Risk Assessment—the Mother of All Uncertainties

Disciplinary Perspectives on Uncertainty in Risk Assessment

JOHN C. BAILAR III[a] AND A. JOHN BAILER[b,c]

[a]*Department of Health Studies and Harris School of Public Policy, University of Chicago, 5841 South Maryland Avenue, MC 2007, Chicago, Illinois 60637, USA*

[b]*Department of Mathematics and Statistics, Miami University, Oxford, Ohio 45056, USA*

[c]*National Institute for Occupational Safety and Health, Education and Information Division, Risk Evaluation Branch, 4676 Columbia Parkway, Cincinnati, Ohio 45226, USA*

ABSTRACT: Uncertainty in the detection and evaluation of chemical hazards to health leads to challenges when conducting risk assessments. Some of the uncertainty has to do with data, some with incomplete understanding of processes, and some with the most fundamental ways of viewing the questions. True variability—across space, in time, or among individuals—complicates the search for understanding many important aspects of risk. A few statistical and toxicologic tools are available to assess uncertainty. Three methods of classifying uncertainty are briefly discussed. In addition, our disciplinary background may influence how we view and discuss variability and uncertainty. We rarely know as much as we think we do (and not just in risk assessment). Great uncertainty is likely to remain an important part of risk assessment for some decades to come.

INTRODUCTION

Uncertainty in the detection and estimation of the impact of chemical hazards to health is extensive and difficult to deal with. Much of our presentation here is likely to be familiar; however, we hope to present familiar things in new contexts and to stimulate rethinking about their implications. Some of the uncertainty has to do with inaccurate and incomplete data, some with incomplete understanding of natural processes, and some with the most fundamental ways of viewing the matter. Uncertainty in risk assessment is commonly associated with issues such as the selection of concentration-response models, or extrapolating across exposure conditions, species, or routes of exposure. In contrast, variability is risk assessment usually relates to interindividual variation. Here we interweave general discussions of uncertainty and variability in risk assessment with disciplinary perspectives on these two topics.

Addresses for correspondence:
[a]John C. Bailar III: 773-834-1242 (voice); 773-702-1979 (fax); e-mail: jcbailar@midway.uchicago.edu
[b]A. John Bailer: 513-529-3538 (voice); 513-529-1493 (fax); e-mail: ajbailer@muohio.edu

TABLE 1. Measured responses in an experiment in which two chemicals were administered (hypothetical data)

		Chemical A dose (mg)		
		0	20	40
Chemical B dose (mg)	0	5 units	15	40
	300	10	30	70
	600	40	60	90

To motivate this discussion, consider a simple example of conceptual uncertainty that arises when statisticians, epidemiologists, and toxicologists encounter the same data set. In this example, we explore how synergy is to be defined, including negative synergy, or antagonism.

TABLE 1 shows hypothetical results from a study of the effects of giving two toxic agents, A and B, simultaneously. Nine tests were performed in a 3×3 factorial design. Assume that sample sizes are large enough and observations are accurate enough to take these figures at face value. Is there synergy at the central point in this table (20 mg A, 300 mg B)? The statistician, using a linear model, will note that the baseline response (untreated) is 5 units, that 20 mg of A, alone, increases the response by 10 units, and that 300 mg of agent B increases the response by 5 units, so that if the agents act independently the center point, with 20 mg of A and 300 mg of B, will be increased over background by 5 + 10 = 15 units. However, the data show a larger increase, of 25 units. Thus, the statistician who uses the customary additive approach to the study of such data is likely to conclude that there is positive synergy between 20 mg of A and 300 mg of B. The epidemiologist may be more at ease with multiplicative models (e.g., relative risk or odds ratio models), and will see that 20 mg of A increases the response by a factor of three over background, 300 mg of B increases risk by a factor of two, so that the combination of independent effects would result a sixfold increase, and this is just what is observed. That is, the agents act independently on a multiplicative scale. Although these two views differ, both focus on the combination of response rates, and both depend on some assumption about the additivity of responses. The natural view of the toxicologist is quite different, because it focuses on additivity of doses rather than responses. If 40 mg of A alone produces a 40 unit response rate, and the same is true for 600 mg of B, one might expect a mixture of $x\%$ of A and $(100 - x)\%$ of B to have the same outcome. However, here half a dose of A (20 mg) plus half a dose of B (300 mg) produces only a 30 unit response, not 40 units, so that the agents do not just fail to be synergistic, they are in fact antagonistic. (As an aside, the three other non-central combination treatments might suggest synergy to the toxicologist, which further complicates this informal assessment.) TABLE 2 summarizes these results. Is there synergy? That depends on one's view of *synergy*, which may be related to intended uses of the concept, or may be simply a matter of custom within relevant disciplines. These differences are conceptual and not a matter of data, biologic process, or scientific understanding.

TABLE 2. Summary of the results of the disciplinary perspectives on synergy

Model type (scientist)	Joint action?
additive outcome (statistician)	synergy
multiplicative outcome (epidemiologist)	independence
additive dose (toxicologist/ pharmacologist)	antagonism

Although these three views are mathematically incompatible except under extreme conditions of little practical interest, each makes a great deal of sense in its own context, and the world would be poorer if we had to give up any of them. This example, used to illustrate these perspectives, is clearly contrived and selectively viewed to illustrate a point. As some of our toxicology colleagues have noted, the notion of synergy defined by dose additivity would not be immediately applied in a situation where the dose-response pattern for each chemical given separately was non-linear as we presented in our example. In reality, a toxicologist would consider the expected mechanism of action. For example, two chemicals competing for the same receptor site leads to predictions of antagonism, whereas independent sites of action would lead to dose additivity. Finally, we are not advocating the compartmentalizing of an individual approach by discipline. Our drawings of *statistician*, *epidemiologist,* and *toxicologist* are caricatures rather than portraits. We hope that these exaggerated disciplinary definitions will suggest how different training and experience may result in different evaluations of the same data pattern. This has clear implications for risk assessment, an interdisciplinary exercise of serious societal impact.

In the same manner, there may be substantial differences in the ways that statisticians and others view broader concepts of uncertainty. In fact, any discussion of risk assessment must examine bias, or nonrandom uncertainty. In many biological problems, questions of bias typically dominate the questions of randomness or variability that make up most of statistics. These questions of bias lead to the awkward, but common, result that two or more risk assessments of the same chemical hazard may result in wildly different conclusions. It is not uncommon to see risk estimates that differ by three orders of magnitude or more (that is, more than 10^3), and in the case of dioxin the variation among agencies of the U.S. federal government was at one time 14 orders of magnitude, although the subset used in regulation showed much less variation.[1] That is real uncertainty.

Another example is provided by the analysis of four separate risk assessments of the chemical Tris.[2] The risk assessments were for the use of Tris as a flame retardant in children's sleepwear. These risk assessments were made at about the same time (over a period of about three weeks) by investigators who knew each other and were in frequent contact by telephone and mail. They had access to the same databases and used similar procedures, and yet still reached the estimates shown in TABLE 3. Despite the close communication and the great overlap in both data and technical understanding, these estimates differ by almost three orders of magnitude. The primary differences were in the ways that human exposure were estimated and the ways that exposure was used to assess risk. Four estimates, if they were independently derived

TABLE 3. Added lifetime risk of kidney cancer per million children using Tris-treated sleepware

Author	Lifetime risk
Hooper and Ames	17,000
Bayard	180 (combined)
	300 (male)
	60 (female)
Schneiderman	52
Harris	≥ 7

and if each had a distribution with the correct median, would have one chance in eight of all falling on the same side of the true value. Of course, these four estimates used much of the same data and so are not independent. Thus, there is a large chance that even the range here might not include the true risk. Fortunately, in this case, the exact value did not matter. The Commissioners of the Consumer Product Safety Commission decided that even the lowest estimate of risk from Tris-treated sleepwear was too high, and they acted promptly to ban Tris from the marketplace. As an aside, the manufacturers recognized the problem with equal clarity, and moved even faster, so that Tris was off the market even before the Commissioners could ban it.

Nonrandom uncertainty utterly swamps any contribution from randomness in risk assessment. In the case of a new chemical, for instance, we may need to extrapolate from animal studies to human outcomes, from high to low doses, from one route of administration to another, and from lifetime exposure to information on once-only exposures. In the face of such uncertainty, it hardly matters whether 3, or 8, or 20 animals from a total of 60 have some particular outcome. This is important because a false sense of security can be a result of a refusal to acknowledge (and attempt to quantify) the uncertainty in risk predictions. We have all heard of the call for one-handed risk assessors (to avoid "on one hand this, on the other hand that"), but a false sense of security can have major harmful effects on both governmental regulation and public acceptance of regulations. Since great uncertainty is always with us in this field, it is crucial that we learn how to conduct risk assessment in the face of it.

DISCIPLINARY PERSPECTIVES ON UNCERTAINTY

During the recent meetings of a committee that examined the foundations of risk assessment (primarily of chemical hazards),[3] with special reference to regulation by EPA but with much broader implications, the eminent scientists serving on the committee viewed uncertainty from a variety of different perspectives. The statistician's concepts of variance and bias, with some allowance for averaging over populations that reflect some true variability from one subject to another, was by no means a consensus. These different views are instructive, and worth understanding for effective communication across the different disciplines reflected in the risk assessment community. TABLE 4 highlights our perspective regarding how uncertainty in risk assess-

ment is often considered and classified by statisticians, risk assessors, and toxicologists.

In general terms, disciplinary perspectives are influenced by background and training. (An implication of this observation is that our comments in this paper are *biased* by our own disciplinary training as statisticians. Note also that our TABLE 4 does not result in classifications that are either mutually exclusive or collectively exhaustive in defining these perspectives. With these caveats acknowledged, we boldly proceed.) The statistician tends to explore systems by the application of probability and statistical models. Variability is viewed in light of random departures from a specified model whereas all other uncertainty is typically viewed as a bias associated with differences between the model and the state of nature. Toxicologists often classify uncertainty as being associated with either a parameter or a model. To the toxicologist, parameter uncertainty (e.g., for some component of a physiologically-based pharmacokinetic model) is more likely to be viewed as mea-

TABLE 4. Uncertainty and variability as viewed by statisticians, risk assessors, and toxicologists

Statistician	Risk Assessor	Toxicologist
1. Variances and covariances are used to evaluate subject, technician, season, and other sources of variability (*true* variation).	1. Hazard identifications—are exposure E and disease D independent given characteristics and context C, where C represents confounders or other risk factors (gender, age, etc.) Is $f(E,D\|C) = f(E\|C)\, f(D\|C)$?	1. Parameter uncertainty—measurement error (variance and bias) and surrogate data (species extrapolation, e.g., using effects in mice to estimate human risk).
2. Uncertainty is reflected in bias—unadjusted (or even unrecognized) confounders, miscalibration of instruments, etc.	2. Hazard characterization—examine the probability distribution of D given E and C [$f(D\|E,C)$]—uncertainty rests in low dose extrapolation and species extrapolation.	2. Model uncertainty—gaps in understanding disease processes or the disposition of xenobiotics—e.g., linear nonthreshold model for carcinogenesis or the construction of physiologically-based pharmacokinetic models.
	3. Exposure characterization—the joint distribution of E and C [$f(E,C)$]—uncertainty arises from scant or absent information about the joint occurrence of E and C.	3. *True* variation—across space, time, individuals (e.g., biologic repair mechanisms) and concerns related to concurrent exposures (confounding, synergy, effect modification, etc.).
	4. Risk determination—can we determine the likelihood of adverse responses as a function of dose—$f(D)$. Note that other kinds of conceptual complexities and political imperatives occur when *integrating out* E and perhaps C.	

surement error, in contrast to the statistician's commonly employed construct of sampling variation. Most toxicologists would also be sensitive to the biological variability of the systems being modeled and they might worry about using a single set of parameter values for making statements about a population. The toxicologist's concern about model uncertainty are analogous to the statistician's concerns about bias in model selection. The statistician and toxicologist would agree that interindividual differences reflect variability, and they are likely to concur that differences in other factors (e.g., space or time) would also contribute to variation in outcome and should be incorporated in an analysis. Some important problems of exposure (confounding, effect modification, and synergy) would be described by a statistician as bias in model specification although the toxicologists might consider this as variation.

A risk assessor may see uncertainty and variability in light of the components of the risk assessment process. In our TABLE 4, we have discussed the risk assessor's considerations of uncertainty and variability in light of probability distributions (no real surprise, given the bias that we bring as statisticians to this discussion). Hazard identification involves evaluating whether hazard exposure (E) and disease (D) are independent given characteristics and context (C). In statistical terms, the question is whether $f(E, D|C) = f(E|C)f(D|C)$. This is a question of partial correlation (after removing the effects of C) in the relationships between E and D. Exposure characterization explores the joint distribution of E and C, $f(E,C)$, where uncertainty arises from scant or absent information about states of nature, expressed in the joint distribution of E and C. Hazard characterization examines the probability distribution of D given E and C, $f(D|E,C)$] with the greatest uncertainty often resting in low dose extrapolation and species extrapolation. Risk determination determines the value of $f(D)$, the frequency or severity of adverse responses, under the specified distribution of conditions as a function of dose, where the conceptual complexities and political imperatives expressed by E and C are integrated out. Note that other disciplines and perspectives could have been considered here. For example, a physician or an epidemiologist may work to a model more like those for infectious disease, where many individuals are exposed but individual factors determine who gets the disease.

REFLECTIONS ON UNCERTAINTY

Uncertainty can be defined as a lack of precise knowledge about the state of nature. This creates practical problems in determining how to assess and deal with, first, the uncertainty itself, and second, estimates of risk that necessarily embody great uncertainty. Since major uncertainty is always with us in this field, it is crucial to learn how to conduct risk assessment in the face of it.

EPA decision-makers have long recognized the usefulness of uncertainty analysis, but they have made only slow headway in replacing *ad hoc* procedures based on a few simple but sweeping assumptions with procedures based on information about the range of risk values consistent with biologic mechanisms of carcinogenic or other toxic effects, the current knowledge of biology and chemistry, or even actual exposures in some group. This is beginning to change as risk assessment based on understanding of biological mechanisms increases, but replacing a few heroic as-

sumptions with a vast number of individually much smaller assumptions will not necessarily reduce uncertainty or get us any closer to true answers. For example, if there is uncertainty about which of the biologic models is correct, great refinement, in either—or even both—may do little to reduce uncertainty in the risk estimates.

One way to examine uncertainty in risk assessment is to classify sources of uncertainty according to the step in the risk assessment process in which they occur (see the *Risk Assessor* column in TABLE 4), including such matters as the failure to identify some hazards, incompatible results from different, but similar studies, estimation of parameters of biological response, the lack of some critical exposure data, and balancing the use of poor epidemiologic data against better but less relevant animal data. This is often done, and now commonly categorized under four headings that may be designated hazard identification, hazard characterization, exposure-response characterization, and risk determination.[4] (These terms vary a bit from author to author, though the four-way breakdown is widely used.) This still does not deal in a satisfactory way with all major uncertainties. For example,

- relating ambient exposures to internal tissue doses,
- extrapolation across mammalian models,
- extrapolation across different routes of exposure,
- extrapolation from lifetime exposure (of workers or animals) to intermittent, and usually lower, exposures of the general population,
- appropriate averaging times for exposure (instantaneous, day, year, etc., where the best answer may depend on biologic half-life).

Another approach, which the Committee[3] adopted, has already been developed as one basis for EPA regulations. This begins with three sets of terms, for parameter uncertainties, model uncertainties, and true variability. Each of these may include important aspects of the two schemes already mentioned—variance/bias and phase of analysis—but the basic approach is quite different. Uncertainties in parameter estimates may arise from measurement error (including such things as random errors in analytic devices, as well as systematic bias), the use of generic or surrogate data in lieu of direct analysis of the parameter to be estimated, misclassification of subjects, random sampling error, and other kinds of non-representativeness.

Model uncertainty arises because of gaps in the scientific theory that is needed to make predictions about risk on the basis of causal inferences. An example is provided by the controversy about whether the linear, non-threshold model for carcinogenesis is sufficiently accurate to be used in setting *conservative* limits to exposure to carcinogens. Evidence suggests that it is not.[5] Other kinds of model uncertainties include errors in understanding relationships and oversimplified models of reality. Important variables may be omitted or perhaps not even recognized as relevant at the time the model is used. The model may fail to account for nontrivial correlations, or miss potentially important confounders or effect modifiers. An important example has to do with the extent of aggregation used in the model; the modeler may either underaggregate (as in considering effects on separate cells or tissues rather than the body as a whole) or overaggregate (as in including persons with a variety of prior exposures that may affect their individual responses). Another example of model uncertainty arises in the use of physiologically-based pharmacokinetic (PBPK) models to relate exposure to internal doses.[6]

True variability—across space, in time, or among individuals—complicates the search for a single value that captures some important aspect of risk. Examples include changes over time in the emission of toxic agents and person-to-person differences in susceptibility. In addition, changes over time may occur in individual susceptibility due to age, diet, or other exposures.

Some scientists now believe that Monte Carlo methods offer a separate and special approach to uncertainty analysis. This is not true, of course; Monte Carlo methods are no more than a computational means to estimate outcomes of complex and mathematically intractable models developed in other ways, not a method in themselves.[7] However, it seems that this perspective is so deeply embedded in the culture of risk assessment that views, as well as terminology, about Monte Carlo methods will be difficult to change. This is not to suggest that Monte Carlo methods are not useful.

The need for Monte Carlo simulation is evident from the TABLE 5, taken from the Committee report.[3] It shows some of the key variables in risk assessment for which probability distributions might be needed. Although not all of these may be needed in every risk assessment, this is by no means a complete list. Furthermore, both distributional forms and parameters of these forms are not generally known, with some forms far from Gaussian, so that the convolutions become mathematically difficult and of unknown reliability. Clearly, computations in risk assessment may become very highly complex. A further problem is that the goal of some environmental regulatory agencies is to protect the most sensitive members of the most sensitive group with a high degree of certainty, with overall allowable risks not greater than, say, 1 in 100,000 or 1 in 1,000,000. This requires untestable assumptions regarding extrapolation from experiments with small numbers of animals, or even from epidemiologic studies with mere tens of thousands of persons. As a final observation on the use of Monte Carlo methods, sampling of different model variables (*parameters*) is often conducted by assuming that the variables are independent. Although this may be true for many variables, correlations clearly exist for others. For example, in physiologically-based pharmacokinetic models, volumes of compartments are probably positively correlated but partition coefficients are probably independent.

Bayesian methods of statistical analysis, which deal with subjective probabilities and the totality of relevant knowledge from *other* sources, have many attractions. There is some history of using subjective approaches to probability distributions in risk assessment, although this history seems to be focused on well-identified components of the risk assessment rather than on the outcome of the assessment as a whole. The NAS Committee Report[3] noted the following:

> Objective probabilities might seem inherently more accurate than subjective probabilities, but this is not always true. Formal methods (Bayesian statistics) exist to incorporate objective information into a subjective probability distribution that reflects other matters that might be relevant but difficult to quantify, such as knowledge about chemical structure, expectations of the effect of concurrent exposure (synergy), or the scope of plausible variations in exposure. The chief advantage of an objective probability distribution is, of course, its objectivity; right or wrong, it is less likely to be susceptible to major and perhaps undetectable bias on the part of the analyst; this has palpable benefits in defending a risk assessment and the decisions that follow. A second advantage is that objective probability distributions are often far easier to determine. However, there can be no rule that objective probability estimates are always preferred to subjective estimates, or vice versa. (NRC, 1994.)[3]

TABLE 5. Examples of variables in risk assessment for which probability distributions might be needed[a]

Model component	Output variable	Independent parameter variables
transport	air concentration	chemical emission rate; stack exit temperature; stack exit velocity; mixing heights
deposition	deposition rate	dry-deposition velocity; wet-deposition velocity; fraction of time with rain
overland	surface-water load	fraction of chemical in overload runoff
water	surface-water concentration	river discharge; chemical decay coefficient in rivers
soil	surface-soil concentration	surface-soil depth; exposure duration; exposure period; cation-exchange capacity; decay coefficient in soil
food chain	fish concentration	water-to-fish bioconcentration factor
	plant concentration	plant interception factors; weathering elimination rate; crop density; soil-to-plant bioconcentration factor
dose	inhalation dose	inhalation rate/body weight
	ingestion dose	plant/soil ingestion rates
	dermal-absorption dose	exposed skin surface area; soil absorption factor; exposure frequency; body weight
risk	total carcinogenic risk	inhalation/ingestion carcinogenic potency factors; dermal-absorption carcinogenic potency factors

[a]Modified from TABLE 6 in Reference 3.

The report noted that there are substantial differences between these approaches, particularly the inability to include aspects of certainty for which there is no quantitative measure, or even to include an extra *fudge factor* to allow for sources of uncertainty not yet thought of. Overall, to explore the difference between parameter uncertainty and model uncertainty, consider the following example in which cancer risk is treated as a parameter to be estimated. In situation A, 3 of 60 animals in a test group develop cancer. This outcome may be regarded as having a distribution very close to Poisson, and the observed frequency of 5% (3/60), has ordinary 95% confi-

dence bounds of 1.03% to 14.6%. Larger samples will ordinarily lead to narrower confidence bounds. In situation B, the chemical under test is thought to follow one of six well-identified chemical pathways, but it is not known which, and these are considered to be about equally likely. The pathways have different means, but the sum of the probability distributions of risk is close to that for the single Gaussian distribution in situation A. Even if sample sizes in B are very much larger than in A, major uncertainty will remain because the level of risk is determined almost exclusively by the biologic situation, and we do not know which biologic model is correct. In more familiar terms, if risk depends on gender, any single population-risk estimate will refer to a hypothetical person who is unknown and will not provide an accurate estimate for any population member. In either situation A or B, it would be mathematically correct to say something like the following: "The expected value of the estimate of the number of annual excess cancer deaths nationwide caused by exposure to this substance is five per million persons; the lower and upper confidence bounds on this estimate are 1.03 per million and 14.6 per million." However, the risk manager who is developing regulations to protect the public health might respond to these situations in quite different ways because of the difference in the source of the uncertainty. The problem is that the summary statement in Situation B, dominated by model uncertainty, obscures important information about the state of scientific knowledge.

In a more extreme example, every knowledgeable scientist may agree that some substance is likely to pose no risk to a population, but that there is one chance in a thousand that it will cause a major disaster. The mode, the median, the 99-percentile are all quite firmly at zero risk. How should the risk regulator respond? Risk managers and the public should be given an opportunity to understand the sources of the controversy, to appreciate why the subjective weights assigned to each model have their given values, and to judge for themselves what action is appropriate when the various theories, of which at most one can be correct, predict such different outcomes. Simply put, it is not in the best interest of either the decision maker or society to treat fundamentally different kinds of predictions as quantities that can be averaged and compared directly without considering the effects of each prediction on the decision it leads to. There may also be quite profound impacts on needs for further information; in one case, one needs a substantially larger sample size, and in the other, one needs additional fundamental research on the mechanisms involved.

There is still another fundamental issue to deal with here. Philosophers, including philosophers of statistics, have noted that the probability of any unobserved event must be either zero or one; the problem is that we do not know which. This is of course built into the foundations of statistics, whether one's basic approach begins with concepts of relative frequencies, or strength of belief, or axioms of probability. The calculated probability is thus a measure of our ignorance. However, this matter is not well understood by the public, and it seems that large numbers of people do not look on probabilities in the same way that statisticians do. This has led to a large literature on risk perception and risk communication, matters that should be well understood by any statistician who may want to enter this field.[8]

Finally, it is quite fair to say that we rarely know as much as we think we do. We observe our data, go through our statistical computations, and find confidence bounds; we then add uncertainty for this, that, and the other thing; to come up with

wider limits for plausible risks, but even then it seems that we fail to account for how little we truly know about the world around us. Recall that the four estimates of risk associated with Tris exposures ranged over nearly three orders of magnitude. Each of the four reports on these risk assessments included extensive discussion about the uncertainties, with an estimate of how far off the final assessment might be. Each of these statements of uncertainty was on the order of a factor of ten. Since orders of magnitude are multiplicative, even these four experienced and knowledgeable risk assessors, working from much the same data base, differed by close to one hundredfold more than any one of them thought plausible in terms of his own analysis. This situation is not at all uncommon. (The observation that we never know as much as we think we do is most certainly not limited to risk assessment. It seems to be pervasive throughout science and throughout our daily lives. The discrepancy between what is known and what is thought to be known probably varies with investigator as well.) This subjectivity allows for investigator bias, whether conscious or subconscious, to enter into analyses.[9]

When uncertainty is great, it is important to devise philosophies, political agendas, computational procedures, and regulations that accommodate this uncertainty. An approach at one extreme would be to assume that regulations should not deal with any risk until it has been adequately demonstrated to be present. (*Adequate* is of course subjective.) This puts the focus on the lower end of the range of uncertainty. This has been described as the "wait for dead bodies" method. Another approach is to focus on the high end, and regulate and otherwise treat the risk as if it were at the greatest level not excluded by a generous interpretation of present information. (This could be a counsel of paralysis—progress would be much slower, but we would be safe from the possible hazards of innovation and risk would simply vanish—except perhaps, the risk of death from boredom.) It might be argued that excessive risk regulation may increase the costs of doing business for some industries which, in turn, may in turn decrease productivity and even lead to economic conditions under which workers lose jobs. Unemployed workers may then be at greater risk of adverse health outcomes. Thus, overly conservative assumptions in risk assessment may result in increasing other risks.

A third approach would be to pick some measure of central tendency. For this, the mean might be preferred over the median, because the median might often be zero though there is an appreciable likelihood of a very substantial harmful effect. For example, in Bailer and Dankovic,[6] the distribution of risks from a Monte Carlo simulation of methylene chloride was bimodal where the two modes differed by three orders of magnitude. (The different modes corresponded to different simulated patterns of cancer response data. One pattern suggested a linear component of dose-response was possible and the other was strictly nonlinear.) In this example, the estimated median risk was significantly less than the estimated mean risk. It is likely that no measure of central tendency is uniformly best for all contexts. This matter requires considerably more public discussion. In this discussion, however, we would strongly discourage the use of logarithmic or geometric distributions, orders of magnitude, and the like for communicating risk. The reason is that many people tend to equate an order of magnitude at one point of the scale with an order of magnitude elsewhere, whereas the difference in their practical effects may be immense—in fact, orders of magnitude. Furthermore, using logarithmic or similar distributions for the

purpose of stabilizing variance and/or making observed distributions more tractable may lead to unwarranted confidence in results. The practical effects of such steps are to substantially reduce the impact of high observations, whereas those high observations are arguably of most concern to us. They ought to be studied in detail and have a major impact on the result.

We all deplore the great uncertainty in risk assessment, yet there seems to be little careful analysis of how much certainty is needed. Does it matter if we are off by a factor of two? A factor of 10? Or 100 or more? We do seem to find value in what we do now, which may be off by a factor of 1,000. In the face of this great uncertainty, the approach that has generally been used by the EPA and other regulatory agencies is to develop *default options* for model uncertainty. These are assumptions that certain approaches to modeling will be used and taken to be correct unless there is evidence to the contrary. Examples are, to assume that the dose-response relation is linear down to a dose of zero, that extrapolations from small rodents to humans should be on the basis of surface area or body weight raised to a specified exponent, and that an agent toxic by one route of administration is toxic by another. In practice, however, it has been extremely difficult to persuade any agency that the evidence is strong enough to overturn any default option; default options are generally considered *conservative*, in that their results are thought to indicate more toxicity than is likely (though not the most extreme results). The effect of this is that the actual level of public protection may often be somewhat higher than the formal estimates would indicate, though the default options may sometimes be quite seriously in error in the other direction.[5] Another effect of reliance on default options may be to reduce the appearance of uncertainty in the risk estimate. If default options are treated as having no uncertainty whereas real data always include uncertainty, then increasing the number or scope of default options might appear to make the answer more and more precise (less variable). Thus, the use of default values, if not accompanied by associated uncertainty, may have the effect of masking much of the nonrandom uncertainty. Returning to our consideration of disciplinary perspectives, not every discipline will necessarily see this as bad. From a legal standpoint, and perhaps from a risk management standpoint, it is easier to make a case for regulating a hazard that is assumed to be precisely known (even if it is a biased estimate) than for regulating a hazard with a very uncertain risk. From either a mechanistic (toxicologic) point of view, or from a statistical point of view, we tend to see a realistic appraisal of the true uncertainty as a good thing. It is not clear that the entire risk assessment community would agree.

Although safety above the estimated level may be most prudent, many people, particularly those allied with chemical manufacturers and users, often complain that we should instead be using our best estimates. It is not clear what *best estimate* means, especially in the face of serious model uncertainty. If knowledgeable scientists are 99% sure that the risk from some exposure is zero, but believe there is a 1% chance that it may be as high as one in one thousand persons, should we assume that the risk is nil? We think not. Even at this point, debate still remains. One response in one thousand exposed individuals is assumed to represent significant impact. Additionally, can risk be reduced in light of feasibility constraints (economic and technological) without inducing other outcomes that may effect health adversely (e.g., unemployment, or substituting an untested alternative).

Risk assessment is a process of analysis, not a specific kind of research and not a result, and it must be viewed as a process that is subject to much uncertainty. Our objective in this paper has been to consider uncertainty and variability in risk assessment from the perspective of the different disciplines and actors involved in the process. After this discussion, we further explored how uncertainty and variability are reflected in typical risk assessments. We closed our reflections with observations on how these risk assessment uncertainties should be communicated to risk managers and the public alike. Although we are encouraged with advances in mechanistic understanding in disease processes and in statistical methods for incorporating uncertainty in risk assessments, we believe that the evaluation and exploration of uncertainty are likely to remain an important part of risk assessment for some decades to come.

ACKNOWLEDGMENTS

The authors thank Drs. James Deddens, David Dankovic, and Eileen Kuempel, whose critiques, comments, and suggestions have led to improvements in the presentation in this paper.

REFERENCES

1. ANDERSON, P.D. 1988. Scientific origins of incompatibility in risk assessment. Stat. Sci. **3:** 320–327.
2. BAILAR, J.C., J. NEEDLEMAN, B. BERNEY & J.M. MCGINNIS. 1993. Assessing Risks to Health: Methodologic Approaches. Auburn House, Westport.
3. NATIONAL RESEARCH COUNCIL, COMMITTEE ON RISK ASSESSMENT OF HAZARDOUS AIR POLLUTANTS. 1994. Science and Judgment in Risk Assessment. National Academy Press, Washington, D.C.
4. MCGINNIS, J.M., J.C. BAILAR & R.G. TARDIFF. 1986. Determining Risks to Health: Federal Policy and Practice. Auburn House, Westport.
5. BAILAR, J.C., E.D.C. CROUCH, R. SHAIKH & D. SPIEGELMAN. 1988. One-hit models of carcinogenesis: conservative or not? Risk Analysis **8:** 485–497.
6. BAILER, A.J. & D.A. DANKOVIC. 1997. An introduction to the use of physiologically-based pharmacokinetic models in risk assessment. Statistical Methods in Medical Research **6:** 341–358.
7. SPECIAL ISSUE. 1996. Commemoration of the 50th Anniversary of Monte Carlo. Human and Ecological Risk Assessment **2**(4).
8. NATIONAL RESEARCH COUNCIL, COMMITTEE ON RISK PERCEPTION AND COMMUNICATION. 1989. Improving Risk Communication. National Academy Press, Washington, D.C.
9. KOZLOWSKI, L.T. *et al.* 1996. Smokers are unaware of the filter vents now on cigarettes: results of a national survey. Tobacco Control **5:** 265–270.