

From: Chemical Risk Assessment and Occupational Health: Current Applications, Limitations, and Future Prospects, Edited by C. Mark Smith, David C. Christiani, and Karl T. Kelsey; Foreword by J. Donald Millar. Westport, CT: Auburn House, 1994. pp. [181]-186.

17.

Limits of Epidemiological Evidence and Uncertainty Analysis in Occupational Risk Assessment

James M. Robins

I haven't been directly involved in risk assessment over the last few years, partly out of frustration with the exercise. Thus, I see myself now as being more of a commentator from the outside than from inside of the process. Although many others have claimed that the risk assessment emperor has no clothes, I will be echoing some of these claims from my own perspective. Even though most of my research now is highly technical, I'm going to take a much more sociological point of view as to why we do risk assessments and how various factors determine what industry, government, and academics do in this area. I think the fundamental problem is that risk assessment methodologies are elaborate mouse-traps that don't catch mice (risks)—at least not the mice (risks) that one can see (quantify or measure). I believe this is true especially for environmental risk assessment and, to a lesser degree, for occupational risk assessments as well.

Risk managers (e.g., at the Environmental Protection Agency [EPA] and the Occupational Safety and Health Administration [OSHA]) need to have ways to deal with potential chemical risks. Considerable pressure has been exerted on these managers to develop analytical approaches to assessing such risks. The fact is that there are a lot of chemical exposures occurring out there; there are a lot of decisions that must be made about them, so people need *rules* or at least a process to guide their regulatory efforts. Thus, a variety of risk assessment rules have been established over the last 10 or 15 years, and many of these rules have evolved more as a response to political pressure than to the science involved. Many people, from both sides of the spectrum, don't like these rules. These include people who are worried about risks and those more worried about economic profit.

As we all know, it's become pretty easy to poke holes in these rules. Both sides point out that the risk assessment mousetraps that have been built aren't catching any mice that we can see. Many of these rules were just created, ad hoc, to get something done, to provide some guidance, or to establish a procedural framework from which to make decisions. There have been many ideas on how to improve these mousetraps. These have included suggestions for building better exposure and dose response models; conducting sensitivity analyses; or replacing the present EPA policy of considering the worst possible exposure scenario by measuring instead the actual distribution of exposure in *real communities* and averaging, by Monte Carlo techniques, over this distribution; and many others. What's interesting is that both sides of this debate have recently taken to the idea of *uncertainty analysis*, often relying on Bayesian statistics, which proffer risk profiles rather than risk estimates.

THE BAYESIAN APPROACH

What is the idea behind these efforts to quantify uncertainty in risk assessment? Bayesian statistics is a set of ideas that says no matter how uncertain you are about something—you could be ready to throw up your hands in despair—all you have to do is say how uncertain you are and what the costs and benefits of being right versus wrong are, and the machinery of Bayesian statistics will tell you what to do: a rule in the face of near total ignorance! What a system! It sounds scientific—and there are computers that will compute it! Unbelievable.

There are a lot of sincere people, frustrated and embarrassed over the current risk assessment system, that have come together to try uncertainty analysis. With uncertainty analysis, rather than ignore the uncertainties, the risk assessor now tries to take them all into account. If one doesn't know the right dose response and exposure model, he can say with a certain probability it is this; with a certain probability, it is that.

What are the proper magnitudes for these probabilities? We don't know? So we'll get a bunch of *experts* together and persuade them to give us a set of probabilities, and we'll put all this in an uncertainty analysis and, eventually—if we get a good Bayesian statistician—calculate a risk distribution. So we're building a better mousetrap? Maybe, but we don't have any more real information, do we? So we're still not catching mice (risks) we can see (quantify). But it's hard to say we're not improving things; I mean, we're taking all this uncertainty into account. We are trying to be honest. But no one can show us a single mouse we've caught because we don't have the science that can count the trapped mice.

So, are these Bayesian ideas good? A famous Bayesian, Jimmy Savage, said, "You can't enjoy the Bayesian omelet without breaking the Bayesian eggs." Now, what did he mean? He meant that if you have a lot of uncertainty and you want to know what to do, you aren't going to be able to figure out what to do in the face of uncertainty unless you are able to establish probability

distributions quantifying how uncertain you are about different things. If you are willing to establish these distributions, then the Bayesian has an algorithm for you to decide what to do. So when you don't know whether something is a chemical carcinogen, instead of saying you don't know, you estimate (guess?) that with a 12 percent probability there is a relative risk of 3 to 3.5; with a 14 percent probability, there's a relative risk of 2.5-3.0; and so on. That's breaking the Bayesian egg; you've now expressed uncertainty in a quantitative way. You haven't learned anything new about the world—you've just introspected yourself about how uncertain you are, and can now do a computation and enjoy the Bayesian omelet; that is, you get a rule for anything you want to decide.

Fortunately, there's an idea in Bayesian statistics called *calibration*. Essentially, you take your prior beliefs about the world, act on them, and then check to see if you are calibrated by simply asking, How are your prior beliefs doing? Do they have anything to do with the real world? For example, consider all those chemical exposures to which you assigned a prior probability of 20 percent that the relative risk for a certain cancer was between 1.02 and 1.10. Say there were 400 such exposure scenarios. If you are calibrated, then 20 percent of these 400 (or 80 of them) will actually have relative risks lying in that range. Therefore, to determine whether we are calibrated, we need feedback about the actual state of the world.

Unfortunately, as far as I can see in risk assessment, we're generally not getting that kind of feedback; we're just getting fancier and more complex ways to quantify our uncertainty. Our understanding of the science is coming too slowly and incrementally to allow for effective calibration. At present, the lab science is either too far removed from our human health concerns or predicts human risk too poorly. Epidemiology, limited by confounding and bias, is too blunt an instrument to quantify small risks at low doses. Maybe some year there will be a big scientific breakthrough that will change this, but what we're talking about now is a fundamental lack of useful information or definable (and doable) experiments that will answer the questions, Are we calibrated? Are we in fact making the right decisions? The point is that all this fancy Bayesian machinery can't answer that question. It merely says, If you quantify how uncertain you are, I'll tell you what to do, but I won't tell you that you are calibrated or that you are making appropriate decisions . . . unless you get good feedback data.

Well, does this Bayesian approach make any sense? I would argue that it's probably not such a great idea, when you have no explicit mechanisms with which to even attempt to calibrate the process. That's the problem in most environmental health regulation. We don't know what we're doing. We don't have a smoking gun for most low-level exposures that will tell us how many people are getting hurt. This lack of feedback is very different from a lot of other areas where such analysis is used. Much of the same Bayesian machinery, computer programs, and ideas have been applied to the world of business decision making, operations research, and medical decision making, where the prior beliefs, at least potentially, can be calibrated.

The importance of feedback and calibration is illustrated by the following story. In World War II, soldiers in training were taught to shoot at flying targets in an attempt to make them better at shooting down foreign planes. Unfortunately, the trainees weren't told how many they got right—how many targets they hit—until the end of the practice session, and none of them ever improved. They never became good shots until they started getting feedback after each shot. That's the difference. In business decision making people get data, they make a decision, things succeed or fail. One learns from experience and tries again. What I'm saying is that there's not much chance in risk assessment to learn from experience.

SOCIOPOLITICAL CONTEXT OF BAYESIAN ANALYSIS

So *why*, you might ask, *is everybody so hot on this?* Let's talk first about regulators. They need rules. Many of their old rules have lost their credibility because they are so uncertain. Now, these regulators can just state their (or more accurately their experts') uncertainty and proceed from there. So it appears they're trying to be more rational, and yet they still have a decision-making procedure.

Now, let's take advocacy people from both sides. Let's first take the people who advocate a cleaner environment. They might make the following argument, which I and my coauthors made in a 1985 paper (Robins *et al.*, 1985):

An argument in favor of the adoption of a Bayesian approach is that it may persuade the scientific community to require that experts report explicitly the distribution of their subjective probabilities. Thus, experts would be put in the position of having to say explicitly whether or not they believe, for example, that there is a 5% chance that a particular chemical exposure will cause 500 excess deaths. Requiring experts to report explicitly their probability distributions will insure that their willingness to accept a substantial amount of risk to the public cannot be hidden behind a cloak of apparent scientific objectivity. The public may be much less willing than the expert to accept such a level of risk. Further, when the public has the opportunity to see the wide range of expert opinion as to the number of excess deaths to be expected from a particular exposure, it will become apparent that in many cases the available scientific data are open to quite conflicting interpretations. The Bayesian approach provides a means for rational decision making in the face of such uncertainty. Additionally, it provides a means for the systematic examination of the subjective beliefs of the experts and regulators chosen by those who hold political power. (P. 327-28)

However, for the same advocates, this approach may be a problem in the courts. Courts don't like uncertainty all that much and usually require a finding that an association of exposure and disease is more probable than not rather than just possible. A 5 percent probability that things could be really bad may have a big public impact and influence regulation, but it could be a losing argument in a damage suit.

Now let's consider advocacy groups who oppose increased environmental regulation. Many such advocates think that Bayesian statistics will make regulations less stringent by focusing on the final average. Now, let me explain what I call the *final average*. The first part of the Bayesian analysis is to derive one's (or one's experts') subjective probability estimates. The final part involves using these estimates to decide what action to take. For example, here you may average a 5 percent probability that the chemical is harmful with a 95 percent probability that it is harmless. Even if there is a small chance that there is a big risk, when you perform this final averaging, the likely risk may appear small, and you might decide to do nothing about the exposure.

This suddenly allows the EPA not to have to take the worst-case exposure scenario, the worst-case dose response extrapolation. Suddenly things seem less risky because now we are averaging over the larger probability that there isn't much risk (but keep in mind that these uncertainty estimates are often really just guesses). I think this is one of the reasons people who don't think chemicals pose a significant risk are also coming to recommend Bayesian analysis—because they are focused on doing the final average. So it's strange that both sides of the political spectrum, as far as I can tell, are now recommending the same move away from current approaches, but for very different reasons.

Now, you might ask, *what is the effect of all this?* My point is that we don't know. This whole thing is yet another mousetrap that doesn't catch the mice we see. For all we know, environmental regulations on carcinogens have had the greatest payoff on heart disease. Who knows? Basically, what has the whole last 20 years of environmental regulation been about? Generally, it's been about keeping corporations from polluting too much. Throughout this entire period of argument back and forth about carcinogenic and other risks, it seems to me, particularly at the environmental level, that we are just too uncertain to know what these regulations have done at all. It might be a great idea to liberalize our regulations by computing the final average and, in effect, saying things probably aren't so bad, if in fact we didn't need all these regulations. On the other hand, it's a very bad idea if we do need them. The problem is that nobody knows. So in that sense, this Bayesian thing is kind of a sham. We just don't have the data, no matter how quantitative our uncertainty (i.e., ignorance).

To build a better Bayesian mousetrap has the potential to become a veritable cottage industry for academia and consultants. The regulators want good mousetraps, and a *good* Bayesian mousetrap requires sophisticated computer models and software that only experts can produce (and perhaps, understand); and there are a large number of hungry experts in academia and elsewhere willing to do so. Those experts that will enter this field are, like all experts, in danger of becoming mesmerized by their fancy tools—their elegant models and programs. It takes an incredible amount of hard intellectual work to build an elegant Bayesian mousetrap. The builder of the mousetrap, with so much hard work and *career chips* invested, is in danger of forgetting that his fancy Bayesian mousetrap still does not catch mice.

It can take years to do a highly sophisticated uncertainty analysis of the risk and benefits of even a single chemical in a very restricted exposure setting. But out there in the community, there are thousands of chemicals in thousands of exposure settings. Thus, the choice to put huge human and capital resources into these elegant, individualized academic projects may not be a wise approach to risk assessment, particularly if this choice is relied on in lieu of developing wide-ranging generic regulations for suspected carcinogens and other toxics.

The final major problem with the Bayesian mousetrap (and with most risk assessment based on cost-benefit analysis) is that it's rife with opportunities for manipulation by purely ideological argument. A major example is the recent *poor-is-riskier* movement. The poor-is-riskier argument goes something as follows: Rich nations and rich people are healthier than poor nations and poor people. Thus, wealth leads to health. Regulation, on the other hand, impedes the accumulation of wealth and, thus, leads to ill health. In fact, so much ill health may result from diminished economic growth due to regulation that the direct health benefits of regulation may be more than offset by a negative impact on gross national product (GNP). The basic engine driving the poor-is-riskier argument is simple empirical correlations between GNP and mortality rates, based on time trend or ecological data. If a first-year epidemiology student drew causal conclusions about the relationship between poverty and health from such gross correlation data, they would fail EPI 101. Yet professors of economics and public health conclude that poor is riskier from the same data in major public policy and government journals. I can only believe that this is due to the fact that there is a powerful audience for this message who want it *out there* for ideological reasons, no matter how weak its scientific basis may be.

Moreover, for the sake of argument, suppose it were the case that environmental regulation of toxic hazards results in a relative slowdown in the rate of GNP growth, which in turn leads to higher mortality rates. Is it then the job of public health professionals to argue that such regulations should not be carried out; or is it our job to argue that they should be carried out, but with sufficient improvements in health care coverage, job security, and unemployment benefits to offset any effects of economic adversity on health? Somehow, the same economic and public health professors alluded to above always seem to neglect to mention the latter option—much less to champion it.