Risk Analysis, Vol. 30, No. 3, 2010 DOI: 10.1111/j.1539-6924.2009.01336.x

Perspective

Conundrums with Uncertainty Factors

Roger Cooke*

The practice of uncertainty factors as applied to noncancer endpoints in the IRIS database harkens back to traditional safety factors. In the era before risk quantification, these were used to build in a "margin of safety." As risk quantification takes hold, the safety factor methods yield to quantitative risk calculations to guarantee safety. Many authors believe that uncertainty factors can be given a probabilistic interpretation as ratios of response rates, and that the reference values computed according to the IRIS methodology can thus be converted to random variables whose distributions can be computed with Monte Carlo methods, based on the distributions of the uncertainty factors. Recent proposals from the National Research Council echo this view. Based on probabilistic arguments, several authors claim that the current practice of uncertainty factors is overprotective. When interpreted probabilistically, uncertainty factors entail very strong assumptions on the underlying response rates. For example, the factor for extrapolating from animal to human is the same whether the dosage is chronic or subchronic. Together with independence assumptions, these assumptions entail that the covariance matrix of the logged response rates is singular. In other words, the accumulated assumptions entail a log-linear dependence between the response rates. This in turn means that any uncertainty analysis based on these assumptions is ill-conditioned; it effectively computes uncertainty conditional on a set of zero probability. The practice of uncertainty factors is due for a thorough review. Two directions are briefly sketched, one based on standard regression models, and one based on nonparametric continuous Bayesian belief nets.

KEY WORDS: Benchmark dose; LOAEL; NOAEL; uncertainty factors

1. INTRODUCTION

The method of uncertainty factors[1] harkens back to the engineering practice of safety factors. If the reference load for an engineered structure is X, then engineers may design the structure to withstand load 3X, using a safety factor of 3 to create a margin of safety. If the structure functions in a corrosive environment, another factor would be multiplied to guarantee safety, and another factor for heat, another factor for vibrations, etc. The choice of values is simply based on "good engineering practice," that is, what has worked in the past. Although safety factors are still common in engineering, they are giving way to probabilistic design in many applications. The reason is that compounding safety factors quickly leads to overdesigning. Compounding safety margins for spaceflight systems may render them too heavy to fly. As our understanding of the system increases, it becomes possible to guarantee the requisite safety by leveraging our scientific understanding of the materials and processes. That of course requires formulating clear probabilistic safety goals, and developing the techniques to demonstrate compliance.

The engineering community has never sought to account for uncertainty by treating safety factors

*Address correspondence to Roger Cooke, Resources for the Future, Washington, DC, and Department of Mathematics, Delft University of Technology, Delft, The Netherlands; cooke@rff.org.

as random variables and assigning them (marginal) distributions; such an approach would not counteract the overdesigning inherent in safety factors. Many authors, including the recent NAS report *Science and Decisions* (pp. 159 ff)[2] have advocated just such a probabilistic approach.

EPA's flagship resource for risk of exposure to hazardous chemicals, the Integrated Risk Information System (IRIS), uses uncertainty factors as safety factors. For noncancer risks, the goal is to derive a reference dose (RfD) or reference concentration (RfC). These "reference values" have been based on a "no observed adverse effect level" (NOAEL) or lowest observable adverse effect level (LOAEL). The reference dose methodology is applied in several programs within EPA, including acute reference exposure (ARE), acute exposure guideline level (AEGL), Office of Pesticide Programs procedures, Office of Water (OW) Health Advisories (HA), and the Agency for Toxic Substances and Disease Registry (ATSDR) Minimal Risk Levels (MRL). These programs have developed different approaches to setting acute, short-term, and longer-term reference values. Efforts are underway to incorporate these different values within the integrated risk information system (IRIS) database.

As pointed out by the WHO,[3] the NOAEL is not a biological threshold, and may either under- or overestimate a true biological threshold. More recently, the benchmark dose (BMD) or lower confidence limit of the benchmark dose (BMDL) have been used as a point of departure for deriving reference values. The effective dose at which r% of the exposed subjects respond, ED_r is another indicator sometimes used as a point of departure. The regulatory history is sketched in Reference 4. According to Reference 5, the current definition of the RfD is

> ***RfD:*** *an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily oral exposure to the human population (including sensitive subgroups) that is likely to be without appreciable risk of deleterious effects during a lifetime. It can be derived from a NOAEL, LOAEL, or BMD, with UFs generally applied to reflect limitations of the data used. Generally used in EPA's noncancer health assessments.*

The same document proposes a revised definition.

> ***Reference Value:*** *an estimate of an exposure, designated by duration and route, to the human population (including susceptible subgroups) that is likely to be without an appreciable risk of adverse health effects over a lifetime. It is derived from a BMDL, a NOAEL, a LOAEL, or another suitable point of departure, with uncertainty/variability factors applied to reflect limitations of the data used.*

The reference value is obtained by dividing a point of departure by uncertainty factors[1] (a typical default value is 10) to account for uncertainty from various sources. These sources, with notion used in IRIS are

1. extrapolation from LOAEL to NOAEL (UF_L),
2. data sparseness (UF_D: extrapolation from poor to rich data contexts)
3. interspecies extrapolation (UF_A: extrapolation from animal to humans)
4. sensitive human subpopulations (UF_H), and
5. subchronic to chronic dosage (UF_S: extrapolation from subchronic to chronic dosage).

The definition makes use of probabilistic and quantitative language as "likely" and "appreciable risk." Although this suggests a quantitative interpretation, none has been proposed to date. To indicate how indeterminate words like "appreciable risk" and "likely" can be, consider the "Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties"[7] issued by the Intergovernmental Panel on Climate Change, reproduced as Table I.

Does exposure at the reference value entail a 66% probability of being without appreciable risk? Reference value exposure to 100 substances acting independently would then entail a 10^{-18} probability of avoiding "appreciable risk." Being "virtually certain" for each of 100 independent chemicals would still entail only a 37% chance of avoiding appreciable risk. Such examples underscore the inevitability of quantifying risk.

[1] The IRIS database[6] currently defines "uncertainty/variability factor (UFs)" as: "One of several, generally 10-fold, default factors used in operationally deriving the RfD and RfC from experimental data. The factors are intended to account for (1) variation in susceptibility among the members of the human population (i.e., interindividual or intraspecies variability); (2) uncertainty in extrapolating animal data to humans (i.e., interspecies uncertainty); (3) uncertainty in extrapolating from data obtained in a study with less-than-lifetime exposure (i.e., extrapolating from subchronic to chronic exposure); (4) uncertainty in extrapolating from a LOAEL rather than from a NOAEL; and (5) uncertainty associated with extrapolation when the database is incomplete."

Table I. Likelihood Table from IPCC Fourth Assessment Report on Addressing Uncertainties

Terminology	Likelihood of the Occurrence/Outcome
Virtually certain	>99% probability of occurrence
Very likely	>90% probability
Likely	>66% probability
About as likely as not	33–66% probability
Unlikely	<33% probability
Very unlikely	<10% probability
Exceptionally unlikely	<1% probability

There has been much work at giving a probabilistic interpretation of the UFs. Many authors[8–18] envisage what might be called a *Random Chemical* approach. Several authors adduce properties based on log-normal distributions. Insightful studies[19,20] suggest that uncertainty factors are independent log-normal variables. Combining uncertainty factors involves multiplying the median values, and combining the "error factors"[2] according to the formula $K_{S\times A} = \exp(1.6449 \times \sqrt{(\sigma_S^2 + \sigma_A^2)})$, where σ_S^2 and σ_A^2 are the variances of $\ln(UF_S)$ and $\ln(UF_A)$, respectively. Thus, $UF_S \times UF_A$ is a log-normal variable with median$(UF_S \times UF_A)$ = median$(UF_S) \times$ median(UF_A), and 95 percentile given by median $(UF_S \times UF_A) \times K_{S\times A}$. If UF_S and UF_A each have an error factor of 10, then the error factor of $UF_S \times UF_A$ is not 100 but 25.95. Several authors suggest that multiplying uncertainty factors might overprotect.[3]

[2] Factors that multiply and divide the median of a log-normal distribution to obtain the 5 and 95 percentiles are termed error factors in the technical risk literature.[23]

[3] "One of the crucial assumptions affecting how uncertainty factors (UFs) are operationally implemented is that they are independent of each other. This assumption of independence has led to the conclusion that the collective uncertainty is the product of all the individual uncertainty factors" (Ref. 14, p. 44). "Because not all true differences are expected to be at their extremes simultaneously, reducing an observed exposure value by a product of default uncertainty factors may lead to undue conservatism" (Ref. 19, p. 190). "Sound combination of extrapolation factors still is an unresolved task in risk assessment. In case of simple multiplication to an overall extrapolation factor this may lead to overly conservative human limit values, if all maximal single factors are used simultaneously. In addition, multiplication of single extrapolation factors would only be correct for independent parameters" (Ref. 3, p. 97). "Because of the application of various uncertainty factors that are multiplied with each other, the standard method for deriving acceptable human limit values is generally considered to be conservative. Indeed, when each individual uncertainty factor by itself is regarded to reflect a worst case situation, their product, i.e. the overall uncertainty factor, will tend to be overly conservative" (Ref. 12, p. 787).

Recent proposals from the National Research Council[2] are based on the random chemical approach, and inherit its problems. For data on response ratios see References 21 and 22.

Section 2 identifies assumptions that the current practice imposes on a probabilistic interpretation of uncertainty factors. The random chemical model is formulated in Section 3, and Section 4 shows that the logged covariance matrix of response rates is singular. Suggestions for moving beyond uncertainty factors are explored in Section 5.

2. ASSUMPTIONS UNDERLYING UNCERTAINTY FACTORS

A probabilistic model for uncertainty factors would consist of a sample space and a set of random variables together with assumptions on their joint distribution such that relations between these random variables reflect the operations that practitioners perform with uncertainty factors. Under the random chemical model, uncertainty factors are random variables reflecting response ratios of randomly sampled toxic substances. Distributions of these ratios are used to draw inferences about new or untested chemicals. The operations performed with uncertainty factors entail assumptions on their joint distribution. The prevailing assumptions are presented informally in this section, precise versions are used in Section 3.

We consider a point of departure (POD), which may be either a BMD, a LOAEL, a NOAEL, or an effective dose for response of r% of the population (ED_r). The relevant assumptions are ("UFs" denotes "uncertainty factors," not to be confused with "UF_S"):

1. The UFs are independent random variables.
2. The extrapolation expressed by a UF is independent of other extrapolations. That is, the UF for extrapolating from poor to rich data contexts does not depend on whether the extrapolation concerns chronic or subchronic dose. The extrapolation from subchronic to chronic dosage does not depend on whether this is applied to humans or animals, or to the poor/rich data contexts,[4] etc.

[4] Subchronic exposure is defined as:[6] "Repeated exposure by the oral, dermal, or inhalation route for more than 30 days, up to approximately 10% of the life span in humans (more than 30 days up to approximately 90 days in typically used laboratory animal species)." Chronic exposure is defined as: "Repeated exposure

3. The conditional distribution of a human reference value, given an observed (unextrapolated) POD, is obtained by dividing the observed POD by the product of the UFs corresponding to the required extrapolations.

3. RANDOM CHEMICAL MODEL

We illustrate the random chemical model for the extrapolations poor-to-rich, subchronic-to-chronic, and animal-to-human. Other extrapolations could serve equally well to illustrate the issues. Suppose we randomly sample toxic chemical t from a representative set of toxic substances. For each t, we imagine that we have values for the POD for animals at each dosage regime (chronic, subchronic; C, S), for each data context (poor, rich; P, R), and for two species (animal, human; A, H). Hence if $HC(t)$ is the reference value for t, that is, the chronic dose that is likely to be without appreciable lifetime risk, we may always write an equation of the form:

$$HC(t) = ASP(t)\frac{ASR(t)}{ASP(t)}\frac{ACR(t)}{ASR(t)}\frac{HC(t)}{ACR(t)}. \quad (1)$$

If we consider t as a randomly drawn chemical, which, following convention, we denote with upper case T, we can write Equation (1) as an equation of random variables:

$$HC(T) = ASP(T)\frac{ASR(T)}{ASP(T)}\frac{ACR(T)}{ASR(T)}\frac{HC(T)}{ACR(T)}. \quad (2)$$

by the oral, dermal, or inhalation route for more than approximately 10% of the life span in humans (more than approximately 90 days to 2 years in typically used laboratory animal species)." U_A is defined as follows: "Use an additional 10-fold factor when extrapolating from valid results from long term studies on experimental animals when results of studies of human exposure are not available or are inadequate" and U_C: "Use an additional 10-fold factor when extrapolating from less than chronic results on experimental animals when there are no useful long-term human data." These definitions would not tell us how to extrapolate from human subchronic studies, nor from animal subchronic to animal chronic. Such extrapolations are surely less prevalent, though not excluded, as the definitions of subchronic and chronic dosages make clear. The absence of separate uncertainty factors, for example, for subchronic-to-chronic for animals and for humans, suggests that Equation (1) is implicitly assumed. Indeed, Swartout *et al.* (Ref. 11, p. 275) write: "Within the current RfD methodology, UF_C [here, UF_S] does not consider differences among species, endpoints, or severity of effects, the same factor is applied in all cases." Because of the rarity of relevant human data, the same authors suggest the use of other endpoints as surrogates in estimating U_A.

If we now write:

$$UF_D = \frac{ASP(T)}{ASR(T)}, \quad UF_S = \frac{ASR(T)}{ACR(T)}, \quad UF_A = \frac{ACR(t)}{HC(t)}, \quad (3)$$

then it appears that have interpreted UFs as random variables based on the random chemical model:

$$HC(T) = \frac{ASP(T)}{UF_D \times UF_S \times UF_A}. \quad (4)$$

Equation (4) reflects an assumption noted by many authors, namely, that the UFs as random variables must be independent. However, Equation (4) involves more assumptions that have not received ample attention. Suppose we observe for a given toxic substance t that $ASP(t) = 30$ [units]. We use $30/(UF_D \times UF_S \times UF_A)$ as the *conditional* distribution of $HC(T)$ given $ASP(T) = 30$. This must hold for any observed value, thus $ASP(T)$ must be independent of $UF_D \times UF_S \times UF_A$.[5] Barring pathological situations, this means that $ASP(T)$ is independent of each of these UFs.

Note that these conditionalization assumptions always apply from "more easily measured to more difficult to measure." They do not apply in the other direction. Letting "$\perp$" denote independence, if $X \perp X/Y$ and also $Y \perp X/Y$, then it is easy to show that X/Y must be constant.[6]

4. IMPLICATIONS OF ASSUMPTIONS

These independence assumptions are quite strong and have consequences. Consider the statement that:

$$ASP(T) \perp UF_D. \quad (5)$$

Note that $ASR(T) = ASP(T)/UF_D$. Taking logs of both sides:

$$Ln(ASR(T) = Ln(ASP(T)) - Ln(UF_D). \quad (6)$$

[5] Rehearsing elementary probability, suppose we wish to express the uncertainty in random variable X, given that another random variable Y takes the value y. One way of modeling this is to write X as a function $g(Y, Z)$ of Y and some other random variable Z. *If Z is independent of Y*, then the conditional distribution $(X \mid Y = y)$ of X given $Y = y$ is the distribution of the random variable $g(y, Z)$. If Y and Z are not independent, then we must use the conditional distribution $(Z \mid Y = y)$ when computing the distribution of g. In this case $X = HC(T)$, $Y = ASP(T)$, $Z = UF_A(T) \times UF_S(T) \times UF_D$, and $g(Y, Z) = Y/Z$. To justify the type of conditionalization for which the UFs are intended, we must have $ASP(T)$ independent of $\{UF_A(T), UF_S(T), UF_D(T)\}$.

[6] Take logs, write out the covariances, and infer that $\sigma^2_{\ln(X)-\ln(Y)} = 0$.

$ASP(T) \perp UF_{\mathrm{D}}$ entails also $Ln(ASP(T) \perp Ln(UF_{\mathrm{D}})$; and that:

$$VAR(Ln(ASR(T)) = VAR(Ln(ASP(T))) + VAR(Ln(UF_D)). \quad (7)$$

This says that the variance of logged animal PODs based on subchronic dosage and rich data sets is strictly *greater than* the variance of logged animal PODs based on a subchronic dosage with poor data contexts; the same holds for chronic dosage. This seems counterintuitive. Similarly, $ASP(T) \perp UF_{\mathrm{S}}$ implies that $VAR[Ln(ACP(T))] > VAR[Ln(ASP(T))]$, and the same holds for rich data contexts. This means that the variance of the logged animal chronic values is strictly greater than the variance of the logged animal subchronic value, which is at odds with statements in the Cancer Guidelines.[7] This underscores the importance of identifying underlying assumptions.

We now demonstrate the singularity of the covariance matrix of logged terms in the UFs. It suffices to consider only the extrapolation from poor to rich data contexts, and from subchronic to chronic dosages, all for animal PODs. We could equally well have chosen UF_{L} and UF_{A}; the argument would be identical. To lighten the notation in this section, let SP denote the animal value of the POD under subchronic dosage in a poor data context, CR the animal POD under chronic dosage in a rich data context, and let SR and CP be defined similarly. Each of these is considered as a random variable whose values are determined by randomly sampling a toxic substance from a representative set.

The basic equations are:

$$CR = SP \times (SR/SP) \times (CR/SR) = SP/(UF_{\mathrm{D}} \times UD_{\mathrm{S}}) \quad (8)$$

$$CR = CP \times (CR/CP) = CP/UF_{\mathrm{D}}. \quad (9)$$

The same UF is applied for extrapolating from poor to rich data contexts, regardless whether the dosage is chronic or subchronic. This means that UF_{D} has a distribution that can be measured in two ways:

$$UF_{\mathrm{D}} \sim \frac{SP}{SR} \quad (10)$$

$$UF_{\mathrm{D}} \sim \frac{CP}{CR}. \quad (11)$$

Similarly, the UF for extrapolating from subchronic to chronic is used for both rich and poor data contexts, hence:

$$UF_{\mathrm{S}} \sim \frac{SP}{CP} \quad (12)$$

$$UF_{\mathrm{S}} \sim \frac{SR}{CR}. \quad (13)$$

We use the following notation:

$$A = VAR(\ln(SP)),$$

$$B = VAR(\ln(SR)),$$

$$C = VAR(\ln(CP)),$$

$$D = VAR(\ln(CR)).$$

If two variables are independent, then also the logs of these variables are independent. If X is independent of Y, then X is also independent of $1/Y$. The following independence statements are assumed by the UF methodology, under a probabilistic interpretation:

I.1 $SP \perp SP/SR$ (enable conditionalization in Equation (8))
I.2 $SP \perp SR/CR$ (enable conditionalization in Equation (8))
I.3 $SP/SR \perp SR/CR$ ($UF_{\mathrm{D}} \perp UF_{\mathrm{S}}$)
I.4 $SP \perp CP/CR$ (enable conditionalization in Equation (8) with Equation (11))
I.5 $CP \perp CP/CR$ (enable conditionalization in Equation (9))
I.6 $CP \perp SP/SR$ (enable conditionalization in Equation (9) with Equation (10))
I.7 $SR/CR \perp CP/CR$ ($UF_{\mathrm{S}} \perp UF_{\mathrm{D}}$)

If two variables are independent, then their covariance is zero. Let $Cov(X,Y)$ denote the covariance of X and Y. Taking logs of I.1–I.7, and setting covariances of independent variables equal to zero, using the linearity of covariance, we find:

[7] "Uncertainty usually increases as the duration becomes shorter relative to the averaging duration or the intermittent doses become more intense than the averaged dose. Moreover, doses during any specific susceptible or refractory period would not be equivalent to doses at other times."[3,4,24]

Table II. Log Covariance Matrix

	Ln(SP)	Ln(SR)	Ln(CP)	Ln(CR)
Ln(SP)	A	A	A	A
Ln(SR)	A	B	A	B
Ln(CP)	A	A	C	C
Ln(CR)	A	B	C	$B + C - A$

I.1 $\Rightarrow$ $A = Cov(\ln(SP), \ln(SR))$,
I.2 and I.1 $\Rightarrow$ $A = Cov(\ln(SP), \ln(CR))$,
I.3 and I.1 and I.2 $\Rightarrow$ $B = Cov(\ln(SR), \ln(CR))$,
I.4 and I.1 and I.2 $\Rightarrow$ $A = Cov(\ln(SP), \ln(CP))$,
I.5 $\Rightarrow$ $C = Cov(\ln(CP), \ln(CR))$,
I.6 and I.4 $\Rightarrow$ $A = Cov(\ln(CP), \ln(SR))$,
I.7 and I.6 and I.5 and I.3 $\Rightarrow$ $D = B + C - A$.

We bring these relations together in the following covariance matrix.

Observe that the second plus third rows, minus the first row equals the fourth row. The determinant of this matrix is zero, meaning that there is a linear relationship between the variables. Using $VAR(X + Y) = VAR(X) + VAR(Y) + 2Cov(X,Y)$, with the values in Table II it follows that:

$$VAR(\ln(CR) + \ln(SP) - \ln(SR) - \ln(CP)) = 0. \tag{14}$$

In other words:

$$CR = SR \times CP/SP. \tag{15}$$

Rearranging, we may write this as $CR/CP = SR/SP$. This says that the two random variables in Equations (10) and (11) must actually be the same variable. This means that if we actually knew the values $SP(t)$, $CP(t)$, and $SR(t)$, we could *compute* $CR(t)$. Similar singularities arise if we consider other pairs of UFs.

5. HOW FORWARD?

The first step forward is to realize that the current system will not admit a probabilistic interpretation and that fundamental changes are required if we wish to account for uncertainty in reference values. Ultimately, for important chemicals, we should like to combine animal data, *in vitro* human data, and epidemiological data from natural experiments to derive dose-response relations with uncertainty quantification. However, there will always be a need to evaluate new potentially toxic chemicals based on their similarity to other chemicals and meager experimental data. We should like a simple method that:

1. Yields predictions of toxicological indicators with uncertainty via a valid probabilistic mechanism.
2. Is based on a random chemical model, regarding a new chemical as a random sample from a reference distribution of chemicals.
3. Is nondisruptive.

This last desideratum is very important, and has received insufficient attention. Regulatory bodies cannot turn on a dime, but must contend with a legacy of accepted practice. The foregoing suggests that a probabilistically valid inference system will be quite different from the current system. Nonetheless, to meld with current practice, it must initialize on the current system, and allow this system to evolve in a measured fashion.

We can use the IRIS database to obtain a reference set of substances, perhaps supplemented with structured expert judgment. These reference values serve to define a population of toxic substances from which a new substance is regarded as a random sample. The idea is simply to discard the *UF* methodology, but to retain the *results* of that method as a reference distribution to initialize the new system. The reference distribution can further evolve under the new regime, but the changes in toxicity indicator values will be nondisruptive. There are at least two ways of leveraging such a reference set to draw inferences on new substances: standard log-linear regression and nonparametric Bayesian belief nets.

5.1. Standard (Log) Linear Regression

To render this discussion more intuitive, we focus on animal (A) and human (H) responses at chronic (C) and subchronic (S) dosages. Suppose we are interested in predicting $\ln(HC)(t^*)$ for a new toxic substance t^*, and we can estimate $Cov(\ln(HC), \ln(AC))$, and $\sigma^2(\ln(AC))$ from a large data set of toxic substances. Translating all variables to have mean zero, the linear least squares predictor (llsp) of $\ln(HC)$ based on $\ln(AC)$ would be:

$$llsp(\ln(HC)(t^*)|\ln(AC)(t^*)) = Cov(\ln(HC), \ln(AC)) \times (1/\sigma^2(\ln(AC))) \times \ln(AC)(t^*).$$

If $\ln(AC)(t^*)$ is known, this is a predictor of $\ln(HC)(t^*)$, not a random variable. We might try to think of $\sigma^2(\ln(AC))/Cov(\ln(HC), \ln(AC))$ as an uncertainty factor, but these would not behave as uncertainty factors in the IRIS methodology. To give one illustration, suppose we observed $\ln(AS)(t^*)$ in addition to $\ln(AC)(t^*)$. The IRIS method would not use the additional information regarding $AS(t^*)$, but would simply use $AC(t^*)/U_A$. However, following the theory of linear models, we should have:

$$\begin{aligned} &llsp((\ln(HC)(t^*)|\ln(AC)(t^*),\\ \ln(AS)(t^*)) &= llsp(\ln(HC)(t^*)|\ln(AC)(t^*))\\ &\quad + llsp[\ln(HC)(t^*)|\ln(AC)(t^*)\\ &\quad - llsp(\ln(AC)(t^*)|\ln(AS)(t^*))]. \end{aligned}$$

In other words, we add the llsp of $\ln(HC)(t^*)$ based on $\ln(AC)(t^*)$ to the llsp of $\ln(HC)(t^*)$ based on the residual $\ln(AC)(t^*) - llsp(\ln(AC)(t^*) \mid \ln(AS)(t^*))$.

We can extract the variance of the llsp; for arbitrary random vectors X, Y, we have (the variance and covariance terms now denote matrices):

$$\sigma^2(llsp(Y \mid X) = Cov(Y, X)(\sigma^2(X))^{-1}Cov(X, Y).$$

We also obtain the variance of the residual as:

$$\sigma^2(Y - llsp(Y \mid X)) = \sigma^2(Y) - \sigma^2(llsp(Y \mid X).$$

Note that the last two variances do not depend on the value of the conditioning variable X. In general, the variance of the residual is not the conditional variance of Y given X, and we do not get the conditional distribution of Y given X. In many cases, we may actually be interested in the distribution of human values given observed animal values, especially if these observed values are in the tails of their respective distributions. Such considerations drive us in the direction of Bayesian belief nets.

5.2. Nonparametric Continuous Bayesian Belief Nets

Nonparametric continuous Bayesian belief nets (NPCBBNs) build a joint density.[25] A full description is inappropriate here; suffice to say that NPCBBNs are based on empirical marginal distributions and empirical rank correlations; they build a joint density by modeling the copula that realizes these empirical rank correlations. In principle any copula can be used, but in practice the joint normal copula is preferred, as it enables rapid conditionalization. Roughly, this means that the rank dependence structure is modeled as that of a joint normal distribution. This hypothesis can be tested, based on the sampling distribution of the determinant of the normal rank correlation matrix. The marginal distributions and rank correlations are not modeled, but taken directly from data; hence they do not form part of the hypothesis being tested. If the normal copula hypothesis is not rejected, then we can proceed to conditionalize any variable on any set of values of other variables.

5.2.1. Using NPCBBNs for Inference on Toxicity

A BBN is a mechanism for encoding inferences. Drawing an inference from data is performed by conditionalizing a BBN on the data; the distribution captured in the BBN is updated with current data. BBNs are simply a perspicuous way of visualizing complex inference patterns. Continuous nonparametric BBNs enable this inference based on an initial data set. With the normal copula, very large problems with very complex inference structures can be handled analytically, so that the inferences are effectively instantaneous. The marginal distributions and rank correlation structure is read from the data; the only assumption is that these rank correlations are compatible with the normal copula. This is a much weaker assumption than that underlying standard regression models, namely, that error terms are independent and normally distributed.

To compare the BBN approach to standard regression modeling, suppose that we are interested in some probabilistic response in animals (A) and humans (H) at chronic (C) and subchronic (S) dosage. As before, this might be an RfD, RfC, BMD, or an ED_{10} or indeed any other indicator. It could be an ED_{10} for subchronic dosage of animals, ED_{50} for chronic dosage for animals, an RfD for chronic human dosage, and a BMD for subchronic human dosage. All that matters is that we have a list of, say, 100 toxic substances filled in as illustrated in Table III. The units in each column must be the same, but need not be the same across columns.

When the inference engine is "seeded" in this way, we can apply it to some new chemical that we regard is a random sample from the population of chemicals from which the chemicals inTable III are randomly drawn. We guarantee thereby that the inferences for new chemicals will reflect the relations between the probabilistic response variables captured in Table III.

Table III. Illustrative Input for BBN

	Probabilistic Response: Animal, Subchronic	*Probabilistic Response*: Animal Chronic	*Probabilistic Response*: Human Subchronic	*Probabilistic Response* Human Chronic
Toxic Chem. 1	20 [AS units]	15 [AC units]	4[HS units]	2[HC units]
Toxic Chem. 2	45[AS units]	30[AC units]	30[HS units]	26[HC units]
Toxic Chem. 3	30[AS units]	35[AC units]	20[HS units]	14[HC units]
...	...	...	...	...

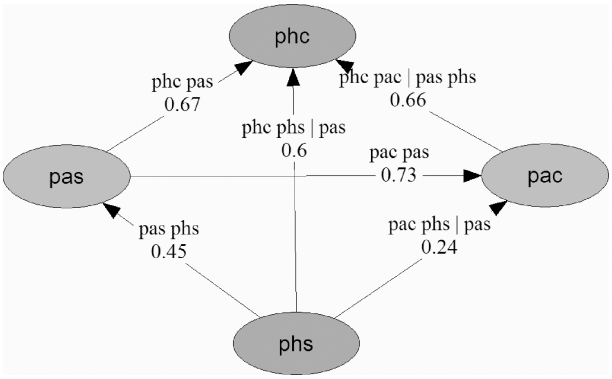

Fig. 1. Bayesian belief net for *P*, Animal, Subchronic (*AS*); *P*, Animal, Chronic (*AC*)); *P*, Human, Subchronic, ((*HS*)); and *P*, Human, Chronic (*HC*)), distributions over set of (fictitious) toxic substances; (conditional) rank correlations as inferred from Table II are shown.

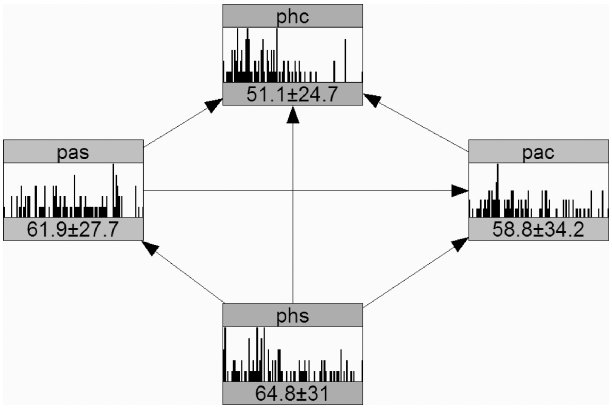

Fig. 2. Same information as Fig. 1, but with means and standard deviations.

Fig. 1 shows the result of reading these (fictitious) data into an NPCBBN. The conditional rank correlations are inferred from the relations in Table III.[25] Fig. 2 shows the marginal histograms, with means and standard deviations. Figs. 3–5 show how this BBN can be conditionalized to draw inferences for new toxic substances, regarded as members of the same population as in Table III.

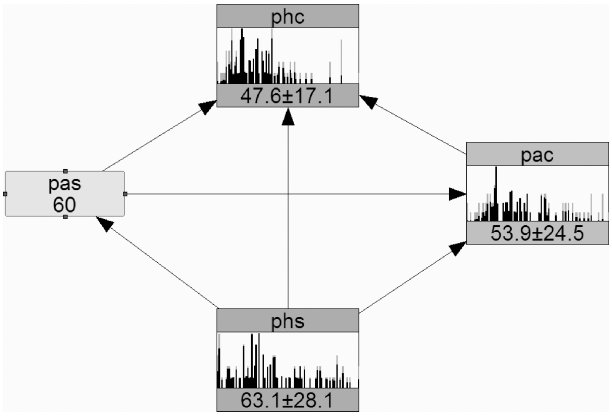

Fig. 3. Distribution in Fig. 2, conditionalized on observing $AS = 60$.

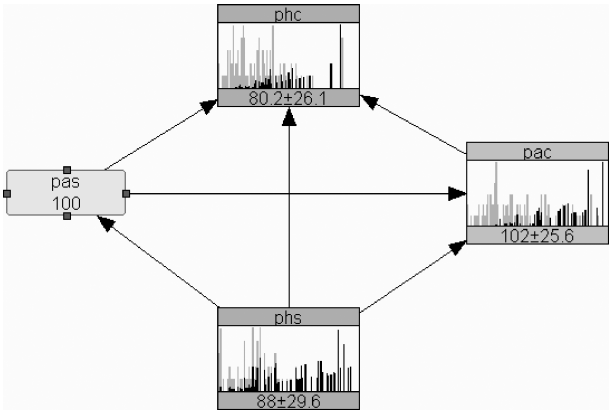

Fig. 4. Distribution in Fig. 2, conditionalized on observing $AS = 100$.

Confronted with a new substance, our inferences from test results may be based on the assumption that this chemical is randomly drawn from the toxicity distribution captured in the BBN. Suppose for a new substance, we observe that $AS = 60$, our mean value for *HC* shifts from 51.1 (Fig. 2) to 47.6 (Fig. 3). The conditional distribution of *HC* giving this information is also available and is shown as a histogram; the original distribution is shown in gray. The standard deviation of *HC* has shifted from 24.7 to 17.1.

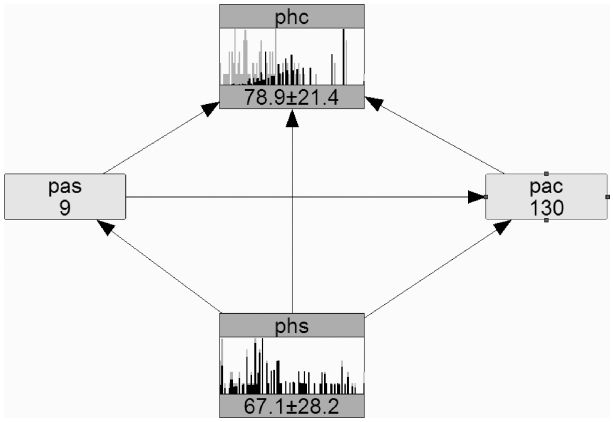

Fig. 5. Distribution in Fig. 2, conditionalized on $\ln(AS) = 9, AC = 130$.

Suppose we conditionalize on $AS = 100$, a very high value (Fig. 4), the mean and standard deviation of *HC* are 80.2 and 26.1, respectively. In contrast to the log-linear regression model, we obtain the full conditional distributions, and we note that the conditional variances are *not* constant. Conditionalization of *AS* on a typical value (60) lowers the conditional uncertainty of *HC* relative to the unconditional uncertainty. Whereas conditionalizing *AS* on a large value (100) increases the conditional uncertainty. In Fig. 5, we conditionalize on a very high value of *AC*, 130. In combination with the very low value of *AS*, 9, this constitutes a very unlikely combination. Nonetheless, because we have built a multivariate density function, we can conditionalize on this unlikely situation and find the *HC* has mean and standard deviation of 78.9 and 21.4, respectively. *AC*'s influence on *HC* is stronger than that of *AS*. Indeed, the correlation on the arrow between *AC* and *HC* in Fig. 1 is the conditional correlation of *HC* and *AC* given *AS*. The unconditional correlation between *AC* and *HS* is 0.825.

6. CONCLUSION

UFs were introduced to create a margin of safety. Attempts to give them a probabilistic interpretation in terms of response ratios of random chemicals encounter insuperable obstacles. Under prevailing assumptions, the logged response ratios have a singular covariance matrix. Any Monte Carlo analysis based on the probabilistic interpretation of UFs is therefore ill-conditioned. Claims that UFs overprotect, as well as recent NRC proposals for probabilistic interpretations, are equally ill-conditioned. A thorough review would not be inappropriate. The effort to quantify uncertainty in dose response should skirt UFs altogether and focus on dose-response modeling. However, the need for rapid and simple inference based on a random chemical model will persist. A new rapid inference system should (1) be as easy to use as the current system; (2) should derive toxicological indicator values, with uncertainty, in a probabilistic valid fashion; and (iii) should initialize on current practice. Two lines of attack have been sketched here, one based on standard regression modeling, and one based on Bayesian belief nets. Undoubtedly, more good ideas will emerge, once the inevitability of thorough-going reform is recognized.

REFERENCES

1. Lehman AJ, Fitzhugh OG. 100-fold margin of safety. Association of Food and Drug Official on USQ Bulletin, 1954; 18:33–35.
2. National Research Council of the National Academies. Science and Decisions: Advancing Risk Assessment. Washington DC, 2009.
3. Kalberlah F, Schneider K, Schuhmacher-Wolz U. Uncertainty in toxicological risk assessment for non-carcinogenic health effects. Regulatory Toxicology and Pharmacology, 2003; 37:92–104.
4. Dourson ML, Stara JF. Regulatory history and experimental support of uncertainty (safety) factors. Regulatory Toxicology and Pharmacology, 1983; 3:224–238.
5. US Environmental Protection Agency. A review of the reference dose and reference concentration processes. Available at: EPA/630/p-02/002f.final report, Accessed in December 2002.
6. Integrated Risk Information System (IRIS). Available at: http://www.epa.gov/ncea/iris/help_gloss.htm#u, Accessed on November 7, 2009.
7. Intergovernmental Panel on Climate Change WMO UNEP. 2005 Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties. Available at: http://ipccwg1.ucar.edu/wg1/Report/AR4_UncertaintyGuidanceNote.pdf, Accessed on November 11, 2009.
8. Abdel-Rahman M, Kadry AM. Studies on the use of uncertainty factors in deriving RfDs. Human Ecological Risk Assessment, 1995; 1(5):614–624.
9. Vermeire T, Stevenson H, Pieters MN, et al. Assessment factors for human health risk assessment: A discussion paper. Critical Reviews in Toxiocology, 1999; 29(5):439–490.
10. Baird SJS, Cohen JT, Graham JD, Shlyakhter AI, Evans JS. Noncancer risk assessment: A probabilistic alternative to current practice. Human and Ecological Risk Assessment, 1996; 2:79–102.
11. Swartout JC, Price PS, Dourson ML, Carlson-Lynch HL, Keenan RE. A probabilistic framework for the reference dose (probabilistic RfD). Risk Analysis, 1998; 18(3):271–281.
12. Slob W, Pieters MN. A probabilistic approach for deriving acceptable human intake limits and human health risks from toxicological studies: General framework. Risk Analysis, 1998; 18:787–789.
13. Evans JS, Baird SJS. Accounting for missing data in noncancer risk assessment. Human Ecological Risk Assessment, 1998; 4:291–317.

14. Calabrese EJ, Gilbert CE. Lack of total independence of uncertainty factors (Ufs): Implications for the size of the total uncertainty factor. Regulatory, Toxicology and Pharmacology, 1993; 17:44–51.
15. Calabrese EJ, Baldwin LA. A toxicological basis to derive generic interspecies uncertainty factors for application in human and ecological risk assessment. Human and Ecological Risk Assessment, 1995; 1(5):555–564.
16. Hattis D, Baird S, Goble R. A straw man proposal for a quantitative definition of the RfD. Drug and Chemical Toxicology, 2002; 25(4):403–436.
17. Kang S-H, Kodell RL, Chen JJ. Incorporating model uncertainties along with data uncertainties in microbial risk assessment. Regulatory, Toxicology and Pharmacology, 2000; 32:68–72.
18. Pekelis M, Nicolich MJ, Gauthier JS. Probabilistic framework for the estimation of adult and child toxicokinetic intraspecies uncertainty factors. Risk Analysis, 2003; 23(6):1239–1255.
19. Kodell RL, Gaylor DW. Combining uncertainty factors in deriving human exposure levels of noncarcinogenic toxicants. Annals of the New York Academy of Sciences, 1999; 895:188–195.
20. Gaylor DW, Kodell RL. Percentiles of the product of uncertainty factors for establishing probabilistic risk doses. Risk Analysis, 2000; 20:245–250.
21. Rhomberg LR, Wolff SK. Empirical scaling of single oral lethal doses across mammalian species based on a large database. Risk Analysis, 1998; 18(6):741–753.
22. Rhomberg LR, Lewandowski TA. Methods for identifying a default cross-species scaling factor. Prepared for Risk Assessment Forum, U.S. Environmental Protection Agency, 1200 Pennsylvania Avenue, N.W. Washington, DC, April 2, 2004.
23. WASH-1400 Reactor Safety Study. Washington, DC: U.S. Nuclear Regulatory Commission, 1975.
24. Guidelines for Carcinogen Risk Assessment. Washington, DC: Risk Assessment Forum, U.S. Environmental Protection Agency, EPA/630/P P3/001F, March 2005.
25. Hanea AM, Kurowicka D, Cooke RM, Ababei DA. Ordinal data mining with non-parametric continuous Bayesian belief nets. Accepted for publication at Computational Statistics and Data Analysis, 2008.