human threshold would be expected to involve risks in this same range (*i.e.*, 1 to 10%) for human populations.

Table 4. Impact of Relative Heterogeneity on Ratio of ED_f in Humans and Laboratory Animals

Relative Heterogeneity	$ED_{f,h}/ED_{f,a}$ for Various Values of f		
σ_h/σ_a	f = 0.01	f = 0.001	f = 0.0001
1.2	1.5	1.8	2.0
1.5	2.6	3.5	4.5
2.0	5.0	8.5	13.0

Historically, regulatory agencies have been hesitant to use thresholds which implicitly involved risks this high. For example, in their derivation of the average to sensitive human "uncertainty factor," Dourson and Stara (1983) use a z_f of 3 stating that "this places the median response in the general range expected for a potential sensitive subgroup of the population under study." Note that, under the assumption that individual thresholds are lognormally distributed, setting the RfD 3 standard deviations below the ED_{50} would correspond to protecting about 99.9% of the population. Alternatively, if the RfD is set 3 standard deviations below the NOAEL (assumed to represent about 3% risk), this choice would protect more than 99.999% of the population.

Once a decision has been made about the target level of risk in humans, f_h, and assumptions have been made about: (i) the level of risk corresponding to the animal NOAEL or LOAEL, f_a; and (ii) the relative heterogeneity of humans and animals, σ_h/σ_a, the average to sensitive human adjustment factor can be easily evaluated using:

$$AF_H = ED_{f_a,a}/ED_{f_h,h} = \frac{ED_{50,a}}{ED_{50,h}} * \frac{\sigma_h^{zf_a}}{\sigma_a^{zf_h}} \qquad (14)$$

In view of the unresolved issues concerning the target level of protection in humans and the relative heterogeneity of humans and laboratory animals, we used two approaches for characterizing AF_H.

The first, which is most similar to that used by Dourson and Stara (1983) in their derivation of UF_H, uses the Weil (1972) data on heterogeneity in rats to directly characterize both σ_a and σ_h and characterizes the difference between $z_{f,h}$ and $z_{f,a}$ as 3. Implicitly this assumes that: (i) there is no difference in heterogeneity between laboratory animals and humans, and (ii) that the adjustment is applied to the NOAEL, and that the target level of protection in humans is on the order of 1/100,000 or lower. The second uses the Weil (1972) data to characterize the heterogeneity in the responses of laboratory animals, σ_a, makes an explicit assumption that humans are 50% more heterogeneous than laboratory animals, *i.e.*, $\sigma_h/\sigma_a = 1.5$, assumes that the risk at the NOAEL is on the order of 3%, *i.e.*, $z_{f,a} = 1.9$, and assumes that the target level of protection among humans is about 1/1000, *i.e.*, $z_{f,h} = 3.1$. In both cases, the distribution of estimates of σ_a is generated by sampling from the empirical distribution of log-probit slopes in the Weil data and converting these to values of σ using Equation 13.

Extrapolation from Subchronic to Chronic Exposure - AF_S

For some chemicals, results are not available from chronic bioassays. In these cases, a NOAEL from a subchronic bioassay may be used as a proxy for the NOAEL that would be expected from a chronic bioassay. Dourson and Stara (1983) used data compiled by Weil and McCollister (1963) for 22 substances which had been tested in both chronic (2 year) and subchronic (30-210 days) bioassays as a basis for justifying the EPA's 10-fold subchronic to chronic "uncertainty factor". These same data, which provide paired chronic and subchronic NOELs (No Observed Effect Levels), were augmented with data for 29 compounds from Lewis (1994). Although the combined dataset included information on both NOAELs and LOAELs, we restricted our attention to the NOAELs.

The distribution of ratios of subchronic to chronic NOAELs (or NOELs) in these data was approximately lognormal, with median 2.1 (*i.e.*, $e^{0.71}$) and geometric standard deviation 2.5 ($e^{0.75}$).[2]

Extrapolation from LOAEL to NOAEL - AF_L

For certain compounds, all of the doses tested in the critical bioassay yield statistically significant results. In these cases, because there is no observed NOAEL, the derivation of the estimated human threshold relies on the LOAEL (Lowest Observed Adverse Effects Level) as a surrogate for the NOAEL. To account for this substitution of the LOAEL for the NOAEL, an additional adjustment factor, AF_L, must be used. By definition, the appropriate adjustment is:

$$AF_L = \text{LOAEL/NOAEL} \tag{15}$$

One approach for estimating the proper adjustment is to examine the relationship between LOAELs and NOAELs in previous studies. Data from Weil and McCollister (1963), which were used by Dourson and Stara (1983) in the derivation of the 10-fold "uncertainty factor", UF_L, and from Lewis (1994) were used to examine the relationship of the LOAEL to the NOAEL. Only data from chronic bioassays were used in our analysis. The ratio of the LOAEL to the NOAEL was computed for each of the 78 LOAEL/NOAEL pairs from the 55 chemicals represented in the combined dataset.

The distribution of LOAEL/NOAEL ratios was approximately lognormal with median of 3.4 (*i.e.*, $e^{1.23}$) and geometric standard deviation of 1.7 ($e^{0.51}$).[3]

Data Completeness Factor - D and Modifying Factor - MF

Our basic analysis excluded compounds for which the EPA had assigned data quality and/or modifying factors. Although we have not considered these adjustments, approaches similar to those outlined above could readily be adapted to characterize the uncertainty introduced by the need to rely on incomplete datasets.

ANALYSIS OF THE PROPAGATION OF UNCERTAINTY

To illustrate the approach, distributional characterizations of the human population threshold dose were computed for each chemical with either a RfD or RfC listed in EPA's Integrated Risk Information System (IRIS) as of June 1994

Table 5. Summary of Adjustment Factors

Adjustment Factor	Symbol	Median	GSD	Data Sources
Animal to Human	AF_A	—[a]	4.9	[b]
Human Heterogeneity	AF_H	—	—	[c]
Basic Approach[d]		2.7	2.3[e]	
Alternative Approach[f]		5.3	1.4[e]	
Subchronic to Chronic	AF_S	2.0	2.1	[g,h]
LOAEL to NOAEL	AF_L	3.4	1.7	[g,h]

[a] Species specific, see Table 1.
[b] Dourson *et al.*, (1992).
[c] Weil (1972).
[d] Assumes animals and humans are equally heterogeneous, and that the target level of protection is on the order of 1 in 100,000.
[e] This is a pseudo-GSD—the square root of the 84th percentile divided by the 16th percentile. The actual distribution, which is not lognormal, is characterized nonparametrically in our analyses.
[f] Assumes humans are 1.5 times as heterogeneous as animals and that the target level of protection is on the order of 1 in 1000.
[g] Weil and McCollister (1963).
[h] Lewis (1994).

(USEPA, 1994). The dataset included 348 chemicals for which an RfD or RfC had been derived from animal bioassays. Of these, 106 had been identified by the EPA as involving incomplete datasets. The vast majority of the 242 remaining cases involved the use of animal data to establish an RfD. These were the focus of our analysis.

For each chemical, the NOAEL (or LOAEL) for the critical study reported in IRIS was used as the basis for our calculations. Data reported in IRIS were used to determine whether the NOAEL (or LOAEL) came from a chronic or subchronic bioassay, and to determine whether additional adjustments were necessary.

For 126 of these compounds NOAELs were available from chronic bioassays. For these, only two adjustments, AF_A and AF_H, were needed to derive estimates of the human threshold. For another 70 compounds, NOAELs were available from subchronic bioassays and one additional adjustment, AF_S, was necessary. For 25 compounds, chronic bioassays had yielded LOAELS. For these chemicals three adjustments—AF_A, AF_H, and AF_L—were needed. For the remaining 10 compounds, subchronic bioassays had yielded LOAELs and all four adjustments—AF_A, AF_H, AF_S and AF_L—were required.

A probabilistic characterization of the human threshold dose for each compound was developed using Equation 2. In these calculations, each adjustment factor was treated as a random variable with a distribution specified in Table 5. It was assumed that all of the adjustments were statistically independent, *i.e.*, no correlations were induced.

Propagation of uncertainty was evaluated using Monte Carlo simulation. Simulations were performed using Crystal Ball™ in an Excel spreadsheet (Decisioneering Inc., 1994; Microsoft Corp., 1992). Runs of one thousand iterations were found to yield replicate distributions with acceptable levels of error in the tails.

RESULTS

This approach has the potential to yield a variety of classes of results. Some results are chemical specific. Others are generic. Both classes of results are described below.

Chemical Specific Result — Acetone

One potential application of the approach would be to analyze the uncertainty in estimates of the human threshold for a specific chemical. To illustrate this use, a probabilistic characterization of the human threshold for acetone has been developed. The NOAEL for acetone is derived from a subchronic bioassay in rats and thus, three adjustments are needed: (i) an interspecies extrapolation, (ii) an adjustment for differential heterogeneity of laboratory rats and humans, and (iii) an adjustment to account for the use of data from a subchronic study as a surrogate for results from a chronic study.

The result is shown in Figure 3. The extent of uncertainty in the estimate of the human threshold is striking—a 90% confidence region for the true human threshold spans the interval from 0.1 mg/kg/day to 70 mg/kg/day. According to our analysis, the median value of the distribution of possible human population thresholds is about 3 mg/kg/day.[4]

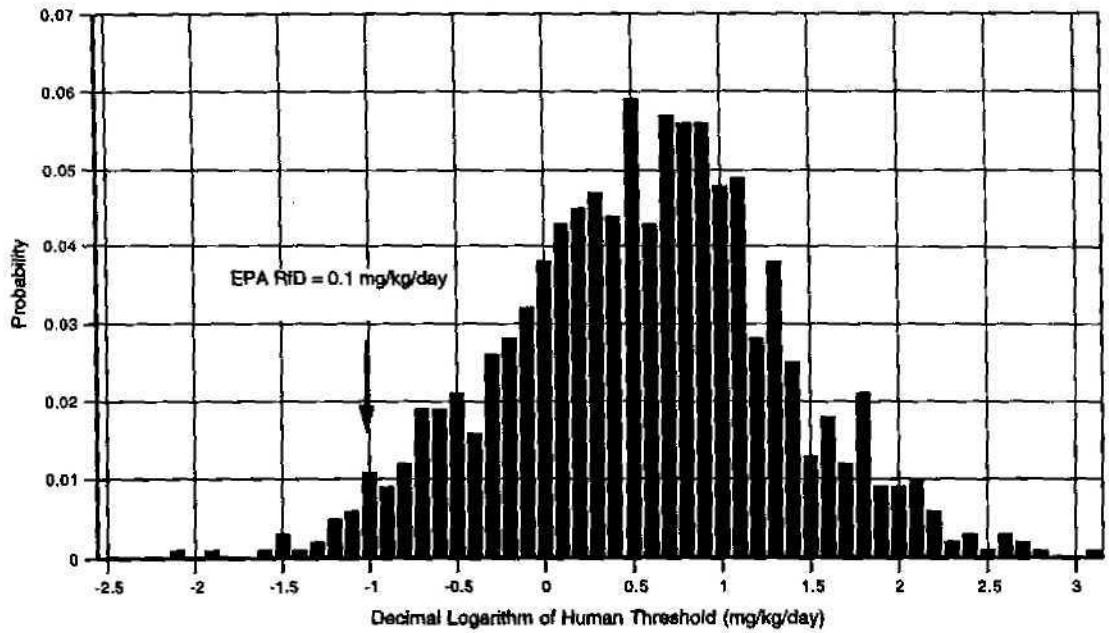

Figure 3. Probabilistic characterization of the human threshold dose for acetone.

The EPA computed the RfD for acetone by dividing the NOAEL (from the subchronic bioassay in rats) of 100 mg/kg/day by three 10-fold "uncertainty factors." These were intended to account for animal to human extrapolation, interspecies variability among humans, and use of a subchronic bioassay to estimate the chronic NOAEL. Using this procedure, the EPA obtained an RfD for acetone of 0.1 mg/kg/day. This value falls just below the 4th percentile of the distribution of plausible values of the human threshold.

Generic Results

In principle, there are an infinite number of possible distributions of the aggregate adjustment factor. However, if we restrict our attention to the default characterizations of uncertainty in the four adjustment factors discussed above, there are only 16 possible results. And of these 16, four are of primary interest: (i) use of a NOAEL from a chronic bioassay, requiring two adjustments, AF_A and AF_H; (ii) use of a NOAEL from a subchronic bioassay, requiring three adjustments, AF_A, AF_H, and AF_S; (iii) use of a LOAEL from a chronic bioassay, requiring three adjustments, AF_A, AF_H, and AF_L; and (iv) use of a LOAEL from a subchronic bioassay, requiring all four adjustments, AF_A, AF_H, AF_S, and AF_L.

The cumulative density functions (CDFs) for the PT distributions for these four combinations of adjustment factors are shown in Figure 4. To simplify the analysis and presentation, the NOAEL was set at 1 mg/kg/day. To derive chemical specific results, it is necessary to: (1) multiply the results by the NOAEL (or LOAEL) for the chemical of interest; (2) divide the results by the appropriate AF_A value for the species tested (*i.e.*, 5.8 for rats, 13.3 for mice); and (3) divide by modifying and data quality factors, if necessary.

In theory, these CDFs could be used by risk managers to discern the likelihood that various proposed levels of exposure were below the human population threshold. Table 6 indicates the degree of adjustment of the NOAEL (or LOAEL), after interspecies scaling, that is needed to achieve 50%, 95%, and 99% confidence that the true human threshold is above a proposed reference dose level.

As might be expected, larger adjustments are necessary for data from either subchronic studies or from experiments in which all doses yielded significant risks. For example, to achieve 95% confidence a NOAEL from a chronic study must be divided by 50, whereas a LOAEL from a subchronic study must be divided by 484. The difference between the required adjustments increases as the level of desired confidence increases.

Comparison of the adjustment factors from the "basic" approach (given in the body of Table 6) with those from the "alternative" approach (given as subscripts) indicates that they are quite similar. It is important to realize, however, that these values result from two quite different assumptions about the relative heterogeneity of animals and humans, and reflect target risks that vary by several orders of magnitude.

How Protective are EPA's Current RfDs?

It is interesting to consider where in these distributions the current EPA RfDs fall. As Figure 5 illustrates, none of the RfDs included in our analysis fall above the 30th percentile of the distribution of possible human threshold doses, and most are below the 10th percentile.

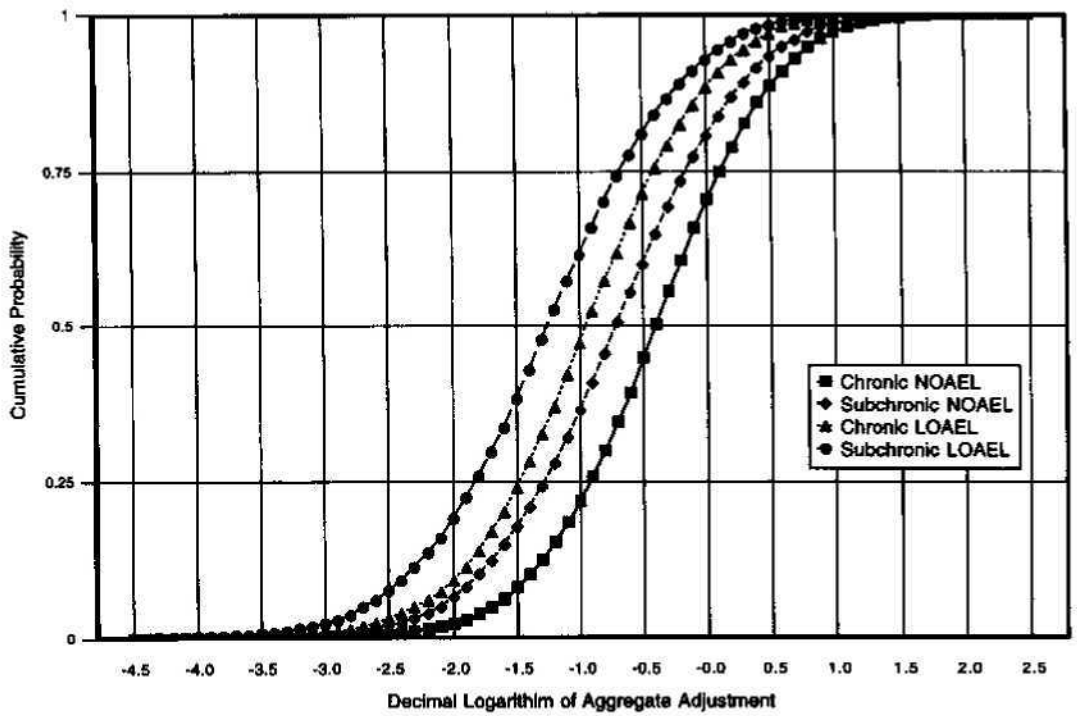

Figure 4. Distributions of the aggregate adjustment factor.

Table 6. Degree of Adjustment of Scaled NOAEL (or LOAEL) Required to Achieve Stated Confidence Level[a]

	Source of Data			
Level of Confidence	NOAEL Chronic	NOAEL Subchronic	LOAEL Chronic	LOAEL Subchronic
50%	$3_{(5)}$	$5_{(10)}$	$9_{(17)}$	$18_{(31)}$
95%	$50_{(63)}$	$126_{(184)}$	$192_{(253)}$	$484_{(642)}$
99%	$220_{(194)}$	$586_{(623)}$	$825_{(835)}$	$2261_{(2359)}$

[a] Values given in the body of the table are derived using the "basic" approach for analyzing differential heterogeneity. Those given in parentheses as subscripts reflect the "alternative" approach.

Table 7 examines this same issue in a different way — asking what fraction of RfDs are at or below the 5th percentile of the distribution. For these, a risk manager would have at least 95% confidence that the true human threshold dose was above the RfD.

Table 7. Fraction of RfDs within Lower 5% of Distribution of Potential Human Threshold Values

	Species		
All	Mice	Rats	Dogs
56% (129/231)	23% (3/13)	39% (57/146)	98% (64/65)

Note that of the 231 RfDs evaluated in our study, 56% were below the 5th percentile of the distribution. However, this fraction varied strongly, depending on the source of data supporting the RfD. For RfDs derived from studies of mice, only 23% of the values were in the lowest 5% of the distribution. In contrast, for RfDs derived from studies of dogs, nearly 98% of the values were in this region.

These species-dependent differences in apparent protection arise because the EPA uses the same uncertainty factor to extrapolate from animals to humans regardless of the species involved. If surface area scaling is appropriate (as our analysis assumes), different adjustments are necessary for different species.

DISCUSSION

All attempts to estimate human population threshold doses based on studies of toxicity in laboratory animals are fraught with uncertainty. The degree of uncertainty in the extrapolation is itself uncertain, but depends in part on the nature of the animal study. Generally, a NOAEL from a chronic animal study provides the strongest

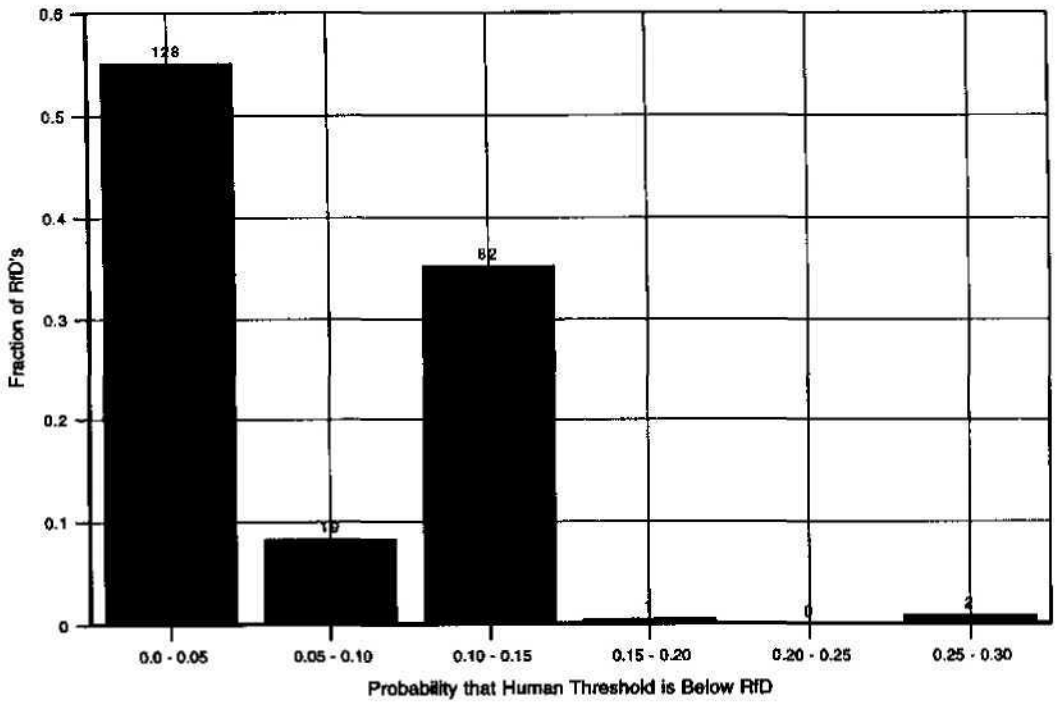

Figure 5. Location of EPA RfDs in the distribution of population thresholds.

evidence for estimating the human threshold. For many chemicals, the chronic NOAEL is not available and must be estimated from either a subchronic NOAEL or a chronic LOAEL. These additional extrapolations involve additional uncertainty.

In the face of uncertainty, any approach for deriving a point estimate of the human population threshold (or any related quantity, such as the RfD, an acceptable daily dose, etc.) inherently involves both scientific assessment and policy judgment. This mixture of science and policy is problematic for decision-makers because the relative impacts of scientific assessment and policy judgment are not readily evident.

This paper presents an alternative approach for noncancer risk assessment which attempts to probabilistically characterize the human population threshold for adverse effects, and argues that this approach is preferable to current approaches because it provides decision-makers information about the uncertainty in estimates of the human population threshold, and encourages them to explicitly consider the social trade-offs inherent in decisions about appropriate levels of human exposure to noncarcinogens. In this way, the new approach attempts to clearly distinguish scientific assessment of uncertainty in estimates of the human population threshold from policy judgments about acceptable levels of risk and appropriate degrees of conservatism in establishing exposure limits or selecting pollution control alternatives.

We have illustrated the approach using much of the same data that is currently used to inform EPA's approach to RfD determination. Clearly, there are limitations in these data. As noted below, the approaches that we have taken to the interpretation of these data are meant to be illustrative rather than definitive.[5]

The extrapolation from animals to humans assumes that when doses are normalized to surface area, equivalent risks occur. The average to sensitive human extrapolation uses data from studies of rats as a proxy for the heterogeneity of human populations. The LOAEL to NOAEL extrapolation relies on the historical ratios of LOAELs to NOAELs as a measure of the needed adjustment. The subchronic to chronic adjustment does not differentiate chemicals according to mechanism of toxicity or other factors which might influence the dynamics of dose delivery or disease development.[6] Clearly all of these assumptions are open to question.

One might be tempted to conclude that these limitations of data and theory present insurmountable barriers to the implementation of the new approach proposed in this paper. However, it is important to recognize that these same limitations impinge on any approach for estimating the human population threshold from animal data. In fact, it is these uncertainties that our probabilistic approach is designed to accommodate. Thus, it is not unreasonable to argue that the new approach is at least as good as, and likely better than, the current approach in the face of uncertainty — and perhaps that the relative advantage increases with increasing uncertainty.

While the new approach may be logically superior to the current approach, it will require decision-makers and the public to squarely face several new issues. They will no longer be able to defer to the judgments of scientists and risk assessors about acceptable levels of risk or about appropriate degrees of confidence that exposures are below human population thresholds for adverse effects. Instead, decision-makers and the public will have to directly confront these tough value-laden issues themselves.

In view of these concerns, we do not recommend an abrupt transition from current, point-value based, approaches to this new probabilistic approach. Rather, we recommend that initially the new approach be applied in parallel with the standard approach and that during this transition period: (i) workshops be held in which decision-makers and the public explore the relative strengths and weaknesses of the new approach, and (ii) toxicologists, epidemiologists, and risk assessors work diligently to improve the approaches for characterizing the uncertainty inherent in each of the four key extrapolations. More specifically, these scientists should consider that:

(i) It is by no means clear that surface area extrapolation is the best default approach for scaling from animals to humans. First, it has been demonstrated that currently available data cannot statistically distinguish between surface area and body weight scaling (Watanabe, Bois, and Zeise, 1992). Second, biologically based measures of "delivered" dose or "biologically effective" dose, reflecting chemical- and species-specific information about pharmacokinetics and pharmacodynamics, may offer great potential to improve interspecies scaling and thereby to reduce the uncertainty inherent in any simple allometric adjustment.[7] Certainly, additional work to explore these issues and to better characterize the uncertainty in such extrapolations is justified.

(ii) Similarly, the reliance on variation across chemicals in heterogeneity of response of rats as a measure of human heterogeneity is particularly weak. As the comparison of results from our basic and alternative approaches indicate (see Table 6), the assumptions made about the differential relative heterogeneity of populations of laboratory animals and humans can substantially influence one's view about the degree of protection afforded by any specified exposure limit. Additional work, both theoretical and empirical, is needed to adequately characterize human variability in sensitivity to chemical toxicity; to better understand the determinants of such variability; and to explore the impact of our limited understanding of this issue on our ability to estimate human population thresholds.

(iii) The use of historical LOAEL/NOAEL ratios to estimate a NOAEL from a LOAEL is not ideal. By definition, the LOAEL and NOAEL are restricted to be one of the doses selected for testing in the study. Typically, doses are spaced in fixed intervals (*e.g.*, 3-fold or 10-fold) and are established with reference to preliminary studies. Thus, the observed distribution of LOAEL/NOAEL ratios primarily reflects the historical frequency of use of various dose spacings. The recent work on the use of benchmark doses, as a supplement to (and eventual replacement for) NOAELs and LOAELs, may largely eliminate the need for the LOAEL to NOAEL extrapolation and can easily be accommodated in our framework (Barnes *et al.*, 1995; Kimmel and Gaylor, 1988).

(iv) The idea that the ratio of the chronic NOAEL to the subchronic NOAEL is the same for all chemicals and is independent of both the length of the subchronic study and the nature of the compound and effect of interest is questionable. Clearly as our understanding of the dynamics of exposure and toxicity advances, this default approach can be refined considerably.

Obviously as chemical (or chemical class) specific knowledge improves, it will be possible to both reduce uncertainty in estimating the human threshold and to better characterize the extent of uncertainty as a function of the nature and amount of information used as the basis for the estimate of the human threshold. However, until such information is developed, an approach based on generic adjustments and probabilistic characterizations of the uncertainty inherent in such generic adjustments would seem appropriate.

In summary, this paper has proposed and illustrated a new approach for noncancer risk assessment. Although the new approach is itself imperfect, we believe that it represents a logical next step in the continual evolution and refinement of regulatory approaches for characterizing noncancer risks.

Acknowledgements

We thank Drs. Michael Dourson, Jeff Swartout, and Steven Lewis for sharing their data and for many helpful and interesting discussions. Drs. Baird, Evans, and Shlyakhter were supported in part by a US EPA Cooperative Agreement (CR 81090-02-3) with Harvard for the Program on Environmental Health and Public Policy.

Endnotes

1. The methods and discussions presented here apply to both the RfD and RfC, but for ease of reading, the term RfD will be used throughout this paper.

2. When the data were separated by route of exposure, a marginally significant difference in the means of the distributions was found (oral—$\mu_{\ln x}$ = 0.62, $\sigma_{\ln x}$= 0.70, n = 56; inhalation—$\mu_{\ln x}$ = 1.06, $\sigma_{\ln x}$ = 0.87, n = 15; t = 2.05 with 69 df). In subsequent analyses, the characterization of the subchronic to chronic adjustment might be improved by treating oral and inhalation exposures separately.

3. The means of the LOAEL/NOAEL ratios were quite similar for the two exposure pathways (oral—$\mu_{\ln x}$ = 1.22, $\sigma_{\ln x}$ = 0.55, n = 61; inhalation—$\mu_{\ln x}$ = 1.27, $\sigma_{\ln x}$= 0.28, n = 17; t = 0.36 with 76 df).

4. These results for acetone were all derived using the "basic" approach for addressing differential heterogeneity. An analysis of the differences between these results and those obtained with the alternative approach is presented in the discussion of generic results.

5. For example, we have used lognormal distributions in the initial probabilistic characterizations of several of the adjustment factors. This choice was made after visual inspection of log probability plots and calculation of simple measures of goodness of fit. No formal comparison was made of the lognormal

characterizations with other equally plausible and perhaps better fitting alternative distributions. Future work should consider these aspects of the analysis more critically.

6. Furthermore, the historical data on ratios of subchronic to chronic NOAELS does not account for any differences in the severity of the disease end points measured in the subchronic and chronic studies.

7. The RfC human equivalent dose developed by the EPA (Jarabek *et al.*, 1990) is one example of a target tissue dose.

References

Adolph, E.F. 1949. Quantitative relations in the physiological constitutions of mammals. *Science.* 109, 579-585.

Altman, P.L. and Dittmer, D.S. 1962. *Growth Including Reproduction and Morphological Development,* Biological Handbooks, Federation of American Societies for Experimental Biology, Washington, DC.

Barnes, D.G., Daston, G.P., Evans, J.S., Jarabek, A.M., Kavelock, R.J., Kimmel, C.A., Park., C., and Spitzer, H.L. 1995. Benchmark dose workshop: criteria for use of a benchmark dose to estimate a reference dose. *Regul. Toxicol. Pharmacol.* 21, 296-306.

Barnes, D.G. and Dourson, M. 1988. Reference dose (RfD): descriptions and use in health risk assessments. *Regul. Toxicol. Pharmacol.* 8, 471-486.

Beck, B.D., Conolly, R.B., Dourson, M.L., Guth, D., Hattis, D., Kimmel, C., and Lewis, S.C. 1993. Improvements in quantitative noncancer risk assessment. *Fundam. Appl. Toxicol.* 20, 1-14.

Calabrese, E.J. 1985. Uncertainty factors and individual variation. *Regul. Toxicol. Pharmacol.* 5, 190-196.

Crouch, E.A.C. 1983. Uncertainties in interspecies extrapolations of carcinogenicity. *Environ. Health Perspect.* 50, 321-327.

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

Davidson, I.W.F., Parker, J.C. and Beliles, R.P. 1986. Biological basis for extrapolation across mammalian species. *Regul. Toxicol. Pharmacol.* 6, 211-237.

Decisioneering Inc. 1994. Crystal Ball Version 3.0 - User Manual. Decisioneering Inc., Boulder, CO.

Dourson, M.L., Knauf, L.A. and Swartout, J.C. 1992. On reference dose (RfD) and its underlying toxicity data base. *Toxicol. Indus. Health* 8, 3, 171-189.

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

Evans, J.S., Graham, J.D., Gray, G.M., and Sielken, R.L., Jr. 1994a. A distributional characterization of cancer risk from formaldehyde exposure. *Risk Anal.*14, 25-34.

Evans, J.S., Gray, G.M., Sielken, R.L., Jr., Smith, A.E., Valdez-Flores, C. and Graham, J.D. 1994b. Use of probabilistic expert judgment in uncertainty analysis of carcinogenic potency. *Regul. Toxicol. Pharmacol.* 20, 15-36.

Farland, W. and Dourson, M. 1993. Noncancer health endpoints: approaches to quantitative risk assessment. In: *Comparative Environmental Risk Assessment,* (Cothern, C.R., Ed.) Boca Raton, Fl, Lewis Publishers, pp. 87-106.

Freireich, E.J., Gehan, E.A., Rall, D.P., Schmidt, L.H., and Skipper, H.E. 1966. Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. *Cancer Chemother. Rep.* 50, 4, 219-244.

Gaylor, D.W. 1992. Incidence of developmental defects at the No Observed Adverse Effects Level (NOAEL). *Regul. Toxicol. Pharmacol.* 15, 151-160.

Gold, L.S., Sawyer, C.B., Magaw, R., Backmann, G.M., de Veciana, M., Levinson, R., Hooper, N.K., Havender, W.R., Berstein, L., Peto, R., Pike, M.C., and Ames, B.N. 1984. A carcinogenic potency database of the standardized results of animal bioassays. *Environ. Health Perspect.* 58, 9-319.

Habicht, F.H., II, Guidance on Risk Characterization for Risk Managers and Risk Assessors, Office of the Administrator, U. S. Environmental Protection Agency, Washington, DC, February 26, 1992.

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, 415-426.

Hattis, D. and Silver, K. 1994. Human interindividual variability—a major source of uncertainty in assessing risks for noncancer health effects. *Risk Anal.* 14, 421-431.

Hertzberg, R.C. 1989. Fitting a model to categorical response data with application to species extrapolation of toxicity. *Health Phys.* 57, Suppl. 1, 405-409.

Hertzberg, R.C. and Dourson, M.L. 1993. Using categorical regression instead of a NOAEL to characterize a toxicologist's judgment in noncancer risk assessment. Proceedings from Second International Symposium on Uncertainty Modeling and Analysis, April 25-28, 1993. Los Alamitos, CA, IEEE Computer Soc. Press, pp. 254-261.

Jarabek, A.M., Menache, M.G., Overton, J.H., Jr., Dourson, M.L. and Miller, F.J. 1990. The U.S. Environmental Protection Agency's inhalation RfD methodology: risk assessment for air toxics. *Toxicol. Indus. Health*, 6, 279-301.

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

Lehman, A.J., and Fitzhugh, O.G. 1954. 100-fold margin of safety. *Assoc. Food Drug Off. U.S. Quart. Bull.* 18, 33-35.

Leisenring, W. and Ryan, L. 1992. Statistical properties of the NOAEL. *Regul. Toxicol. Pharmacol.* 15, 161-171.

Lewis, S.C. 1994. Personal communication of results from CMA report.

Lewis, S.C., Lynch, J.R., and Nikiforov, A.I. 1990. A new approach to deriving community exposure guidelines from "no-observed-adverse-effect-levels." *Regul. Toxicol. Pharmacol.* 11, 314-330.

Microsoft Corp. 1992. *Excel version 4, User's Guide.* Microsoft Corporation, Redmond, WA.

Renwick, A.C. 1993. Data-derived safety factors for the evaluation of food additives and environmental contaminants. *Food Add. Contam.* 10, 3, 275-305.

SAB (Science Advisory Board letter to Reilly). 1990. Science Advisory Board's comments on the use of uncertainty and modifying factors in establishing Reference Dose Levels. EPA-SAB-EHC-90-005.

USEPA (U. S. Environmental Protection Agency). 1994. *The Integrated Risk Information Service (IRIS) Online.* Office of Environmental Health and Environmental Assessment, Environmental Criteria and Assessment Office, Cincinnati, OH.

USEPA (U. S. Environmental Protection Agency). 1989. *Exposure Factors Handbook.* EPA/600/8-89/043.

Watanabe, K., Bois, F.Y., and Zeise, L. 1992. Interspecies extrapolation: a re-examination of acute toxicity data. *Risk Anal.* 12, 2, 301-317.

Weil, C.S. 1972. Statistics vs. safety factors and scientific judgment in the evaluation of safety for man. *Toxicol. Appl. Pharmacol.* 21, 454-463.

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, 6, 486-491.