Is Risk Assessment Really Too Conservative?: Revising the Revisionists

Adam M. Finkel*

INTRODUCTION

A growing chorus of voices is convinced that quantitative risk assessment (QRA) has evolved into a caricature of itself. According to this view, quantitative estimates of human health risk from environmental pollution (particularly from carcinogenic substances) have become so dependent on unreasonable worst-case assumptions as to be meaningless, alarmist, and counterproductive. The articles in this symposium issue by Dr. Elizabeth Anderson[1] and Dr. Dennis Paustenbach[2] discuss many of the main arguments made by a number of health scientists,[3] policy analysts,[4] and regulators.[5]

* Dr. Finkel is a research fellow at the Center for Risk Management at Resources for the Future in Washington, D.C. He has an Sc.D. in environmental health sciences from the Harvard School of Public Health and a master's degree in public policy from the Kennedy School of Government. He has served as a technical advisor to the National Academy of Sciences, the House Science and Technology Committee, and an environmental research organization in Mexico City, and was editor-in-chief of a weekly newsletter on hazardous waste issues. The contributions of his colleagues Michael Gough, Paul Portney, and John Mankin are gratefully acknowledged. The opinions expressed herein are those of the author and do not necessarily reflect the views of Resources for the Future or the Center for Risk Management.

1. Anderson, *Scientific Developments in Risk Assessment: Legal Implications*, this volume.

2. Paustenbach, *Health Risk Assessments in Toxic Tort Litigation: Opportunities and Pitfalls*, this volume.

3. Maxim, *Problems Associated with the Use of Conservative Assumptions in Exposure and Risk Analysis*, in THE RISK ASSESSMENT OF ENVIRONMENTAL HAZARDS: A TEXTBOOK OF CASE STUDIES 526-60 (D.J. Paustenbach ed. 1989); Sielken, Pitfalls of Needlessly Conservative Aspects of Cancer Risk Quantification and How to Avoid Them, (Jan. 15, 1989) (presented to the American Association for the Advancement of Science, San Francisco, Calif.).

4. Nichols & Zeckhauser, *The Dangers of Caution: Conservatism in Assessment and the Mismanagement of Risk*, in 4 ADVANCES IN APPLIED MICRO-ECONOMICS 55-82 (1986); Huber, *Exorcists Versus Gatekeepers in Risk Regulation*, REGULATION, November/December, 1983, at 23-32.

5. T. Yosie, Speech to the Society for Risk Analysis, Houston, Texas (Nov. 2 1987) (entitled *Science and Sociology: The Persistence of Conservative Assumptions in Risk Assessment*)(as reported in Society for Risk Analysis Risk Newsletter).

The denunciation of QRA as an ideologically-motivated exercise in exaggeration has cropped up from time to time,[6] beginning even before United States federal agencies started to codify their QRA procedures.[7] Only in the past year or two, however, has a critical mass of scientific experts converged to the view that the time has come to reevaluate these procedures, and the first tangible results of this "revisionist" stance are even more recent.[8] Linked temporally and perhaps causally to this growing momentum against "conservatism" in QRA is the view that science has at last begun to offer practical and defensible alternatives to what Anderson calls "overestimated theoretical risk," that is, the production of risk estimates via a process critics claim is rife with scientific shortcomings.[9]

The critics of conservatism are a diverse group. They share the belief that some or all of the inferences central to QRA are overly timid and at variance with new theories or data to the contrary. They differ widely, however, in the scope and intensity of their enthusiasm for specific alternatives, and in their degree of dissatisfaction with the status quo. In order to respond both to Anderson and Paustenbach and to the growing number of published and unpublished critiques of conservatism, I have coined the catchall term "revisionist." I emphasize that this is a stereotype which describes none of these critics precisely, but which is a composite of documented views rather than a "straw man."

My hypothetical revisionist believes that at virtually all phases of QRA, the preoccupation with the "worst case" has driven out scientific rules of reason, and tends to believe that all new theories or data sets, if valid, will likely reduce existing estimates of biologic potency or human exposure. Not everyone who advocates the rollback of particular conservative assumptions is a revisionist, and probably no individual critic of QRA could fall prey to all of the misperceptions I mention in this article. Nevertheless, this stereotype is not so limited that it refers only to some-

6. E. EFRON, THE APOCALYPTICS: CANCER AND THE BIG LIE (1984).

7. E.P.A. Guidelines for Carcinogen Risk Assessment, 51 Fed. Reg. 33,992-34,003 (1986).

8. Shabecoff, *E.P.A. Reassesses the Cancer Risks of Many Chemicals: Hazards Seen as Lower*, N.Y. Times, Jan. 4, 1988, at A1; EPA, Office of Health and Environmental Assessment, Update to the Health Assessment Document and Addendum for Dichloromethane (Methylene Chloride): Pharmacokinetics, Mechanism of Action, and Epidemiology (1987) (External Review Draft EPA/600/8-87/030A).

9. Paustenbach, *supra* note 2, at Section I.

one who would prefer never to assess or control risks in the absence of unequivocal evidence of harm. Finally, from the perspective of an EPA official charged with responding to calls for QRA revision, all of the revisionist arguments must be equally dealt with, even though they come from individuals with diverse perspectives, backgrounds, and motives.

While I advocate certain changes in QRA, I do not share the revisionist disdain for risk assessment as currently practiced. I am not so confident there exists systematic conservatism in QRA, that if so it would be a scourge we must repudiate, or that the new science necessarily provides a more appropriate or even a more credible alternative.[10] I reject as disingenuous the characterization that "[conservatism] allows for policy choices to masquerade as if they were scientific facts."[11] Revisionists also tend to selectively marshall and interpret facts to support or supplant policy choices. Moreover, I question the pejorative description of conservatism as persisting solely because bureaucrats appreciate its rigidity or because they live in fear of the "accusation that their [risk] assessments are not sufficiently health protective."[12] While I agree with many of the revisionists that ideally, QRA would not subsume covert value judgments that ought to be part of the overt balancing process that is risk management, I believe there is no workable alternative to the current mingling of these related activities. I welcome efforts to make the assumptions and uncertainties in QRA more explicit, so that as a society we can discern how conservative regulatory proposals actually are. I do not believe, however, that merely scaling back some or all of these assumptions will achieve a value-free process with the desired separation of risk assessment and risk management.

In short, I believe that some of the critics of conservatism have fallen prey to some or all of a series of misperceptions about the science and craft of QRA. These most often result from insufficient attention to the twin influences of scientific uncertainty and human interindividual variability, so the lion's share of the arguments raised in this paper will concern these two under-appreciated phenomena. Indeed, I will conclude in part that when risk assessment and management are reevaluated in light of uncertainty and variability, the narrowness of the entire debate over

10. *Id.*
11. Sielken, *supra* note 3.
12. Paustenbach, *supra* note 2, at Section III.

whether the risk numbers we generate are too high or too low will be revealed.

At the outset, I emphasize that I do not share the belief, which Paustenbach attributes to Commoner,[13] that QRA is a method for legitimizing the discharge of pollutants in quantities preferred by industry. Like Paustenbach, I look forward to a time when QRA will be even more influential in helping regulators and the public discriminate between significant and *de minimis* risks. Moreover, I share his apparent view that blanket prejudices against the use of QRA threaten to make us all less safe, and may lead us to squander the finite resources available for risk reduction. Nor do I dispute that we need to guard against the reflexive and exclusive use of worst-case assumptions. In fact, neither Anderson nor Paustenbach mentions two very potent arguments against conservatism that I think the scientific and regulatory communities must address.

First, we need to recognize an inescapable paradox about conservatism—in its attempt to impose uniformity on the mechanics of risk assessment, it virtually guarantees marked non-uniformity in the outputs of risk assessment. Some worst cases are simply "worse" than others, in the sense of being less plausible to occur (or less frequent in occurrence), but conservatism tends to obscure these differences. For example, even a seemingly innocuous contrivance such as "assume that at all hazardous waste sites, the maximally exposed individual (MEI) is exposed to the concentration measured directly downgradient at the property boundary of the facility" introduces what Nichols and Zeckhauser term "asymmetric conservatism."[14]

In some cases, the resulting risk estimate will be quite conservative if the geography is such that the MEI lives hundreds of yards away from the boundary and/or is not directly downgradient of the pollutant source. In other cases, however, the assumption may be nearly correct and thus barely conservative. In statistical terms, regulators might then be forced to compare a pair of risk estimates, perhaps one that had only a 1 in 1000 chance of being too low and another that had a 1 in 10 chance of being too low, yet this information about the different character of the estimates would be unavailable. Such a comparison might well result in the

13. *Id.*

14. Nichols & Zeckhauser, *supra* note 4, at 69.

allocation of greater resources to address the relatively less important risk, at the expense of the more significant one.[15]

Second, certain aspects of conservatism can provide a powerful disincentive for needed research. For instance, if one or more animal carcinogenicity bioassays already exists for a substance, and if the relevant regulatory agency is on record stating it will use only the assay yielding the highest estimate of the cancer potency factor, essentially all incentive for an industrial concern to conduct or fund additional animal tests is removed by virtue of this "stacked deck."

I have no doubt that as QRA matures, we will discover that conservatism has caused some of our existing assessments to enormously exaggerate particular human risks. Some fragmentary evidence has already accrued in this regard. For example, the discovery of a protein (alpha$_{2u}$-globulin), arguably unique to male rats and essential for carcinogenesis to occur in the male rat kidney following certain stimuli,[16] suggests that our estimates of the human carcinogenic potency of certain substances (perhaps unleaded gasoline) were qualitatively in error (the true human cancer risks may in fact be zero). In these cases, the discarding of all other test results and the preoccupation with the single positive response may have been the undesirable outcome of a conservative stance.

TABLE I
NINE PERVASIVE MISPERCEPTIONS
ABOUT RISK ASSESSMENT "CONSERVATISM"

1) There is no such thing as an "actual risk."
2) Conservatism is inherently no more or less biased a method than alternative approaches.
3) Only some conservative assumptions are gratuitous.
4) Not all of the inferences we make are in fact conservative.
5) A cascade of truly conservative steps may still yield a reasonable estimate of risk.

15. In statistical terms, the first estimate would lie at the 99.99th percentile of an uncertainty distribution about the (unknown) true population risk, and the other would lie at the 90th percentile of its distribution. See Section X *infra*, however, for a discussion of how eliminating conservatism would not necessarily ameliorate this ranking problem.

16. Charbonneau, Short, Lock & Swenberg, *Mechanism of Petroleum-Induced Sex-Specific Protein Droplet Nephropathy and Renal Cell Proliferation in Fischer-344 Rats: Relevance to Humans*, in 21 TRACE SUBSTANCES IN ENVIRONMENTAL HEALTH 263-73 (1987).

6) Conservative assumptions at some stages of risk quantification may correct for the omission of other stages altogether.
7) More science does not always mean less risk, nor does it always "reduce uncertainty."
8) Costs, like risks, can be biased or *de minimis*.
9) There is a major difference between the pooling of data and the averaging of irreconcilable theories or results; the latter is a perilous process laden with hidden value judgments.

Nevertheless, these objections to current QRA procedures or results pale by comparison to my skepticism that all the revisionist positions are fully thought through. The remainder of this article constitutes a cautionary note warning against hasty or piecemeal revision of existing risk assessment procedures. A closer look at the following nine common misperceptions about QRA and conservatism (see Table I) may cast doubt on Paustenbach's contention that "something went wrong"[17] with risk assessment, and suggest that much of the revisionist "wish list" for the future is objectively incomplete, potentially imprudent, and at best no less problematic than the status quo.

I. THERE IS NO SUCH THING AS AN "ACTUAL RISK"

Some revisionists suggest that all of the problems with contemporary risk assessment stem from the "unreality" of its outputs. They claim that instead, scientists ought to provide risk managers with estimates of "actual risk," or what Anderson calls "real risk."[18] In this view, now that science has begun to develop techniques for making risk estimates more accurate, these should replace conservative estimates, as surely as the Copernican model of the solar system replaced the Ptolemaic model. As Section II of this paper will suggest, the desire to promote decisions based on actual estimates of risk may not be free of subtle and potentially troubling value judgments. But more fundamentally, there exists widespread confusion about the concept of real risk itself, which is, in many ways, nonsensical.

All risks are probabilistic summaries of unknowable future events—they can describe the long-run or average behavior of simple systems, but the behavior of subsystems or of individual

17. Paustenbach, *supra* note 2, at Section II.
18. Anderson, *supra* note 1, at Section I.

members of the system (as well as the short-run behavior of the whole) may diverge markedly from any prediction based on a risk estimate. Consider a simple homogeneous system obeying well-understood but stochastic physical laws—a chunk of the radioactive isotope uranium-238. Science may determine virtually exactly that the half-life of the isotope is 4.51 billion years, so that in that time period, almost exactly half of the atoms in an initially pure sample would have decayed into lighter isotopes. It makes no sense, however, to try to predict when a particular atom in the sample will decay, or whether it will still be a U^{238} atom at a particular time in the future, or even to bank on a precise estimate of how many decay events will occur in a short time period and/or among a small subpopulation of atoms. Yet these are some of the kinds of inferences an actual risk would have to allow us to make in order to be qualitatively better than the "unreal" estimates current QRA provides.

Of course, in moving from this idealized example to the practical area of human health risk assessment, the system becomes even more complicated along several dimensions, further defeating the goal of reaching actual risk. First, there are considerable difficulties even in estimating the true long-run average probability in real situations. The uncertainty in the half-life is primarily due to simple measurement error, which we can reduce further by improving our analytical devices.[19] Uncertainties in exposure estimation and carcinogenic potency assessment take many different forms and can be quite recalcitrant to brute force methods of data acquisition. Often, obtaining enough exposure or potency observations to make sampling error manageable is impossible or quite expensive, and then a host of non-random parameter uncertainties (*e.g.*, the "healthy worker effect" in occupational studies) and fundamental modeling uncertainties (*e.g.*, are Gaussian dispersion or linear dose-response models appropriate?) further confound efforts to converge on a single probability or risk number.[20]

Second, the physicist can rely on every atom of U^{238} being alike—they will not all decay at the same time, but they all face the same underlying chance of decaying at any moment in time.

19. Presumably, there is virtually no binomial or sampling error in the calculation because the number of observations (decaying atoms) is so large.

20. Finkel, Perspectives on Uncertainty in Risk Assessment: A Guide for Decision-Makers (1989) (Center for Risk Management Report), at 9-39.

Human beings, however, exhibit enormous interindividual variability, both in the exposures they receive (due to geographic, lifestyle, and genetic factors)[21] and in their susceptibilities to each exposure (due to other genetic, temporal, and lifestyle factors).[22] So, even if an individual could know his exposure pattern precisely, a potency estimate that may by chance accurately describe his probability of death or disease at a given moment will certainly fail to yield a real risk for some other person, or even for the original person as his age or environment changes.

Finally, for a number to be an actual risk, one must believe that, both for the regulator and the exposed public, the concept of risk can be reduced to an unalloyed numerical value. This conclusion is tempting in light of the classical theory of individual utility, which holds that while people can be risk-averse with respect to an outcome (*e.g.*, decline a chance to flip a coin to win $100 or lose $50, even though the expected value of the wager is positive), rational thought precludes any special weighting of the *probabilities* of an outcome.[23] In other words, risk numbers only modify the ultimate outcome (disease or death), and therefore one ought to view a risk of 10^{-3} as exactly ten times as bad as one of 10^{-4}. Furthermore, one should be indifferent between the following two descriptions of risk, whose expected values are equal: 1) one's risk is determined to be 10^{-5}, and this value is known with certainty; and 2) there is a fifty-fifty chance one's true risk is either 2×10^{-5} or zero.

This view has recently been challenged by researchers concerned with human cognition and perception,[24] who reject on both empirical and theoretical grounds the "linearity assumption"[25] that allows risks to be averaged and otherwise manipulated as if they had no special connotative meanings. Therefore, in the general case when the value of a probability is uncertain,

21. Ott, *Total Human Exposure—An Emerging Science Focuses on Humans as Receptors of Environmental Pollution*, 19 ENVTL. SCI. & TECH. 880-86 (1985); Wallace, *The Influence of Personal Activities on Exposure to Volatile Organic Compounds*, ENVTL. RESEARCH (submitted 1989).

22. Finkel, *Estimating the Extent of Human Variability in Susceptibility to Carcinogenesis*, RISK ANALYSIS (1988) (in press).

23. H. RAIFFA, DECISION ANALYSIS: INTRODUCTORY LECTURES ON CHOICES UNDER UNCERTAINTY 57-61 (1968).

24. Tversky & Kahneman, *The Framing of Decisions and the Psychology of Choice*, 211 SCIENCE 453 (1981); Lopes, *Some Thoughts on the Psychological Concept of Risk*, in J. EXPERIMENTAL PSYCHOLOGY: HUMAN PERCEPTION AND PERFORMANCE 137-44 (1983).

25. K. SHRADER-FRECHETTE, RISK ANALYSIS AND SCIENTIFIC METHOD 157-95 (1985).

calling any single numerical summary of that uncertainty an actual risk may ignore human values and perceptions; even if that probability were somehow known with certainty, it may be similarly unresponsive to compare it with another actual risk, or to convert the risk to a measure of social cost, without considering factors other than the probability itself.

None of this discussion is intended to deny that certain incremental changes in QRA procedures may be motivated by disparities between existing predictions and observational experience, or to deny that they may move particular risk estimates closer to values we deem more likely to be borne out in the future. Proponents of change "oversell" these adjustments, however, if they equate "more likely to be borne out" with "real." First, unlike Copernican celestial mechanics, we will never actually be able to test the accuracy of alternative risk estimates against real human experience, although this alone does not militate against the quest for better numbers. Second, the risk assessment process is sufficiently complex that we cannot be confident that adding a dose of reality at one stage will necessarily improve the reality content of the overall output. Third, uncertainty is sufficiently pervasive that no new theory is likely to yield an estimate that automatically supplants the standard one; such improvements provide more information, not definitive information. Again, the claim that new approaches "provide the regulatory or legal arena with an indication of the extent to which the plausible upper bound may be overestimating risk for particular chemicals,"[26] which may be tantamount to saying they yield real estimates, overstates the case for revision. Ideally, conservative estimates have the virtue of not being advertised and utilized with reference to their reality content. In practice, although EPA often fails to communicate that its risk estimates are construed to be upper bounds, on occasion it overstates the case and implies there is virtually zero possibility that risks could be higher than stated.

II. Conservatism is Inherently no More or Less Biased a Method than Alternative Approaches

Critics of current QRA procedures often equate the various conservative assumptions with the statistical term "bias," a word which in everyday parlance connotes intentional disregard for

26. Anderson, *supra* note 1, at Section II(B).

facts or conventions of fairness. The implication of descriptions such as "assumptions that bias the estimates upward"[27] is that lowering the risk estimates would reduce or remove the bias. Again, just as with real and unreal, some revisionists prefer to frame the debate in semantic terms that contrast laudable versus suspicious motives.

The statistical concept of bias, however, refers to an estimator (*i.e.*, a procedure for producing an estimate) which on average does not equal the parameter it is supposed to estimate.[28] This descriptive notion has nothing to do with evaluating the choice of the estimator itself on prescriptive grounds. Using a specific statistical upper confidence limit (UCL) to describe the results of an animal carcinogenicity bioassay, or basing population exposure estimates on the parameter values applicable to a highly exposed individual, can be criticized as a value-laden estimator. But it is not necessarily a biased procedure, unless the critic can show that the analyst's number is systematically different (higher) than the quantity he believes he is estimating (*e.g.*, the desired UCL on potency or the exposure the individual faces). I have no doubt such true bias occasionally does occur, when analysts or regulators treat numbers meant to be particular UCLs as if they were central or real estimates, and then explicitly or tacitly hedge further towards overcautiousness by inflating these numbers further. Such behavior is a problem of education, not an acknowledgement that the estimators themselves are biased. Besides, true statistical biases are probably more likely to occur with central estimators of risk. It is often easier to estimate, to a desired amount of precision, more extreme values than central ones; the average speed of a sample of 10,000 cars on an interstate highway may be difficult to predict (and highly variable depending on conditions), but one might predict rather confidently that the 500th-fastest car travels at between 72 and 76 mph.

But what of the description of conservatism as "value-laden," a term that can turn pejorative when critics allege these values are covert and are "masquerading as facts"? There is no sense denying the judgmental content of conservatism, but many critics seem not to accept that all summary estimators of an uncertain quantity are value-laden. Summary measures are little more than

27. Nichols & Zeckhauser, *supra* note 4, at 57.

28. J. FREUND, MATHEMATICAL STATISTICS 256 (1971).

ways to interpret facts in light of a subjective calculus of the costs of error—differences among values yield differences in the interpretive calculus, but all values can masquerade as (or preferably, enrich) the facts to the same degree. For example, choosing to describe an uncertain quantity by the 95th percentile of its probability distribution merely reflects the conscious or tacit evaluation that an error of underestimation (the five percent chance the "truth" exceeds the summary value) is nineteen times as bad as an error of overestimation.

Most other summary measures simply strike this balance between probabilities and social costs in a different way. For example, the lognormal distribution is one of the most common mathematical formulations of uncertainty used in risk analysis;[29] it can be summarized as a best estimate with a "factor of X" uncertainty surrounding it in both directions.[30] One could use this best estimate (which is the statistical median) to describe the risk, and this would embody the value judgment that the costs of the two types of errors are exactly equivalent (as the probability of each error is fifty percent when the median is chosen). Another common estimator is the mode of the uncertainty distribution, the single value deemed more likely to occur than any other. The maximum likelihood estimator (MLE) advocated by several revisionists[31] is, in certain contexts such as sampling error in animal bioassay data, the mode of the relevant uncertainty distribution. The mode reflects a different value judgment—that one should minimize the probability of an error, without regard to its type (over- or underestimation) or its magnitude.

Indeed, none of these probability-based estimators consider how large the errors might be, except indirectly. As Section V will demonstrate, the mean of an uncertainty distribution is frequently much larger than the median, which in turn is generally larger than the mode. The mean may even be comparable in magnitude to the 95th percentile (or more extreme) "upper

29. Crouch & Wilson, *Regulation of Carcinogens*, 1 RISK ANALYSIS 48 (1981); Finkel, supra note 22.

30. If, as is commonly assumed, the "factor of X" is intended to demarcate a 97 percent confidence region (such that there is a 97 percent chance the true value will lie between the best estimate (B) divided by X and B times X), then there is a 68 percent chance the true value will lie between $B/X^{1/2}$ and $BX^{1/2}$.

31. Sielken, *supra* note 3; Nichols & Zeckhauser, *supra* note 4.

bounds."[32] The mean is the sensible estimator if one believes that the social cost of an error increases with the magnitude of that error (e.g., in the classic "newsboy's problem,"[33] if the true demand is for 100 papers, ordering 10 or 190 papers from the supplier is a more costly mistake than ordering 99 or 101). Estimators even *larger* than the mean may be more appropriate if in addition to caring about the size of errors, there is also a perceived asymmetry in social cost (*e.g.*, an excess newspaper is not as bad as an unsatisfied customer, or an error of overestimating risk is not as bad as an equally large error of underestimation). Therefore, to the extent that some of the central estimates put forth as alternatives to conservatism are not as large as the true mean, and to the extent that even the mean may not capture asymmetries in social cost, these estimators are not only as value-laden as conservative ones—they may be laden with value judgments at odds with those of society as a whole.

III. ONLY SOME CONSERVATIVE ASSUMPTIONS ARE GRATUITOUS

Certain results of a conservative stance are easy targets for revisionist observers. In cases such as the one Paustenbach describes[34] about the dioxin hazard from municipal waste incinerators, a strong case can be made that some risk assessments strain the bounds of credulity. As a practitioner and proponent of QRA, I too am concerned that Congress and the public not lose faith in the process because they construe QRA as an exercise in constructing bizarre hypotheticals, like the child who seems to be getting most of his nourishment from incinerator fallout. Nearly as damaging to the reputation of QRA are instances such as the one Anderson points out,[35] where seemingly obvious questions (how much dioxin is irreversibly bound to soil and thus not bioavailable upon ingestion?) are not asked, and the assessor instead chooses to assume the worst.

32. For example, if the median or "best estimate" is 10.0 in a lognormal distribution with a rather modest "factor of 25" uncertainty, the mean is approximately 36.5, the mode approximately 0.75, and the 95th percentile value approximately 141.0.

33. M. MORGAN & M. HENRION, UNCERTAINTY: A GUIDE FOR DEALING WITH UNCERTAINTY IN QUANTITATIVE RISK AND POLICY ANALYSIS 384-86 (1988)(preprint).

34. Paustenbach, *supra* note 2, at Section IV(D).

35. Anderson, *supra* note 1, at Section II(C).

Such reflexive use of pessimistic assumptions can be termed gratuitous, in that it suggests at least the appearance of laziness on the part of the assessor. To a varying degree, a bit more introspection or research would presumably reveal that such assumptions are simply short-cuts. Moreover, in many cases the assessor would see that if he continued to gather more data, the original conservative estimate would become less and less plausible, in an orderly and predictable fashion.

It would be a serious mistake, however, to malign all of the allegedly conservative aspects of QRA with the same broad brush. In fact, many of the basic assumptions of QRA are conservative both out of respect for how little we know and out of a recognition that as we learn more, the outputs of these procedures may *not* converge towards lower results. Perhaps the best example is the use of the statistical 95th percentile UCL, rather than the MLE, in the analysis of rodent bioassay data. Both the MLE and the UCL are converted into estimators of the carcinogenic potency of the substance; to a very rough approximation, the MLE estimator of potency is the slope of the straight line that gives the "best fit" to the rodent data.[36] Again very roughly, the UCL estimator of potency is the slope of a steeper straight line fitting an alternative set of data points, wherein we hypothesize that if we had repeated the bioassay, the same underlying risks to the animals might have yielded a more pronounced tumor response. If we observe two tumors in fifty rodents, the best estimate is that each rodent had a cancer risk of 0.04 (2 ÷ 50) at that dose. However, we know that even if the true risk was 0.04, if we could repeat the bioassay 100 times, on about five occasions we would observe five or more tumors in this group. If we only had one of those bioassays, we would have concluded that the best estimate of risk was at least 0.1 (5 ÷ 50). In other words, the true risk to each animal could be 0.1 or higher; the chance observation of only two tumors out of fifty is not at all inconsistent with this estimate.

It is well known that depending on the arrangement of the doses and the observed responses, the UCL slope may be many

36. The bioassay data are as follows: at each dose, located on the horizontal axis, the response (on the vertical axis) is the fraction of animals tested which developed tumors in a particular tissue.

times greater than the MLE slope.[37] But this does not mean that the UCL is an outrageous value waiting to be refuted by more data. It is nothing more or less than the lower bound on the exact value we believe we would call the best estimate on five occasions if we repeated the animal experiment 100 more times. Consider this analogy to highlight the distinction between gratuitousness and prudence. If a baseball player approached his team's owner after the first two games of the season and asked for a million-dollar raise because he was batting .800 at the time, the owner would probably think it prudent to wait for the player to amass 100 or more at-bats before caving in. This is because at the time, the owner might believe there was a reasonable chance the player's true average might end up being as low as .200.

We are in the same situation with potential carcinogens, except that bioassays are so expensive and time-consuming that we will never get a "full season's" worth of data on a chemical. While I agree with Paustenbach that the "degree of potential conservatism of the bounding procedure" (that is, the ratio of the UCL to MLE estimators) should be reported in risk characterizations, it is not generally true that "zero risk is as likely as the upper bound value of risk."[38] In many cases, the MLE slope is greater than zero, indicating that we would have to assume the observed data were by chance an unusually strong result of a weak underlying risk (akin to the baseball owner assuming the .800 hitter was actually on a "cold streak" at the time!) to believe that zero potency was a plausible conclusion. Besides, Paustenbach fails to make the distinction between the slope of a linear dose-response function and the potency. Even in cases where an estimator of the linear slope is zero (and the UCL on the linear slope is always positive),[39] this does not mean that the chemical is not a carcinogen, only that its risk decreases more rapidly than linearly at lower doses.[40]

37. T. Thorslund, Estimation of Lifetime Risks Using a Multistage Theory of Carcinogenesis (Feb. 20, 1985)(EPA Carcinogen Assessment Group typescript).

38. Paustenbach, *supra* note 2, at Section IV(B).

39. Guess, Crump & Peto, *Uncertainty Estimates for Low-Dose-Rate Extrapolation of Animal Carcinogenicity Data*, 37 CANCER RESEARCH 3475-83 (1977).

40. This oversight is encouraged by EPA's current reporting procedures, which sometimes treat "potency" as if it refers only to the linear part of the function. If the MLE of the linear part is zero, the MLE for "potency" (really the risk at an arbitrary low dose) is not zero, but comes from the dose squared times the MLE for the "dose-squared term" of the function.

In addition to placing too much confidence in a limited set of data which may not be borne out by further experience, the MLE has the added disadvantage of being extremely fragile to small fluctuations in those data. Paustenbach's example of a one-tumor difference between otherwise identical bioassays conducted in a "New York" and a "San Francisco" laboratory[41] actually demonstrates why EPA and other agencies shun the MLE in favor of the UCL. The exact values for the added risk at a dose of 0.01 mg/kg-day, using the MLEs,[42] are 5.704×10^{-7} in "New York" (or about one chance in 1.7 million) and 9.88×10^{-5} in "San Francisco" (almost exactly one chance in 10,000). These risks do indeed differ by a factor of 174. However, the UCL on added risk is 2.42×10^{-4} in "New York" and 3.02×10^{-4} in "San Francisco," a trivial difference of only twenty-five percent. If anything, this example underscores the folly of using a best estimate that might so dramatically underestimate true risk due to a capricious event (perhaps the rodents in the middle dose group in "New York" were just lucky, or perhaps the pathologist who examined these animals failed to correctly diagnose one (or more) that had tumors). While the UCL may sometimes obscure real differences among chemicals,[43] the MLE is really the more gratuitous estimator, for it ignores how much the risk estimate might differ if more data (or more thorough analysis of existing data) were available.[44]

IV. Not All of the Inferences We Make are in Fact Conservative

In addition to criticizing all of the truly conservative assumptions regardless of their underlying rationales, the revisionist tends to portray all inferences used in QRA as conservative, regardless of whether this description is numerically apt. Statements such as "wherever there was scientific uncertainty, the

41. Paustenbach, *supra* note 2, at Section IV(B).

42. MLE and UCL slopes were calculated via the computer code "MSTAGE87," provided courtesy of E. Crouch.

43. H. Ozkaynak & A. Finkel, Potencies and Unit Risk Values for Suspected Human Carcinogens as Input to Health Risk Assessment (Dec. 18, 1986)(report to Energy and Environmental Systems Division, Argonne National Laboratory).

44. A. Finkel, Computing Uncertainty in Carcinogenic Potency: A "Bootstrap" Approach Incorporating Bayesian Prior Information (Aug. 8, 1988)(report to EPA Office of Policy, Planning and Evaluation). This report suggests alternative ways of characterizing "potency" that do not rely on point estimators such as the MLE or UCL.

most . . . conservative assumption was always chosen"[45] contribute to an exaggerated picture of the risk assessment process as a rote series of timid choices. In some cases, neither the regulator nor the critic knows whether a particular assumption is conservative, yet both now accept the premise that the choice is in fact a biased one. For example, both EPA and its critics now describe EPA's choice of the animal data set from which it calculates its cancer potency factors as a routine of "selecting the most sensitive sex/species combination" available. This statement is objectively true (with the caveat that this sex/strain combination must show a statistically significant tumor increase), yet it conveys too high a degree of confidence that this choice leads to conservative human risk estimates. If EPA generally had access to data on dozens of animal species, and then chose the single sex/species combination yielding the largest potency factor, one might justifiably compare this to the "multiple comparisons" fallacy in epidemiology (wherein data are gathered on so many health effects, without regard to theories of causal association, that by chance at least one effect is likely to show a "significant," but spurious, elevation).[46]

But the connotation of picking the "one bad apple from the barrel" is certainly less justified when one recalls that except in rare instances, EPA has but four sex/species combinations to pore over (males and females, rats and mice). Just because the female mouse may be the most sensitive test animal available does not necessarily imply that it is more sensitive than the human. Paustenbach's accurate statement that current potency estimates are often "more applicable to a rat than a human"[47] is in fact an indictment of the process as nonconservative if in general or in particular cases rats are actually less sensitive than humans.

In several other instances, actual data exist to bolster the argument that particular assumptions are not always conservative. Very few scientists or policy analysts are calling for revision of these procedures to increase their degree of conservatism, yet in at least three major portions of the process, such revision might be justified, if indeed we always wanted to guarantee the most conservative plausible stance:

45. Anderson, *supra* note 1, at Section I.

46. Feinstein, *Scientific Standards in Epidemiologic Studies of the Menace of Daily Life*, 242 SCIENCE 1253 (1988).

47. Paustenbach, *supra* note 2, at Section IV(D).

• *Use of Linear Dose-Response Models.* This inference has been characterized almost universally as the "most conservative plausible model,"[48] yet until recently no attempts had been made to validate this view. Bailar and his colleagues,[49] risk analysts from the Harvard School of Public Health, recently showed that in a significant number of cases, superlinear functions that are steeper at low doses than at higher ones fit observed animal data better than linear functions do. One familiar example where superlinearity has been verified via extensive testing in the low-dose region concerns the human liver carcinogen vinyl chloride (VC). Bailar noted that linear extrapolation using only the control group and the two "high-dose" groups tested in the original bioassay (inhalation of 2500 and 6000 ppm VC) produced a potency estimate nine times lower than that obtained by linear extrapolation using eight "low-dose" groups (between 0 and 200 ppm inhalation exposure). In this case, the "plateau" of response at higher doses is known to result from the saturation of an enzyme system that converts VC to a potent carcinogenic intermediate. Bailar discusses five other biological reasons why the true dose-response relation for a given chemical may "plateau," making it more likely that the EPA procedure yields nonconservative estimates of low-dose potency.

• *Use of Standard Air or Water Dispersion Models.* Although the exposure scenarios risk analysts employ are often conservative (*e.g.*, they assume that all persons drink two liters of tap water daily[50]), the dispersion models used to predict the pollutant concentrations in the water or air are often nonconservative. For example, the various short- and long-term air dispersion models were developed for use in simple terrain, without hills, valleys, or buildings in between the pollutant source and the human "receptor." Such complexities of terrain will cause the models to overpredict exposures in some situations, but underpredict them in others, particularly near the source where concentrations are highest. Also, meteorological conditions not accounted for in existing models, notably the common phenomenon of "fumigation" (the rapid shift from "stable" to "unstable" stratification of the

48. Lave, *Estimating the Risk of Carcinogens*, 1 RISK ANALYSIS 60 (1981).

49. Bailar, Crouch, Shaikh & Spiegelman, *One-Hit Models of Carcinogenesis: Conservative or Not?*, 8 RISK ANALYSIS 485-97 (1988).

50. Maxim, *supra* note 3, at 542.

atmosphere)[51] can cause transient but substantial elevations in pollutant concentration.

Moreover, as Anderson acknowledges, the "source term" (the amount of pollutant emitted or created) in these models may be underestimated. She cites the case of chemical conversion of one pollutant to a more toxic one,[52] a process which can also involve biochemical transformation, as in the generation by microbes of highly toxic methylmercury from emissions of inorganic mercury.[53] Both Anderson and Paustenbach cite examples where the source term is decreasing with time, to show that simply multiplying the current emission rate by the time horizon is too conservative. Counterexamples exist, however, where by virtue of chemical transformation or increased contributions to the source term (as in groundwater contamination from a hazardous waste site where disposal is ongoing) such simple extrapolation is equally nonconservative. Finally, contrary to Anderson's reference to the Tacoma smelter (ASARCO) case,[54] more intensive data collection on emission rates does not necessarily reveal systematic conservatism. In that case, EPA lowered (by roughly four-fold) only its source term estimate for the "process emissions" of arsenic from the main stacks; simultaneously, it discovered that the source term for the "fugitive emissions" from valves, leaks, etc., was underestimated by about a factor of two (since these latter emissions occur much closer to ground level, the net effect on human risk of these adjustments tended towards a balance).[55] The point is not to quibble with the specific example, but merely to show that it is unfair to automatically equate more data with lower exposure estimates.

- *Assumption that Multiple Risks are "Additive" in Nature.* We know that in some cases, exposures to two different substances can have a synergistic effect in elevating excess cancer risk. For example, the cancer rate among asbestos workers with a history of cigarette smoking was found to be more than three times higher

51. F. PASQUILL & F. SMITH, ATMOSPHERIC DIFFUSION 282 (3d ed. 1983).

52. Anderson, *supra* note 1, at Section II(E).

53. *The Impact of Mercury Releases at the Oak Ridge Complex: Hearing Before the Subcomm. on Investigations and Oversight of the House Comm. on Science and Technology*, 98th Cong., 1st Sess. 211-18 (1983).

54. Anderson, *supra* note 1, at Section III(E).

55. Finkel & Evans, Towards Cost-Effective Methods for Reducing Uncertainty in Environmental Health Decision Processes (1985)(proceedings of the Annual Meeting of the Society for Risk Analysis at 543-553.).

FIGURE 1A
PROBABILITY DISTRIBUTION FOR X, ONE UNCERTAIN COMPONENT OF RISK
(X is distributed lognormally, with an uncertainty of a "Factor of 10")

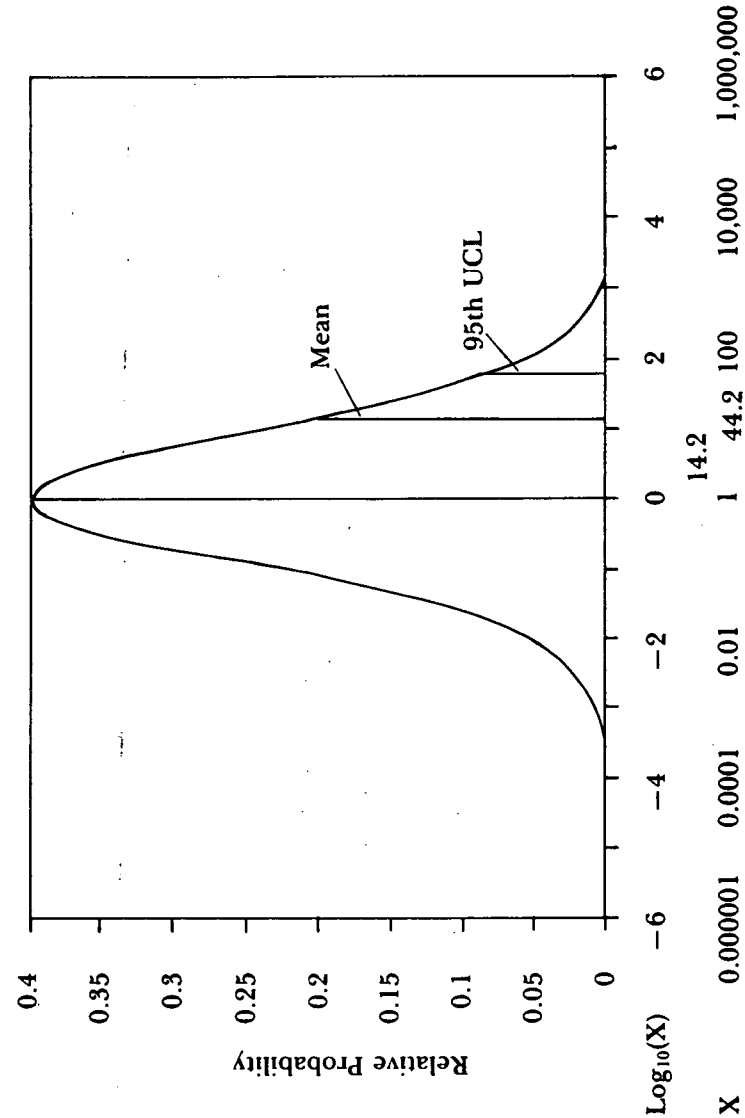
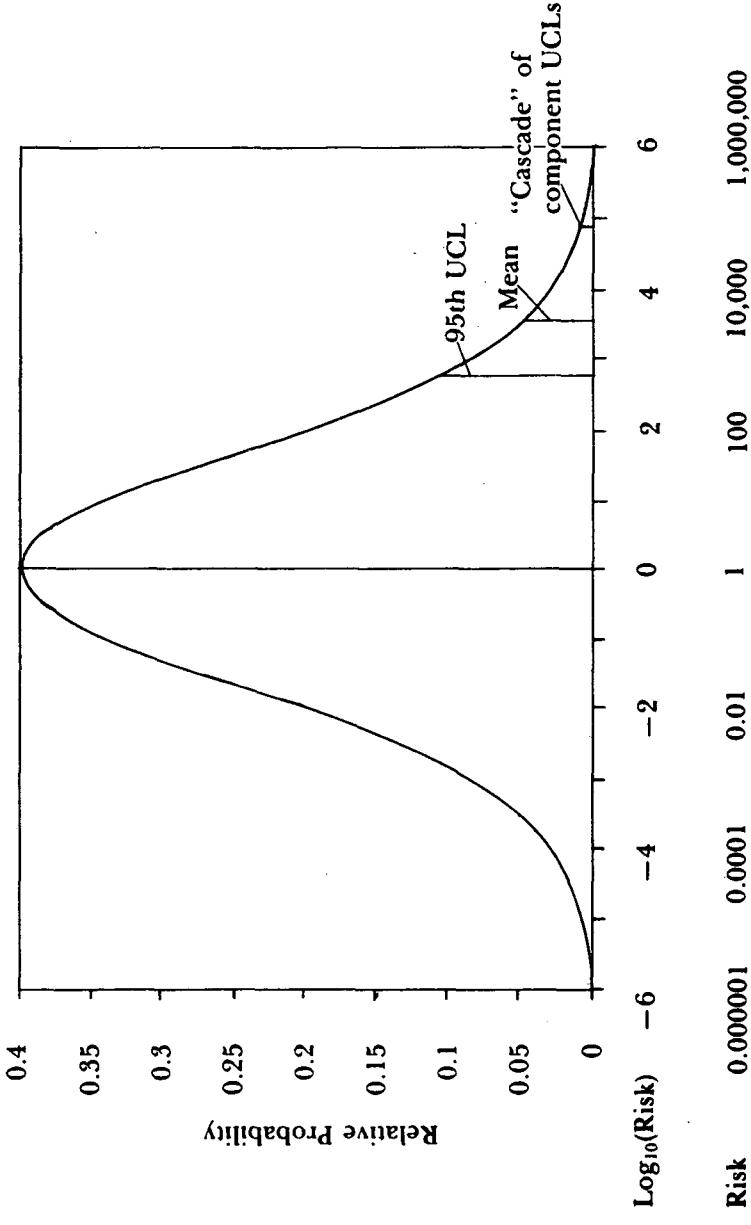


FIGURE 1B
PROBABILITY DISTRIBUTION FOR "RISK," THE PRODUCT OF THREE UNCERTAIN COMPONENTS



than the estimate one would derive by simply adding the excess risks from each exposure separately.[56] Thus, the compartmentalization that is central to QRA may be inherently nonconservative in predicting certain risks, given the multiplicity of exposures we all face (although by the same token, ignoring synergies could cause us to understate the health benefits of removing or reducing a particular exposure).

V. A Cascade of Truly Conservative Steps May Still Yield a Reasonable Estimate of Risk

One of the strongest potential arguments against conservatism is that it pervades the risk assessment process and that therefore our final risk estimates are ultra-conservative products of a cascade of exaggerations. Unfortunately, this valid conceptual point has itself been exaggerated by some of the revisionists. Their basic argument goes as follows: if a risk estimate is derived by stringing together N component estimates (*e.g.*, concentration times potency times number of exposed people, N=3), and each of these is a 95th-percentile conservative estimate, then the risk estimate is conservative to the $(1-(0.05)^N)$ degree.[57] If N=5, for instance, the risk would allegedly be a 99.99996th percentile estimate; because each component has only a 0.05 chance of being too low, the overall estimate has a $(0.05)^5$ chance (or about 1 in 3 million) of being an underestimate.

This is an appealing short-cut that borders on an exercise in sleight-of-hand. The figure of 1 in 3 million represents only a small subset of the possible combinations that could cause the risk estimate to be nonconservative—the ones in which all five of the components are "too low." Clearly, there is no *a priori* reason why any one of the five estimates couldn't be severely nonconservative; the remaining four could then in fact be overestimates and the output might still be "accurate" or even nonconservative. The proper way to address questions of cascading conservatism is to examine the uncertainty distribution of the risk estimate, which reflects the influence of all possible combinations of over- and underestimation in the components. For example, if risk is composed of three lognormally-distributed uncertainties (regardless

56. Selikoff, *Carcinogenic Risk Management in the United States*, in Management of Assessed Risk for Carcinogens 290 (1981).

57. Environ Corp., Alternative Approaches to the Conduct of Carcinogenicity Risk Assessment: A Proposal (July 26, 1987)(draft).

of how broad or narrow these distributions are), and one multiplies the three 95th percentile values together, one can show that the conservative risk estimate lies at the 99.8th percentile of the overall distribution, not the 99.99996th. Both of these percentiles admittedly reveal it is unlikely that the risk estimate is "too low," but the difference between the spurious underestimation probability of 1 in 3 million and the true probability of 1 in 500 (1−0.998) could greatly influence a decisionmaker concerned with possible errors caused by underestimation.

Arguably, the decisionmaker should also be concerned with the magnitude of misestimation as well as its probability, and thus should also evaluate how conservative the chosen estimator is with respect to the actual mean of the uncertainty distribution for risk. The relationship between the mean and any arbitrary UCL does depend on the breadth of the uncertainty distribution—when the uncertainty is a multiplicative "factor of X," for larger X it becomes more likely that the mean will approach or exceed any UCL.[58] Figures 1a and 1b show that if each of three components of risk has a "factor of 100" uncertainty around it, the risk estimate derived by "cascading" the three 95th percentiles together would indeed be much larger than the mean. However, the mean would in turn be much larger than the 95th percentile of the overall uncertainty distribution, so that choosing this latter value would be conservative probabilistically but quite nonconservative in terms of actual consequences. In this example, one could cascade three 90th percentile estimates and still barely be conservative with respect to the mean. Decisionmakers and the public should realize that while it is easy to ridicule a risk estimate for being exaggerated (*i.e.*, quite unlikely to be an underestimate), such estimates may be more reasonable with respect to the average of all possibilities than any less exaggerated ones.[59]

58. For example, suppose the "best estimate" of risk is 1 (death per year), but there is a "factor of 10" uncertainty surrounding this estimate, such that there is a 20 percent chance the true risk is 0.1 and a 20 percent chance it is actually 10.0 (leaving a 60 percent chance the "best estimate" is correct). The expected value or mean of this distribution is 2.62 [(0.2)(0.1)+(0.6)(1)+ (0.2)(10)], more than twice the median value of 1— the symmetry on the multiplicative scale is an asymmetry on the arithmetic scale. If the distribution were continuous instead of having only three possible values, the mean would probably lie at about the 80th percentile.

59. The "nonconservative" nature of probabilistically "conservative" estimates becomes even more pronounced if the uncertainties in the individual components are correlated with each other. For example, the error in estimating the average concentration of a chemical people are exposed to and the error in estimating the number of exposed per-

VI. Conservative Assumptions at Some Stages of Risk Quantification May Correct for the Omission of Other Stages Altogether

One controversial rationale for retaining assumptions one believes to be conservative comes from the recognition that one can generate an unreal output from a series of real inputs, if that series is incomplete. Engineers are familiar with this concept, as they often design structures and devices with margins of safety to protect against possible failure modes that have not yet been identified or quantified. QRA is a complicated multi-step undertaking, so it stands to reason that there might be unanticipated facets in the process that would increase our risk estimates, if only we took them into account. The revisionist would doubtless argue, with some justification, that to hedge against factors or processes we do not know exist smacks of a tautologous conservatism gone awry. However, there exist several tangible examples of factors we know can be important, yet which we omit from current practice because they are particularly difficult to quantify. Three of these factors in particular hardly qualify as fanciful contingencies, and would be likely to increase risk estimates if made part of the routine estimation process:

- *Accounting for "Indirect Exposures."* Until recently, risk assessments typically considered only direct exposures, such as inhalation of contaminated air, ingestion of contaminated water, or dermal absorption from direct contact with the substance of interest. Researchers are increasingly exploring the problems of indirect exposures, and in a number of cases have already found that these exposures can dominate the more obvious pathways. For example, the highest levels of certain volatile organic carcinogens to which a typical individual is exposed may well occur during showering and bathing.[60] If the individual's water supply is contaminated with a particular substance, his exposure may be greater due to inhaling the substance as it volatilizes from hot water than due to drinking the water directly (or inhaling the amounts typically found in indoor or outdoor air). Similarly, exposure to "dioxin" (TCDD) may be relatively greatest from eat-

sons may not be independent— "missing" an important indirect exposure pathway (see Section VI) may cause both estimates to be too low for the same reason.

60. Bogen & McKone, *Linking Indoor Air and Pharmacokinetic Models to Assess Tetrachloroethylene Risk*, 8 Risk Analysis 509 (1988).

ing vegetables that have accumulated the substance following airborne fallout from point sources (or from drinking the milk of cows fed on these plants), and relatively less from direct inhalation of TCDD-contaminated particulates or direct ingestion of TCDD-contaminated soil.[61]

- *Accounting for Changing Human Demography.* Often, risk assessments for point sources of pollution (such as abandoned hazardous waste sites) assume that the number of individuals currently exposed will remain constant during the time-horizon (usually seventy years) of the assessment. As Maxim, a risk assessment consultant, indicates, individuals may change their residences and places of work many times during a lifetime,[62] so this simplification may exaggerate exposures to particular individuals (although, as he acknowledges, it may be an accurate way to estimate cumulative exposure without regard to who bears the risk). In some instances, however, there may be a systematic increase over time in the number of exposed persons and/or the magnitude of their exposures. Today's aggregated exposures from a waste site located in a rural area may be minimal, but future exposures will not be if there is an influx of population or a changing pattern of local water use.

- *Accounting for Variations in Human Susceptibility to Carcinogens.* This factor, which current risk assessments ignore by assuming the point estimate of carcinogenic potency applies equally to all humans, may well be the most significant nonconservative assumption of all. This assumption flies in the face of a growing body of knowledge about several very common genetic, environmental, and lifestyle factors (*e.g.*, nutritional status, stress, concurrent diseases) that can cause interindividual variations in susceptibility[63] of several-fold, and about dozens of relatively rare conditions (*e.g.*, ataxia-telangiectasia, xeroderma pigmentosum) that can yield differences of thousands or millions.[64] Recent study of the effects on cancer risk of human variation in the

61. Travis & Hattemer-Frey, *Human Exposure to 2,3,7,8-TCDD*, 16 CHEMOSPHERE 2331-42 (1987).

62. Maxim, *supra* note 3, at 541.

63. An operational definition of "susceptibility" relates the dose of a carcinogenic stimulus to the risk at that dose. If person A faces a 1-in-100,000 excess risk of cancer from daily exposure to 100 mg of substance X, and person B has the same risk although exposed to only 1 mg/day, then B is 100 times more "susceptible" than A.

64. Finkel, *supra* note 22, at 5. Also, although these recessive genetic conditions are often rare, "heterozygotes" who carry only one copy of the defective gene (and therefore

amount and quality of metabolizing enzymes[65] or in the arrangement and control of oncogene sequences within cellular DNA[66] suggests that even among "normal" people, the risks of cancer initiation due to an identical exposure to an identical stimulus may vary quite substantially. Yet most of our risk estimates are based on the responses observed in small groups of laboratory animals, genetically and environmentally manipulated to be as homogenous as possible in their sensitivity.

Currently, science cannot provide a definitive characterization of how susceptibilities are distributed across the human population, nor of how susceptible the average human is relative to the inbred rats or mice used in testing. The first knowledge gap is yielding to some empirical inquiry, however. A preliminary study using epidemiologic data on lung and colorectal cancer[67] concluded it would be plausible (and not particularly conservative) to describe human susceptibilities to carcinogenesis as lognormally distributed, such that five percent of the population is about twenty-five times more (and five percent twenty-five times less) susceptible than the average person, and one percent each about one hundred times more or less susceptible. This tentative conclusion was supported by another study which instead examined clinical data on human variation in some of the biological processes involved in carcinogenesis.[68]

Figure 2 shows the effect on QRA of this amount of interindividual variability. Each curve shows the ratio of an individual's true risk to the "assessed risk" (the point estimate calculated under the existing assumption that all humans are alike), as one examines the more susceptible fifty percent of the human population. The different curves could represent three different assumptions about how sensitive rodents are with respect to humans. The curve on the left assumes that rodents are more

do not manifest symptoms of the disease) may be at elevated risk, and may constitute significant fractions of the total population.

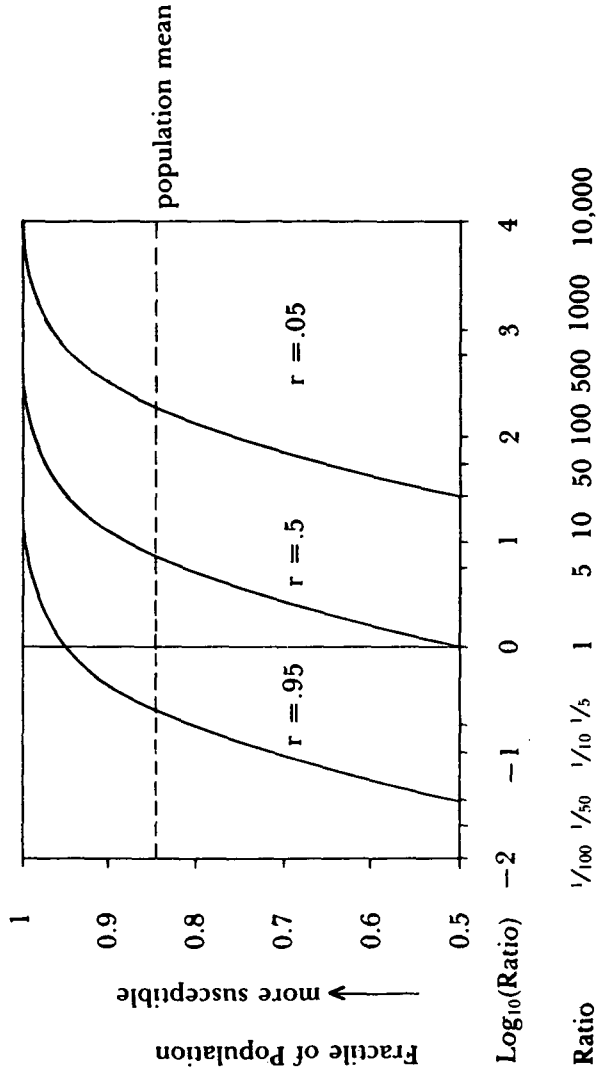
65. Cartwright, *The Role of N-Acetyltransferase Phenotypes in Bladder Carcinogenesis: A Pharmacokinetic Epidemiological Approach to Bladder Cancer*, 2 THE LANCET 842-46 (1982).

66. Krontiris, *Unique Allelic Restriction Fragments of the Human Ha-Ras Locus in Leukocyte and Tumour DNAs of Cancer Patients*, 313 NATURE 369 (1985).

67. Finkel, *supra* note 22.

68. Hattis, Erdreich & Dimauro, Human Variability in Parameters Potentially Relevant to Susceptibility to Carcinogenesis (1986) (report to the Environmental Criteria and Assessment Office, U.S. EPA, MIT Center for Technology, Policy, and Industrial Development, Report #86-4).

FIGURE 2
RATIO OF INDIVIDUAL'S RISK TO "ASSESSED RISK"



sensitive than all but five percent of humans (a nonconservative assumption in the absence of empirical information, one that suggests our current procedures used to extrapolate animal data are quite conservative); the curve on the right assumes all but five percent of humans are more sensitive than rodents. So even if the curve on the left applies, the true risks may be four or more times *higher* than advertised for one percent of the population; if the assumptions are neutral, as in the middle curve, the advertised risks may be a factor of fifty or more too low for 2-1/2 percent of persons.[69] Note also that the mean of all risks falls at the 84th percentile of this putative susceptibility distribution, so that on average, even very conservative procedures yield estimates within a factor of four of the right answer, and neutral procedures will be nonconservative by more than a factor of seven.[70]

The point of this section is to suggest that conservative assumptions in the existing stages of QRA may (deliberately or fortuitously) substitute for the nonconservative omission of other factors. Thus, even if we knew a particular revision would add more reality in one or more of the existing stages, that action could, ironically, make the overall risk estimate less real. It is certainly intellectually unsatisfying to arrive at a correct (or appropriately prudent) answer through a series of flawed calculations (some too high, others too low by virtue of being ignored). But the current focus of many revisionists seems to be to keep the existing and perhaps incomplete framework while scaling back some of the procedures specific to it—that is, rectifying the errors of comission in QRA, not those of omission. In the best of all worlds, as much scientific progress would be made in areas such

69. Another way to look at these assumptions is to put aside the interspecies extrapolation argument and simply note that the point estimate of potency derived from rodent data should be "correct" (*i.e.*, represent the true probability of cancer at a given dose) for some unknown percentile of the human distribution. The left-hand curve just assumes that we have ended up being conservative, such that the estimate is "correct" for the 95th percentile human (that is, it is conservative for most people, nonconservative only for the very susceptible).

70. As suggested in Section III, it is not necessarily true that the use of the 95th percentile UCL in the curve-fitting exercise yields an estimate of potency correct for the 95th percentile human. The former involves an admission of statistical uncertainty; the latter involves the unknown biological relationship between the average susceptibility of the two species. In addition, this analysis could also apply even if the risk estimates were based on human data, because these data generally come from small groups of workers who may be systematically "healthier" than the general population, or at least not exhibit as much variability as the population as a whole.

as human susceptibility and total human exposure as is being made in areas such as rodent metabolism and tumor promotion. This asymmetry is due in part to different amounts of resources applied, and in part to the relative intractability of some of the omitted factors. In my opinion, the burden of proof ought to be on the proponents of revision to suggest that a particular refinement of good science not only improves the reality content of the stage of QRA to which it applies, but improves (or at least does not decrease) the reality of the entire potency or exposure characterization process.

VII. MORE SCIENCE DOES NOT ALWAYS MEAN LESS RISK, NOR DOES IT ALWAYS "REDUCE UNCERTAINTY"

Some critics of risk assessment seem to believe that whenever the "new and improved" science comes into play, lower estimates of potency and risk naturally follow.[71] This is in part a self-fulfilling prophecy, and in part a *post hoc* rationalization. Certainly, there exist cases where new data or new biologically-based observations have provided convincing evidence that particular risk estimates were overly conservative. However, Paustenbach and Anderson paint a picture suggesting that the march of science leads QRA exclusively in one direction, and perhaps suggesting that only resource constraints are holding EPA and the other agencies back from lowering many more estimates that are ripe for reduction. There are several deficiencies with such a characterization:

- *The Available Examples May Constitute a Biased Sample.* The federal regulatory agencies have relatively few resources available to gather more data on substances whose risks they have already assessed. Instead, they rely primarily on academic or industry researchers to propose new models; for new chemical-specific data, they rely primarily on studies conducted or commissioned by corporations or industry groups. In many cases, researchers are obligated to report their findings, whatever their implications, as when testing reveals an unexpected hazard of a substance. However, in other instances, researchers are selective in the research avenues they pursue, the findings they report, or the intensity with which they advocate the application of their work. Epidemiologists, for example, may find it difficult to publish inconclusive

71. Anderson, *supra* note 1; Maxim, *supra* note 3, at 526.

or negative results, and some have suggested they tend (consciously or otherwise) to select situations for study where they believe a statistically significant problem is likely to be found. By the same token, the brief history of the reanalysis of bioassays and exposure studies is perhaps affected by the selective choice and advocacy of research that is likely to yield favorable results.

- *Not All of the Available Examples Support Risk Lowering.* In addition to the examples in Section VI of "new" science that might lead to increases in risk estimates, even the particular scientific refinements supported most heavily by the revisionists can reveal possible nonconservatism when applied to some substances. This aspect of the "new science" is often downplayed by its proponents, however. As an example showing how one cannot prejudge the implications of more detailed biological information, Swenberg and his colleagues at the Chemical Industry Institute of Toxicology recently studied how the shape of dose-response curves for three chemicals would change if data on the prevalence of DNA adducts (chemically modified units of DNA, associated with exposure) were used as a measure of response, instead of merely counting the number of tumors in each group of experimental animals. In one case, the dose-response curve incorporating the new science was essentially unchanged (still linear in character), in one case the new curve was steeper at high doses than at low ones (which might support a lowering of risk at doses to which humans are exposed), and in the third case the new curve was steeper at low doses (which might suggest that raising the old estimate of potency would be appropriate).

- *Even Among the Reevaluations That Have Resulted in Lower Risk Estimates, Some Questionable Interpretation of Data Has Occurred.* EPA's 1987 decision to reduce the official estimate of the potency of methylene chloride ($MeCl_2$) by a factor of 8.8 is a good example of how new and subtle value judgments can accompany new science and influence its conclusions.[72] The new potency estimate emerged from application of a pharmacokinetic (PK) model developed by chemical industry researchers, a complicated mathematical formulation for quantifying how each species (in this case, mouse and human) metabolizes and distributes a given dose of a chemical throughout body tissues. To use this model, nu-

72. EPA, *supra* note 8, at 7-13. The industry researchers argued that their results supported a reduction of 144-fold; EPA declined to accept all of the assumptions underlying this conclusion.

merical values of forty-six biological parameters (twenty-three in each species) must be specified. These parameters include the weight of target organs (lung, liver, etc.), the flow of blood through different tissues, the affinity of the chemical for blood, air, and fat, and the rate at which different enzymes are produced and react with the chemical. Uncertainty surrounds all of these numbers—some rodent parameters are hard to measure, some human parameters cannot be measured, but must be inferred from the rodent numbers, and most of the human parameters vary to an unknown extent across our diverse population. Yet the 8.8-fold reduction emerged from the assumption that all forty-six values could be represented exactly by "best" estimates, which Anderson says are "actual physiological parameters." Thus, QRA becomes more complicated (which the revisionist would applaud according to the "more is better" credo), yet the additional complication of acknowledging the uncertainty in these new parameters, or representing them by conservative estimates, drops out of the process. Two scientists from the National Institute of Environmental Health Sciences recently reanalyzed the $MeCl_2$ data using a plausible range of values for each of the forty-six parameters to reflect their uncertainty and variability, and concluded there was more than a five percent chance that the new PK model, correctly interpreted, actually reveals that an increase in the potency of $MeCl_2$ is warranted.[73]

Similarly, Anderson advocates a subtle shift away from prudence in interpreting models when she suggests that the potency of benzo[a]pyrene could be lowered by substituting a "two-hit" for a linear model.[74] Her justification is that the former model appears to fit the data better, but she does not discuss how confident we can be in that interpretation. It is quite possible that a linear model might fit almost as well, or even fit better if the data

73. Portier & Kaplan, *The Variability of Safe Dose Estimates When Using Complicated Models of the Carcinogenic Process. A Case Study: Methylene Chloride*, in FUNDAMENTAL AND APPLIED TOXICOLOGY (1989) (in press). It is important to note that their conclusion does not follow from a cascade of conservative assumptions. Rather, each parameter was randomly assigned a value from its plausible range, and the PK model was run and a potency estimate derived. This process was repeated 1000 times, and the potency estimates arranged from lowest to highest; the 50th-highest value, which is an estimate of the 95th percentile upper-bound on potency, was higher than EPA's original estimate (and more than 10 times higher than the revised estimate).

74. Anderson, *supra* note 1, at Section II(B).

were analyzed with allowance for the random variation inherent to the small sample of test animals.

Perhaps part of the reason that advocates of new models do not take pains to explore the uncertainty in their application is that common sense tells us that more information always makes us less uncertain. Except when the direct costs (or indirect costs due to delay) of research are disproportionately large, I wholeheartedly endorse investigations into new models and the collection of new data, and agree that risk estimates incorporating this information ought to make QRA more precise. Unfortunately, in the case of QRA we often find that new information raises new questions and imposes new demands for detail in interpretation, even as it provides new answers. More technically, one could say that while new science cannot increase uncertainty, it assuredly can reveal (in hindsight) that we knew less about the situation than we thought. Users of conservative procedures can avoid confronting uncertainty directly because these procedures are designed to be cautious in the face of incomplete knowledge. Advocates of more complicated models that are capable of generating more precise estimates ought therefore to be especially careful that they verify the degree of precision, and that they do not confuse precision with accuracy.[75] Anderson's closing statement ("the more precise the risk assessment, the easier it is for difficult decisions to be made") is a cogent description of what decisionmakers may believe and desire. The truth, however, is that precise estimates that appear accurate (because they do not account for uncertainty) also make it easier to make difficult and wrong decisions.

VIII. Assessing the Costs of Compliance

A fair critique of conservatism in risk assessment would acknowledge that in some cases, risk estimates themselves do not lead to any concrete action on the part of the regulator or the regulated, unless the costs of such actions are also assessed and deemed appropriate and acceptable. As Anderson points out, if the theoretical risk is lowered, then the impetus for remedial action may be considerably decreased, presumably because the action would no longer be worth the cost. The revisionists, however, rarely explore the possibility that the costs of complying

75. An archer shooting at a target would be precise and yet not accurate if all her arrows landed within centimeters of each other, but none was anywhere near the bullseye.

with risk reduction measures are themselves systematically biased. While no detailed comparison exists of the *ex ante* and *ex post* estimates of compliance expenditures, examples showing marked overestimation come easily to mind. Perhaps the most frequently-cited example concerns VC. In 1974, an industry study claimed that to meet OSHA's proposed one part per million standard for VC concentrations in the workplace, companies making polyvinyl chloride (PVC) plastics would either have to raise prices by about eighty percent or shut down entirely. By April 1975, however (after a federal appeals court had upheld OSHA's proposed rule), the major PVC producers were estimating price rises of only about five percent—and by late 1976, the boom in PVC sales and apparent lack of any price increase attributable to compliance caused an industry trade magazine to run a headline entitled "PVC Rolls out of Jeopardy, Into Jubilation."[76] Ashford,[77] a professor of environmental policy at the Massachusetts Institute of Technology, has suggested that compliance expenditure estimates generally fail to take into account three factors that could reduce these estimates: (1) economies of scale that arise as demand for control technology increases; (2) economies due to movement "up the learning curve" of efficient compliance; and (3) technological innovation that yields entirely new techniques for compliance. On the other hand, all parties should bear in mind that compliance expenditures are only one of the inputs to the total social cost of risk reduction measures. Other costs, which may be positive or negative, include employment effects and opportunity costs.

In addition, there occasionally may exist an asymmetry in the way regulatory agencies think about control costs versus risks. Frequently, agencies choose not to act to reduce risks because the individual probabilities of harm are so small as to be *de minimis*, trifling additions to the "ocean of risk" each of us faces daily.[78] Sometimes, exposures that pose *de minimis* risks are sufficiently widespread that the total possible increase in mortality would be tangible, even though no individual would face a palpable or "unacceptable" risk.

76. *PVC Rolls Out of Jeopardy, Into Jubilation*, CHEMICAL WEEK, Sept. 15, 1976, at 34.

77. Ashford, *Alternatives to Cost-Benefit Analysis in Regulatory Decisions*, in MANAGEMENT OF ASSESSED RISK FOR CARCINOGENS 129, 130 (1981).

78. Travis & Hattemer-Frey, *Determining an Acceptable Level of Risk*, 22(8) ENVTL. SCI. & TECH. 873 (1988).

The appealing logic which removes risks from consideration when they are trivially small is often *not* followed when assessing the other side of the risk-cost tradeoff. We live in an "ocean of costs" as well, and yet trivial effects on consumer prices are often treated as full-fledged "costs of control" and aggregated across all affected individuals without reference to the *de minimis* notion. For example, industry argued in the early 1980s that compliance with proposed benzene controls would cause a two percent rise in the consumer price of gasoline. Viewed as a macroeconomic perturbation potentially adding billions of dollars to the national "gas bill," such a cost might not appear justified. Viewed as a collection of *de minimis* changes in the budgets of affected consumers (perhaps on the order of ten dollars per person per year), the costs might be worth incurring, even to reduce small individual risks.

IX. The Pooling of Data Versus the Averaging of Irreconcilable Theories or Results

Central to the thinking of some revisionists is the advocacy of a weight-of-evidence approach to supplant conservative procedures. They characterize conservatism as a rudimentary weight-of-evidence approach, wherein the single most foreboding piece of evidence gets all the weight regardless of the amount or quality of moderating or contradictory data. Paustenbach's reasoning that "when the conclusions reached from high-quality data are overwhelming, spurious data must be de-emphasized or discarded"[79] is quite persuasive, and examples where this maxim is not followed constitute a legitimate indictment against QRA as currently practiced.

The more important and vexing question is what to do in the more frequent situations where data or models conflict, and yet the contrast in quality, validity, or credibility is not overwhelming. In extending Paustenbach's argument to the more general situation, I believe the revisionists may distort the uses and conclusions of the weight-of-evidence concept. The basic pitfall involves mingling and confusing the distinct ideas of ambiguity and the policy response to it. In an example detached from policy controversy, most of us would recognize that being offered half a chance (perhaps a coin flip) to win a loaf of bread is not the same

79. Paustenbach, *supra* note 2, at Section IV(A).

as being offered the gift of half a loaf. Yet the "expected value" approach spelled out by Nichols and Zeckhauser,[80] economists from the Kennedy School of Government at Harvard University, Sielken,[81] a statistician, and others equates these two states. In an example adapted from Nichols and Zeckhauser, if current QRA procedures predicted that the cancer risk of exposure to X units of a chemical was 10^{-3}, and an alternative set of procedures (perhaps using a dose-response model with a threshold) predicted zero risk at this dose, regulators might decide the "true" risk was somewhere between zero and 10^{-3}. If they deemed the models equally likely to be correct, the "official" risk estimate would then fall midway between the two values (or 5×10^{-4}); if the threshold model was instead deemed three times more likely to be true, the official estimate would become the weighted average value of 2.5×10^{-4}. In other words, they believe we can ascertain and compensate for the "degree of *conservatism*" (an issue involving the magnitude of error) by factoring in the degree of *belief* that a lower estimate is correct (an issue involving the probability of error).

I acknowledge that we should consider changing some of today's default procedures to incorporate via these averaging techniques certain high-quality data we have hitherto excluded. Such refinements could remove some of the research disincentive discussed earlier, as well as improve the plausibility of some of the component steps of QRA. In particular, I support efforts to: (1) use more of the available rodent bioassay data via the pooling of experiments or the averaging of potency estimates obtained from comparable studies; (2) relax our preoccupation with the "maximally exposed individual," so that information about the range of human exposures (as in Paustenbach's Table 1) is not lost; and (3) incorporate "reality checks" into modeling efforts, as Maxim[82] and others have suggested. In effect, invoking physical or mathematical principles to rule out certain extreme combinations of hypothetical model results allows the assessor to legitimately reduce the level of conservatism via high-quality data to the contrary.

None of these refinements invites the misleading, illogical, and potentially manipulative conclusions that can result from an indiscriminate advocacy of averaging as a way to combat conservatism.

80. Nichols & Zeckhauser, *supra* note 4, at 72-76.
81. Sielken, *supra* note 3.
82. Maxim, *supra* note 3, at 546.

Fundamental scientific controversies and the search for scientific truth simply do not lend themselves to resolution by "the average of right and wrong answers."[83] Apparently based on his belief that cancer promoters "probably" have "practical threshold[s],"[84] Paustenbach argues that we should lower potency estimates for a particular substance that is "probably" a promoter. Perhaps it should be lowered substantially, by accepting the threshold model unquestioningly, or perhaps lowered to a lesser extent by averaging the conclusions of the default and alternative procedures. Since the former method is generally imprudent in the face of scientific uncertainty, revisionists often support the latter approach as a more palatable compromise position. In an article[85] and testimony before the EPA Science Advisory Board (SAB),[86] I recently suggested four reasons why averaging is generally an untenable policy:

- It is highly sensitive to the very subjective assignment of weights to different theories or results. I find it hard to imagine a procedure more dependent on "policy choices masquerading as scientific facts" than one which presents point estimates of risk that hinge on the (presumably unpublished) opinions of the assessors or reviewers as to whether the new science is as likely, half as likely, or ten times more likely to be true.
- The average often represents an impossible or nonsensical state of nature. Obviously, there is no such thing as a chemical with "half a threshold," although in special cases it may be efficient (over the long run) to act as if such chimeras could exist.[87] The nonsensical aspect of averaging should always promote caution—recall King Solomon's wisdom in demonstrating that one obvious compromise over the custody of a baby was worse than either alternative.
- The average may exist but be irrelevant. The parable of a person remaining with his feet in a fireplace and his head in a freezer (because his average body temperature was comfortable) is an instructive analogy to cases where uncertainty over risk

83. Silbergeld, *Agency Drops Science for a Wishful Finger-in-the-Wind*, L.A. Times, Jan. 18, 1988.

84. Paustenbach, *supra* note 2, at Section IV(B).

85. Finkel, *Dioxin: Are We Safer Now Than Before?*, 8 RISK ANALYSIS 161 (1988).

86. A. Finkel, Comments on EPA's Proposed Revision of the Carcinogenic Potency of TCDD (Nov. 29, 1988)(presented to the Science Advisory Board Review Panel on Dioxin).

87. Finkel, *supra* note 85, at 164.

could be obscured by averaging. For the reasons elaborated in Section I, citizens (and hence decisionmakers) may desire to know whether they are being asked to bear (or to pay to remove) "half a risk" or "half a chance at double the risk."

- By putting a premium on ambiguity, the averaging process encourages interested parties to make incomplete arguments, promotes strategic posturing, and sets a precedent whereby agencies may have to repeatedly expend resources examining or refuting new theories and models (or revived versions of dormant ones).

Contrary to Anderson's interpretation,[88] EPA's tentative decision to downgrade the potency estimate of TCDD by a factor of seventeen was not based on the application of a new mathematical model to estimate the potency of a promoter. Even if EPA had used this rationale, it could have been criticized on the grounds that the biological parameters of this model are at least as uncertain and variable as those needed to use PK models, but at least this would have been a defensible scientific "judgment call" that the time had come to abandon the default procedures for this particular chemical. Instead, EPA's first rationale (in its November 1987 draft proposal) justified the new estimate as "[n]ear the mid-point of a range" bounded by the predictions of linear and threshold models of dioxin's carcinogenicity.[89] Both that draft and the latest revision[90] state explicitly that the two-stage or promoter model is too scientifically premature and statistically variable to use at this time.

The June 1988 proposal retained the same seventeen-fold decrease, but justified it as "consistent with" three "default" estimates produced by different United States federal agencies, instead of using the (European) threshold-based estimates as inputs to the averaging procedure. At this writing, an SAB expert panel is drafting a report to EPA which reportedly will conclude that there is no clear scientific basis for moving the estimate of potency downwards (*i.e.*, that the weight-of-evidence also includes some indications that the original estimate might be appropriate or not conservative enough). At present, the dioxin case is there-

88. Anderson, *supra* note 1, at Section II(B).

89. E.P.A. Dioxin Task Force, A Cancer Risk-Specific Dose Estimate for 2, 3, 7, 8-TCDD 44 (1987).

90. E.P.A. Dioxin Task Force, A Cancer Risk-Specific Dose Estimate for 2, 3, 7, 8-TCDD 44-45 (June 1988) (EPA/600/6-88/007(A)(a).

fore a study of how the presence of fundamental scientific ambiguity does not necessarily lead to the partial or selective abandonment of conservative procedures, despite the allure of arguments that call for hedging towards lower risk estimates until the controversies are resolved.

X. Conclusion And Recommendations

Anderson, Paustenbach, and I all agree that the craft and science of QRA must continue to evolve. They are candid that their frustration with current procedures reflects their dual roles as scientists and as defenders of (or advisors to) corporations. As a scientist, I too am frustrated by the formulaic nature of some of the "conservative" assumptions, and applaud Anderson and Paustenbach for helping to push three items in particular to the top of the national QRA research agenda: (1) the need to glean more information from rodent bioassays, particularly as regards PK data that can help explain why substances do not seem to affect certain strains and sexes, instead of focusing only on the observed tumor response and on the most affected animal subgroup; (2) the need to elucidate the distribution of human exposures in all its richness, instead of focusing all our energy on how high exposure might possibly be for a single (perhaps hypothetical) maximally exposed individual; and (3) the need to recognize that animal tests at high doses are not always predictive for humans at low doses, that the qualitative issue of whether a substance is a human carcinogen can on a case-by-case basis sometimes yield to detailed scrutiny using techniques of toxicology, biophysics, and (increasingly) molecular biochemistry. I am certain, however, that reasonable scientific experts also have cause for frustration with some of the formulaic and nonconservative aspects of QRA, particularly the inattention to issues of human susceptibility, of total human exposure through a complex web of pathways in the outdoor and indoor environments, and of synergy among exposures.

Therefore, purely with reference to the few scientific truths and observational evidence at our disposal, the case for wholesale abandonment of conservatism is tenuous at best. If this entire debate were just about objective evidence, I am sure that each of us could write another article composed solely of particular examples and counter-examples, and the stalemate would continue. We could each recount descriptions of how hindsight has re-

vealed "false negatives" or "false positives" with regard to carcinogenicity or cancer potency, and revealed gross overestimation or underestimation of exposures.

I could cite in detail the work of Allen, Crump, and Shipp,[91] who have undertaken the only definitive comparison of the few cases (some twenty-three substances) where human epidemiologic data can be used to perform a *post hoc* "reality check" on how well animal data (as currently analyzed and interpreted) can predict cancer potency in humans. They concluded there is no obvious merit to the claim that bioassay-based potency estimates systematically exaggerate human risks—on average the results were quite comparable, although there were "outliers" in both directions. The revisionists would probably counter that this comparison is unfair because it excludes all rodent-positive chemicals for which no evidence of human carcinogenicity exists. At this point, the traditional impasse would occur, with each "side" trying to exploit the unavoidable asymmetry in evidence. The revisionist would use phraseology like "the handful of chemicals which are known to be initiators or mutagens," while the defender of conservatism would counter that there are few rodent carcinogens we "know" not to be initiators (or human carcinogens). Whenever evidence is incomplete and the "glass" is either "half empty" or "half full," "facts" can follow conclusions rather than vice versa; Anderson uses the finding that only one percent of a PCB dose reaches the livers of rodents to argue for "dramatically lower estimates of . . . risk,"[92] whereas a different observer could easily draw the exact opposite conclusion (what if humans distribute more than one percent to their livers?).

Thus, despite the desires of the revisionists to make this a scientific debate, or one pitting "science" against "values," the future of conservatism will largely depend on the clash between divergent ideologies and value judgments. The crux of this ideological difference concerns how one wishes to respond to uncertainty. The stereotypical conservative may tend to be paralyzed by uncertainty, or to exploit it to prevent regulators from making (or even considering) difficult trade-offs between societal risk and social cost. The caricature of a revisionist, on the other hand, is someone who is indifferent to uncertainty in the ways described in

91. Allen, Crump & Shipp, *Correlation Between Carcinogenic Potency of Chemicals in Animals and Humans*, 8 RISK ANALYSIS 531 (1988).

92. Anderson, *supra* note 1, at Section II(C).

this article, and who allows overconfidence in data or theories to dictate precisely the preconceived outcomes of these tradeoffs. Both extremes can lead to profoundly undesirable results—overconfidence and indifference to uncertainty may have contributed to the Challenger disaster,[93] yet without some willingness to consider scenarios other than the worst case, the twenty-four previous Space Shuttle successes might never have been allowed to occur.

I advocate a new synthesis of these polar positions, which may allow QRA to evolve in full awareness of uncertainty. Techniques now exist to quantify uncertainties and variabilities in carcinogenic potency,[94] ambient concentrations,[95] human uptake,[96] and population demographics, and to combine these uncertainties into an uncertainty estimate for risk. Although each of these techniques need more refinement and more "operating experience," they can provide decisionmakers with hitherto-unavailable information on the probabilities and magnitudes of errors of under- and overestimation of risk. A quantitative uncertainty estimate, in the form of a probability density function like those in Figures 1a and 1b, can replace both conservative and "real" point estimates of risk. No longer would we have to argue about whether the numbers are too high or too low, but we could focus on how conservative any particular choice would be, and select an appropriate level of conservatism based on the probability, magnitude, and social cost of potential misestimation in either direction. These uncertainty estimates would also provide the public with an honest and evolving appraisal of the possible range of risk values, and could finally shed light on the ranking problem identified by Nichols and Zeckhauser. One can show that errors in ranking different risks (or different outcomes of controlling a single risk) are endemic even if one has eliminated asymmetric conservatism by reducing each risk estimate to a central value or some other summary measure.[97]

93. Freudenburg, *Perceived Risk, Real Risk: Social Science and the Art of Probabilistic Risk Assessment* 242 SCIENCE 44 (1988).

94. Finkel, *supra* note 44.

95. Freeman, *A Method for Propagating Measurement Uncertainties Through Dispersion Models* 36 J. AIR POLLUTION CONTROL A. 246 (1986).

96. McKone & Ryan, *Human Exposures to Chemicals Through Food Chains: An Uncertainty Analysis*, ENVT'L SCI. & TECH. (1989) (in press).

97. Finkel, *supra* note 20, at 81-84. This report suggests that the risk analyst provide the decisionmaker with an estimate of the uncertainty in the *ratio* of the two risks or outcomes.

I pretend no clairvoyance on how attention to uncertainty and variability will affect risk estimates—but I suspect that it will reveal some of our existing estimates were unduly conservative, while others barely contained any margin of safety or actually were underestimates. I am confident, however, that this measured approach to risk and uncertainty would not lead to the kind of "swing of the pendulum" some revisionists appear to advocate. I think it instructive to look, as Henrion and Fischhoff have,[98] at the history of overconfidence in scientific estimates. Even in trying to estimate measurable, unvarying physical constants such as the speed of light, the historical record shows a distinct tendency for each new estimate to fall *outside* the confidence bounds scientists placed on the existing estimate. In many cases, each new surprise also tended to oscillate above and below what we now believe to be the accurate value. Such overconfidence and overcompensation is merely embarrassing when the uncertain quantity is a physical constant—but the same pattern applied to risk assessment could be tangibly harmful to the public.

In conclusion, while I am reluctant to believe that good science is "a subterfuge designed to accomplish de facto deregulation,"[99] neither do I readily accept that there is "a dark side to conservatism."[100] I look forward to the debate over values continuing, although I am troubled that in practice this debate makes us focus on reassessing the risks of a few dozen chemicals (where scientific information is relatively abundant and the stakes for industry are relatively high), and diverts resources away from EPA's initial evaluation of the tens of thousands of other chemicals in commerce. In a time of budgetary austerity, our common desire for more science might have to give way to more practical considerations. I feel it is wholly appropriate for the revisionists to remind the public of the many success stories of environmental health risk reduction, to emphasize that the United States population enjoys a relatively safe environment and food supply, to place in

Even if one risk is deemed worse than another by all available summary measures (mean, MLE, UCL, etc.), this exercise may reveal a significant probability that the true rank order is reversed. The prioritization decision, like any control decision, then ideally becomes a question of balancing the likelihoods and costs of making the wrong choice.

98. Henrion & Fischhoff, *Assessing Uncertainty in Physical Constants*, 54 AM. J. PHYSICS 791 (1986).

99. Latin, *Good Science, Bad Regulation, and Toxic Risk Assessment*, 5 YALE J. ON REG. 94 (1988).

100. Maxim, *supra* note 3, at 555.

numerical or societal perspective the many relatively large risks we now accept individually or collectively, and to stress how costly it may be (to individual firms or to the national economy) to reduce certain risks. As a decision theorist and public health professional, however, I do object to attempts to trivialize uncertain risks so as to alter the balance between risk reduction and cost, if such attempts are driven by personal value judgments, accomplished by selective use of incomplete evidence, and masquerade as neutral and objective scientific progress.