The Perils of Prudence

How Conservative Risk Assessments Distort Regulation

Albert L. Nichols and Richard J. Zeckhauser

GOVERNMENT AGENCIES charged with regulating health risks—from workplace hazards and nuclear power plants to new drugs and toxic wastes—typically face massive uncertainties about the extent, sometimes even the existence, of those risks. The standard practice for dealing with such uncertainties is to be "conservative"—to rely on assumptions that give high estimates of risk in an effort to avoid unpleasant surprises. Faced with two sets of assumptions, one yielding an estimated risk of 1 case of cancer per 100,000 people, the other a risk of 1 per 10,000, most agencies would use the second set even if it were much less likely to be accurate.

This approach is most firmly entrenched for cancer risk assessment within such federal agencies as the Food and Drug Administration (FDA), the Occupational Safety and Health Administration (OSHA), and the Environmental Protection Agency (EPA). The EPA's new "Guidelines for Carcinogen Risk Assessment," issued this September after lengthy debate in the administration and Congress, make no pretense of seeking a best estimate of actual cancer risks, but rather strive to find a "plausible upper bound" for those risks. The justification is prudence: in the face of uncertainty, the government should proceed with caution and "err on the side of safety."

Albert L. Nichols is associate professor of public policy and Richard J. Zeckhauser is professor of political economy at the Kennedy School of Government, Harvard University. *This article is based on an earlier technical paper appearing in V. Kerry Smith (ed.),* Advances in Applied Microeconomics.

The cumulative effect of following the upper-bound path, using a long series of conservative assumptions, can be monumental overestimates of health risks. The result is more stringent and costly regulation of at least some types of risk than if policy makers were more realistically informed. But apart from creating a tendency toward overcontrol, biased estimates distort the pattern of regulation. Some low-level risks are regulated too stringently while more severe risks are tolerated. The price we pay for risk reduction is too high and, if the discrepancies in stringency are great enough, we may even end up with *more risk* than we would with realistic assessments. Conservatism in risk assessment, in other words, may well lead to a pattern of regulatory decisions that jeopardizes public health and safety.

Assessment vs. Management

Quantitative risk assessment is an increasingly important tool in regulatory decisions involving health and safety. Such assessments attempt to determine whether a substance poses a hazard and, if so, the magnitude of the risk, which is expressed as a probability of injury, illness, or death. Regulators typically weigh risk assessments against other factors (such as the cost of controls) in deciding what action, if any, to take.

Quantitative risk assessment is essential to intelligent regulatory policy. In the absence of explicit estimates of risks, policy is likely to be

determined by the availability and affordability of technology, as opposed to the costs and benefits of regulation. Technology-based regulation tends to require gross overexpenditures to control trivial risks in thriving industries, while ignoring far larger risks in less robust industries. In some cases, particularly those involving carcinogens, Congress may be led to impose zero-risk criteria, as in the Delaney clause of the Food, Drug, and Cosmetic Act, with equally arbitrary results. Quantitative assessments at least make it possible for regulators to strike some balance between risk reduction and other factors, such as the cost of controls.

Measuring risks and deciding what to do about them are distinct activities. "Risk assessment" refers to the primarily scientific enterprise of estimating risks. "Risk management" refers to the formulation of policies to control those risks; relying on political, ethical, and economic judgments, it is informed, but not determined, by science.

The separation of assessment and management is important. It helps shield the scientific process from political manipulation by preventing policy makers from altering risk assessments to fit their desired policy choices. At the same time, policy makers do not have to accept implicit policy choices masquerading as scientific fact. Maintaining the distinction clarifies political debate and accountability.

In practice, however, the line between risk assessment and management is blurred. Fundamental gaps in knowledge make risk assessment extraordinarily imprecise, requiring many choices to be made among competing models and assumptions. In the absence of firm data or scientific consensus, many of these nominally technical decisions end up being implicit choices about values and policies, though they are not acknowledged as such. This obscures the true character of the choices being made and reduces political accountability.

Cancer Risk Assessment and the EPA

Nowhere is the entanglement of risk assessment and risk management more evident than in cancer regulation, where the natural and social scientists, statisticians, and others who prepare risk assessments—whom we shall call "risk assessors" for convenience—characteristically make a long series of "prudent," conservative assumptions. At the EPA, the major producer and consumer of cancer risk assessments, the agency's administrator and political deputies (whom we shall call policy makers) are rarely told the range of plausible estimates, or the most likely estimate, of the risks they are asked to regulate. Instead, they are usually told only the "plausible upper bound" estimate, a term of art indicating that the actual risk is almost certain to be no higher. (In most cases the "plausible lower bound" is zero risk, as the EPA's own Carcinogen Assessment Group admits.)

Most policy makers and many observers of health and safety regulation are aware that cancer risk estimates are meant to be upper bounds rather than best estimates. Far fewer, however, realize how many upward twists are exerted in the assessment process or just how significant the cumulative effects can be.

Most policy makers and many observers of health and safety regulation are aware that cancer risk estimates are meant to be upper bounds rather than best estimates. Far fewer, however, realize how many upward twists are exerted in the assessment process or just how significant the cumulative effects can be. A brief look at the procedures used by the EPA will identify the most important sources of bias in the procedures used by many federal agencies.

Stages in an EPA Risk Assessment. To simplify a complicated process, consider three major steps in an EPA cancer risk assessment: (1) estimating the amount of the to-be-regulated substance currently in the environment (say the emissions level of a pollutant); (2) estimating exposure to the substance; and (3) estimating the cancer risk per unit of exposure.

Emissions. The first step is to estimate how much of the carcinogen is emitted into the environment. Although this is the most straightforward step in risk assessment, it has enough uncertainties to leave room for upward bias. The EPA, for example, often assumes that plants operate at full capacity, though few actually do. It may use out-of-date emissions figures when improvements in technology have lowered emissions. Or it may use unrealistic assumptions

about the lifespan of a substance. (EPA has assumed, for example, that new dyes will be used for 40 years, when in fact the average economic life of such dyes is less than a third of this.)

Such biases are likely to be greatest at the early stages of regulatory development, when the data are weakest. This is well illustrated by the EPA's decision in the late 1970s to assign high priority to regulating benzene emissions from maleic anhydride plants based on emission estimates that were about five times higher than its final estimates.

Exposure. The next step is to link emissions to exposure. The EPA typically begins with a computer model to predict how a substance is dispersed, then combines data on human populations with assumptions about behavior to estimate the size and distribution of the exposed population. Each step requires choices among alternative assumptions and estimates, offering numerous opportunities for the upper-bound route.

Conservatism is most likely to enter dispersion modeling when analyses are done for a single "model" source rather than for specific facilities. In such cases, the tendency is to choose values that reflect adverse rather than typical conditions. In estimating exposure to benzene from the maleic anhydride plants, for example, EPA used data from Pittsburgh, where, in its own words, "meteorological conditions that maximize ground-level concentrations . . . are common." A later critique by the industry, using local meteorological data for each plant, suggested that EPA's estimates were more than 50 percent too high on average.

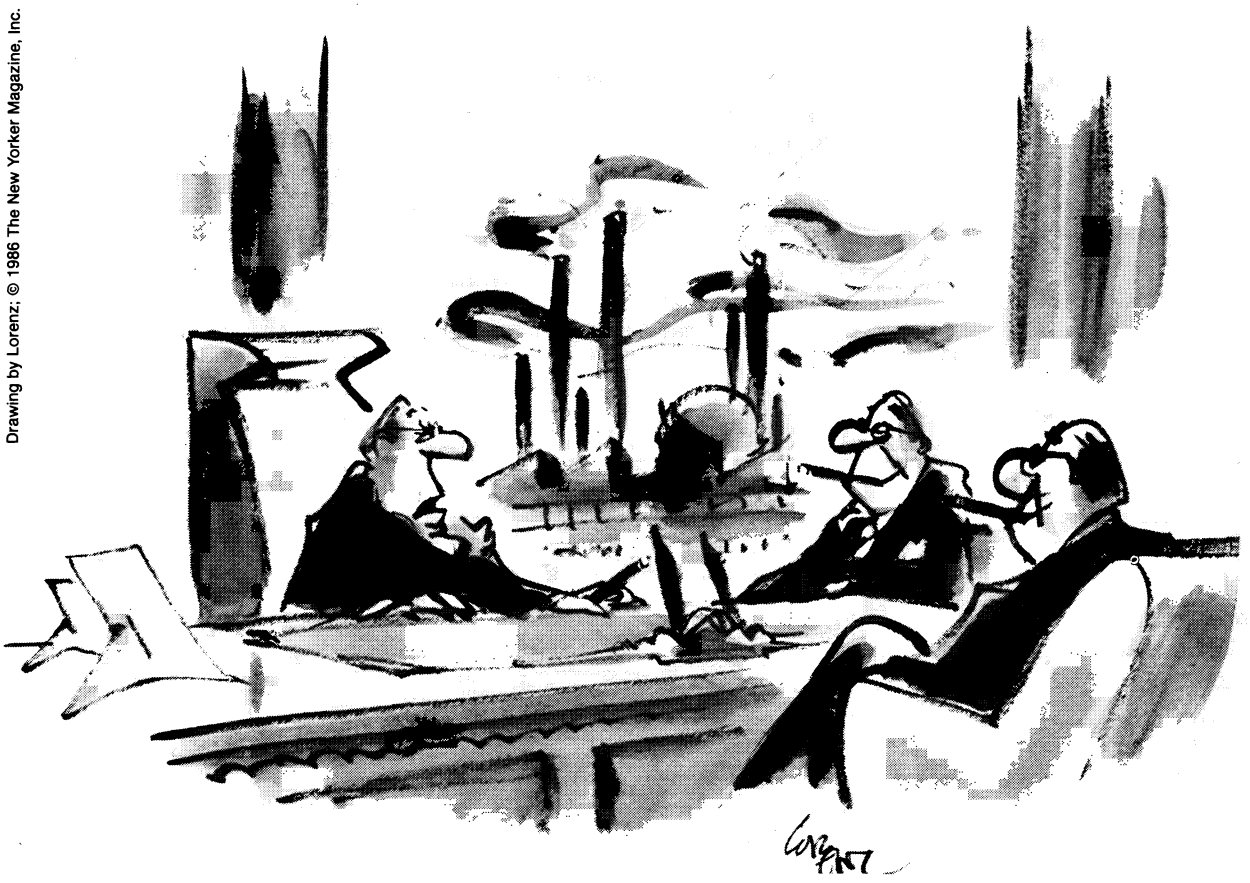
Similar problems arise in other media. The risks posed by hazardous wastes deposited in the ground, for example, depend in part on soil conditions and on how close the facilities are to drinking water supplies. Risk assessments are likely to assume that facilities are located close to drinking water wells and in highly permeable soils. While these assumptions are undoubtedly borne out in some facilities, the vast majority of facilities are likely to cause far less exposure.

The predictions of the dispersion model are then used to estimate human exposure. In the case of air pollutants, this involves multiplying the predicted concentrations at various distances from sources by population estimates. While this may sound straightforward, the estimation procedures now in use are likely to overstate the exposure reductions to be achieved through regulation. For example, dispersion models predict outdoor concentrations, implicitly assuming that people spend all of their time outdoors. But, in fact, most people spend the vast majority of their time indoors, and recent studies suggest that reductions in outdoor concentrations have less than a one-for-one impact on indoor concentrations.

Other aspects of the EPA's exposure assessments reflect similarly unrealistic assumptions about behavior. Estimates of "maximum individual risk," for example, assume that some individuals are born and die at the point of maximum pollution concentration and never leave that spot, even to go to work or school (or even to venture indoors). The possibility of averting behavior is usually ignored. Most estimates of the risk from ground-water contamination, for example, do not account for the possibility that the contamination will be discovered and alternative water supplies used. Failure to recognize that people may take steps to avoid exposure (for example, by buying bottled water when their well is contaminated) leads to overestimates of exposure and hence risk.

Estimates of "maximum individual risk," for example, assume that some individuals are born and die at the point of maximum pollution concentration and never leave that spot, even to go to work or school (or even to venture indoors).

Dose-Response. By far the most problematic step in cancer risk assessment is translating exposure estimates into risk estimates. This is done by constructing a dose-response function which describes the relationship between exposure and the likelihood of contracting cancer. Unfortunately, this function cannot be observed directly; a firm theoretical foundation for constructing it does not exist; and estimating it through controlled animal experiments would be prohibitively expensive, if not impossible. Given that the risks of concern are of low probability, usually on the order of 1 in 10,000 or less over a lifetime (less than one one-hundredth the average risk of dying in a motor vehicle accident), observing the incidence of disease even in very large populations may give little indication of the magnitude of risk. To estimate risk in such circumstances,



"I think we agree, gentlemen, that one can respect Mother Nature without coddling her."

scientists assess the level of risk where the dose, and hence the response, is many times greater. The usual sources of data are laboratory studies of animals exposed to very high doses or epidemiological studies of workers exposed to high concentrations.

Using these data to estimate risks to the general population almost always requires heroic extrapolation—inferring from the excess cancers observed among a few, heavily exposed humans or laboratory animals how many excess cancers there might be among humans whose exposure is vastly lower. Various models have been developed to perform these extrapolations but they are more math than science. Theory provides no clear support for any of them, and empirical tests cannot be used to determine which is most accurate.

EPA and other regulatory agencies typically rely on the "one-hit" extrapolation model (or one of its variants), which most scientists believe produces upper-bound estimates of low-dose risks. This model assumes that even a single molecule of a substance can produce cancer, and predicts that risk is proportional to dose (a linear relationship) at low to moderate exposure. Virtually all other extrapolation models are nonlinear, predicting that risk falls more than proportionately as dose decreases. When applied to the same high-dose data as the one-hit model, the non-linear models predict far lower risks at the doses relevant to most regulatory decisions. Indeed, the risk estimates from the one-hit model are often hundreds or thousands of times higher than those derived from the alternative models.

If risk data are available for more than one exposure level, the EPA uses a "multistage" model, which can be the same as the one-hit model or sharply nonlinear, depending on the data and method used to estimate the model's parameters. EPA, however, employs an estimation method that forces linearity at low doses, even when the data indicate a nonlinear relationship at known exposure levels. Very briefly, the EPA's assessors first find the maximum likelihood estimates (MLE) of all of the parameters, a

standard statistical technique. Thereafter, however, they proceed along the upper-bound route by finding the largest value of the linear term that cannot be rejected at the 95-percent confidence level. In most cases this increases the risk estimate by a factor of two or three over what it would be with the MLE, though in some instances the 95 percent approach can increase low-dose risk estimates by several orders of magnitude.

The MLE approach has problems of its own. Applied to certain kinds of exposure data, it can be highly sensitive to slight data variations and can even yield a negative linear dose-response term. But these difficulties merely reflect our lack of knowledge about how to infer low-dose risks from high-risks; the 95-percent confidence-level approach masks this uncertainty rather than reducing it. In any event, the EPA uses the 95-percent confidence level in all cases, not just where the MLE raises statistical dilemmas. Alternative procedures could avoid these dilemmas without introducing an additional bias toward extreme estimates.

Animal to human. Risk assessments rely on human epidemiological data where possible, but in most cases only studies of laboratory animals are available. There are obvious conceptual difficulties in extrapolating from small, genetically homogeneous animals (often bred to be particularly susceptible to cancer) to much larger, genetically diverse human beings who live far longer. Although most scientists agree that animal carcinogens should be considered likely human carcinogens, they disagree as to how, or even whether, quantitative estimates should be made from animal data. In the absence of a widely accepted model, uncertainty prevails and upper-bound extrapolation techniques tend to rule the day.

A critical quantitative issue is how to compute doses that produce equivalent risks in animals and humans. Dose-response relationships in animals are generally translated to human equivalents on the basis of weight or surface area. Because surface area increases by much less than weight as one moves from mouse to man, surface-area conversion leads to much higher risk estimates than weight conversion; using mouse data the difference is roughly a factor of 13, and using rat data it is about a factor of six. Studies of carcinogens tested in more than one species provide little basis for choosing between the two methods. Not surprisingly, the EPA has embraced the surface-area method and the accompanying higher estimates of risk.

The EPA selects the alternative, giving higher estimates at several other steps in the extrapolation from animals to people. In counting tumors in animal studies, for example, the EPA includes both benign and malignant tumors, arguing that benign tumors may progress to malignancy and that a substance producing benign tumors in animals might produce malignant ones in human beings. Most benign tumors never become malignant, however, so this procedure almost certainly overestimates human risks. It is as if we performed autopsies on every person who died and attributed each death to cancer if the autopsy revealed any tumors—whether or not they were malignant and whether or not they were the cause of death.

It is as if we performed autopsies on every person who died and attributed each death to cancer if the autopsy revealed any tumors—whether or not they were malignant and whether or not they were the cause of death.

A similar choice is made with tests involving multiple species (or strains of the same species) and sexes. Following the path of "prudence," data from the most sensitive animal strain are used. The primary rationale for this choice is to account for the more sensitive segments of the genetically diverse human population. But the fact that some people are less sensitive than average, and others merely average, also should be considered in estimating the overall risk to the population exposed.

Conservatism and Misallocation

If the purpose of using upper-bound assumptions is to provide a margin of safety in risk regulation, it is a badly flawed strategy. Although it may lead to tighter regulation in some circumstances (where risk assessments are most exaggerated), it also leads to less protection in others (where risk assessments are least exaggerated) and diverts scarce resources, including agency attention, from their highest valued uses. Whether conservative risk assessments lead to

policies which increase or decrease risk is an open question; they may well raise both costs and overall risk.

Part of the problem with upper-bound assumptions is that even modest overestimates can easily compound to yield a substantial exaggeration of the overall risk. Each choice, when viewed in isolation, may appear plausible and prudent, but the end result can be an extreme estimate that no longer qualifies as plausible. The problem is that the degree of conservatism applied at each stage accumulates multiplicatively. For example, if a risk estimate is a multiple of five (independent) factors, and risk assessors use a value for each factor just twice its expected value, their estimate will be 32 times greater than the expected risk.

Part of the problem with upper-bound assumptions is that even modest overestimates can easily compound to yield a substantial exaggeration of the overall risk.

Several experts in decision analysis have examined the effects of compounding uncertainties in the case of one substance, perchloroethylene, whose major use is as a dry cleaning fluid. Although they considered only two alternative values for each of three factors, their risk estimates varied by a factor of 35,000—ranging from a low estimate derived from a nonlinear, weight-based extrapolation from a rat study, to a high estimate derived from a linear, surface-area-based extrapolation from a mouse study. Given current scientific knowledge, one cannot determine which set of assumptions gives the best estimate. Standard risk-assessment procedures used by the EPA and other agencies, however, would use the most conservative assumption for each step, and thus would yield an estimate at the extreme end of the range.

Even if one accepts the basic assumptions behind the EPA's methods for estimating carcinogenic risks, there is ample room for controversy in the details of implementation. In the case of benzene, for example, the EPA's Cancer Assessment Group based its estimate on several epidemiological studies of workers exposed to high doses. Two EPA analysts argued that the CAG underestimated the exposure levels of the workers in some of the studies and incorrectly included some deaths of workers who either were not in an original cohort or who had been exposed to other potential carcinogens besides benzene. As a result, they concluded, the CAG estimate was too high by a factor of four. Another critic, a physician retained by industry, argued that the CAG had underestimated the baseline risk for one group of workers, concluding that the CAG estimate was too high by a factor of 10. All these estimates were derived from the same basic model and the same data, differing only in assumptions about "details."

These examples illustrate the degree to which conservatism can easily compound through a series of individually plausible assumptions to yield a result far more exaggerated—and uncertain—than decision makers or the public are likely to realize. They also illustrate that while the choice of a dose-response function is probably the single most problematic area of risk assessment, a series of "little" decisions also matter a great deal.

Misordered Priorities. If the degree of exaggeration were the same across different agencies and types of risk, conservative assessments would give the public an exaggerated notion of the magnitude of risks, but at least those assessments would be useful for comparing risks and helping to set priorities. Unfortunately, the effects of conservatism are not so benign. Differences in the degree of conservatism distort priorities and may well result in lower overall safety.

Some of the differences arise because different agencies, or even different parts of the same agency, follow their own procedures for estimating risks. The risk assessment "principles" issued by the White House Office of Science and Technology Policy in 1984, and the risk assessment guidelines published recently by the EPA, are attempts to eliminate this source of inconsistency.

Uniform guidelines are important, but no matter how rigorously followed they cannot eliminate the inconsistencies caused by conservatism. The most serious problem, which we discuss below, is that cancer assessment guidelines do nothing to redress imbalances in the way carcinogenic and other risks are estimated. Even among carcinogens, however, uniform conservative guidelines leave much potential for misordering risks. If estimates are always based on the most sensitive species tested, for example, the degree of bias will rise with the number of

species tested. Similarly, counting benign as well as malignant animal tumors distorts relative risk estimates because substances differ in the mix of tumors they cause in test animals. More generally, if the techniques applied to animal data are more conservative than those used when human epidemiological data are available, the upward bias will be higher for substances for which the only data come from animal tests.

The fundamental problem with inconsistent assessments is that policy makers are presented with false alternatives. Suppose there are two substances, A and B, and that the true carcinogenic risk of A is twice that of B. Because of differing degrees of conservatism, however, the risk from A is overestimated by a factor of two while the the risk from B is overestimated by a factor of 10. Policy makers are likely to conclude that B is substantially more dangerous and deserves more stringent control than A, when in fact the reverse is true. The resulting regulatory decision would *increase* risk. Where the substances under consideration are economic substitutes (such as competing pesticides or food additives), this problem is magnified since the banning of one leads to an increase in the use of the other.

This differential treatment of risks is institutionalized in the way new substances or activities are evaluated. New products, by having to be proved safe, are subject to much higher regulatory hurdles than existing products, which remain innocent until proved guilty, as Peter Huber has shown ("Exorcists vs. Gatekeepers in Risk Regulation," *Regulation*, November/December 1983). This bias extends to risk assessments as well, because the difficulties in predicting levels and patterns of use for materials not yet on the market provide fertile ground for upper-bound assumptions. Meanwhile, the risks from existing materials that would be displaced by the new ones are likely to go unassessed.

Carcinogens vs. Noncarcinogens. The potential for misallocation is most apparent where regulations address very different types of risks. Some risks are common enough and can be measured with sufficient ease that reliable statistics are available. We have very accurate counts of how many people die in motor vehicle accidents, for example. Predicting the effects of interventions—such as requiring air bags—is more difficult, but still much less uncertain than predicting the health impact of lowering the emissions of a carcinogenic substance. In the debate over air bags, fatality-reduction estimates that differ by a factor of two are regarded as very far apart; alternative estimates for carcinogens routinely vary by factors of 100 or more.

In cases such as air bags, conservative estimation techniques have relatively little effect because the band of uncertainty is relatively narrow; even when there is sharp disagreement about the effectiveness of an intervention, hard data on existing risks limit the degree to which the impact of regulation can be overestimated. In the case of carcinogenic risks, by contrast, *the greater the uncertainty about a given effect, the more likely it is to be overestimated.* As a result, highly uncertain risks, such as environmental carcinogens, are the ones most likely to be overcontrolled. This tendency is reinforced by the fact that the procedures for estimating carcinogenic risks typically are more conservative than those for noncarcinogens.

The fundamental problem with inconsistent assessments is that policy makers are presented with false alternatives.

Lead vs. Benzene. The debate over the EPA's 1985 decision to tighten the limit on lead in gasoline provides another illustration of the potential for distortions arising from asymmetric conservatism between carcinogens and noncarcinogens. The EPA estimated that a sharp reduction in lead in gasoline would reduce by 150,000 the number of children with potentially hazardous levels of lead in their blood, reduce other air pollutants from vehicles whose pollution-control catalysts (on newer "unleaded gas only" vehicles) are destroyed by leaded gasoline, and possibly prevent 5,000 deaths per year from cardiovascular diseases among middle-aged men. The value of these and other benefits exceeded estimated costs by at least three to one.

Critics of the proposed rule argued that reducing lead in gasoline was likely to increase the amount of benzene, a carcinogen. (See C. Boyden Gray, "EPA and the Gasoline Tar-Baby," in this issue.) The basic concern was legitimate. Comparisons of benzene and lead risks were distorted by the radically different degrees of conservatism employed in their estimation. The benzene risk estimate followed the conservative

procedures described above. Based on studies of workers exposed to several hundred parts per *million* of benzene, the estimate extrapolated on a linear basis down to current ambient concentrations of several parts per *billion*, more than 1,000 times lower. Other conservative assumptions biased the risk estimate further.

The lead analysis, by contrast, was limited to risks documented in epidemiological studies of health effects at or near current levels of lead exposure. The risk estimate for children did not include *any* effects below the blood-lead levels defined as potentially dangerous by the Centers for Disease Control, though many studies suggested that damage might occur at lower levels. (The levels in dispute were roughly one-half those at which serious effects are well accepted—not hundreds or thousands of times lower, as with benzene.)

The risk estimate for adults did not include any effects on women, since the study linking blood lead and blood pressure showed a statistically significant result only for men. And because the best data on cardiovascular risk factors were for white males aged 40 to 59, no estimates were included for nonwhites or for white males outside that age range. Had the procedures employed been analogous to those used with carcinogens, the results for white males aged 40 to 59 (or for male rats, for that matter) would have been extended to women, nonwhites, and other age groups, yielding estimates many times higher. The benzene estimates, although based on studies of working age men (some of them Turkish shoe makers who employed high concentrations of benzene in home-based production), were extended to the entire U.S. population.

Because of these differences, comparing EPA's lead and benzene risk estimates is misleading and might have blocked control of lead in order to avoid the exaggerated risks estimated for benzene. Two factors prevented this from happening. First, even with the conservative approach used for benzene, the estimated risks were small compared to those for lead—less than four cases per year for benzene emissions from service stations. Second—and probably more important politically—additional analysis indicated that reducing lead in gasoline would probably *reduce* benzene emissions on net, because fewer catalytic converters would be disabled by misfueling with leaded gas.

The lead-benzene case had a happy ending, but we cannot expect that the bias introduced by asymmetric conservatism will always be so fortuitously overcome. The misplaced concern about benzene and the excessive conservatism used in estimating its risks are symptomatic of what we believe to be a broader problem: The federal government spends more of its marginal resources regulating carcinogens—rather than other safety, health, or environmental hazards—than is warranted by the likely benefits for public health and welfare. John Morrall's review of cost-effectiveness estimates for a broad range of regulations (see John Morrall, "A Review of the Record," in this issue) suggests that cancer risks are in fact greatly overregulated relative to other risks. He finds that the estimated cost per life saved tends to be much higher for "health" (in practice, cancer) regulations than for those promoting safety. A more refined comparison that thoroughly adjusted for differential conservatism in risk assessments would show even more dramatic differences, with much higher estimated costs per life saved for regulation of carcinogens.

The federal government spends more of its marginal resources regulating carcinogens—rather than other safety, health, or environmental hazards—than is warranted by the likely benefits for public health and welfare.

Fear of Cancer

Some observers argue that cancer risks should be assessed more conservatively because people view cancer as a worse way to die than other ways, such as heart attack. But if this preference does exist, the appropriate course of action is to assign greater policy weight to averting cancer deaths than other fatalities, not to overstate the risk of cancer. Consider a policy choice from a very different arena, education. Suppose evaluations of two reading programs show one to be more effective with poor students and the other to be more effective with average students. If we value reading gains by poor students more than those by average students, we would not account for this preference by overestimating the effectiveness of the first program; rather we

would place greater value on gains achieved at the bottom of the scale. This would be reflected in higher estimates of the benefits of the reading program helping poor students and, perhaps, by greater spending on that program. By analogy, we might place greater weight on cancer deaths than on other fatalities in determining the costs and benefits of regulation, but this does not mean we should exaggerate the risks of cancer.

Whether cancer merits "extra points" in public health decisions is an open question. While cancer may be a terrible way to die, it is likely to come later in life than, say, death from an auto accident. If a 20-year-old were offered a choice between a certain reduction in his lifetime risk of death from cancer or auto accidents, he might well opt for the reduction in auto risks. This would save many more years of life, which might more than compensate for the extra suffering associated with dying from cancer.

Alternatives to Upper-Bound Assessment

Critics of current risk assessment procedures often recommend that risk assessors be required to report a range of risk estimates based on alternative plausible assumptions rather than a single estimate based on upper-bound assumptions. It would then be up to policy makers to decide how much weight to give to the different estimates, and how to balance risk reduction against cost and other factors. More complete reporting is an important first step, but it is incomplete and raises problems of its own. Which assumptions, of the almost limitless array of alternatives, should be used to generate the range of estimates? Presented with ranges of estimates that vary widely, on what basis should policy makers choose among them? While the current system suffers from treating some important value judgments as if they were scientific questions, the alternative of presenting huge ranges and no other information would make the opposite mistake—it would shift the responsibility for making scientific judgments to policy makers.

Two additional steps can be taken to address these problems. First, in addition to reporting ranges of risk estimates, scientists should present estimates of *how likely* it is that any particular risk estimate is correct. Second, policy makers should base their decisions about most health risks on the *expected value or mean estimate of the risk, not the upper bound.*

The Expected-Value Approach. To see how the expected-value approach works, recall the example at the beginning of this article: One set of assumptions predicts that the risk of cancer from a chemical is 1 in 10,000, while another set puts the risk substantially lower at 1 in 100,000. Suppose that risk assessors, following our first prescription, report both figures along with their estimate that each one is correct—90 percent for the lower estimate and 10 percent for the higher estimate. How should this information be used?

Under current agency practice the policy maker would use the high estimate, ignoring the very high probability the risk is 10 times lower. An alternative would be to use the *most likely* estimate—to assume that the lower estimate, which scientists believe has a 90 percent chance of being correct, is the right one—but this of course would ignore the possibility that the risk is 10 times higher. The expected-value approach, by contrast, uses all the available information. It is simply the weighted average of the risk estimates, with the weight for each alternative equal to the subjective probability that it is correct. In our example, the expected value of the risk is about 2 in 100,000, which is roughly one-fifth the conservative estimate but almost twice the "most likely" estimate. Although the specific ratios are artifacts of our example, the basic ordering is not. Expected-value estimates of risks will be lower than upper-bound estimates, but generally higher than "best" or "most likely" estimates; the less likely it is that the upper-bound estimate is correct, the lower the expected value.

The expected-value approach, firmly grounded in the theory of rational decision-making under uncertainty, offers several major advantages over present practice.

The expected-value approach, firmly grounded in the theory of rational decision-making under uncertainty, offers several major advantages over present practice. First, it produces risk estimates that differ because of actual differences in likely degrees of risk (or at least, the scientists' best estimation of those risks) rather than because of differences in degrees of exaggeration. This would facilitate comparisons of risks and policies, both within and across agen-

cies, and would aid in setting priorities. Second, it imposes an appropriate division of labor between risk assessors and policy makers. Compared with the current approach, fewer value judgments would be concealed in the risk-assessment process, thereby opening up public debate and improving the accountability of public officials. Third, the expected-value approach mitigates against a natural reaction to risk assessors who always cry wolf: As policy makers become aware that risk assessments are arbitrarily inflated, some are likely to compensate by discounting the size or importance of those estimates, leading to increased reliance on "pragmatic" considerations (such as "affordability" and short-term political pressures) and even greater policy inconsistencies.

Risk Aversion: A Misapplied Argument. Some proponents of current procedures argue that the expected-value approach is inappropriate because most individuals are highly risk averse. Faced with a range of risk estimates, the argument goes, policy makers should rely on the high estimate to take account of aversion to risk (particularly cancer risks). This argument reflects a misunderstanding of both risk aversion and the expected-value approach.

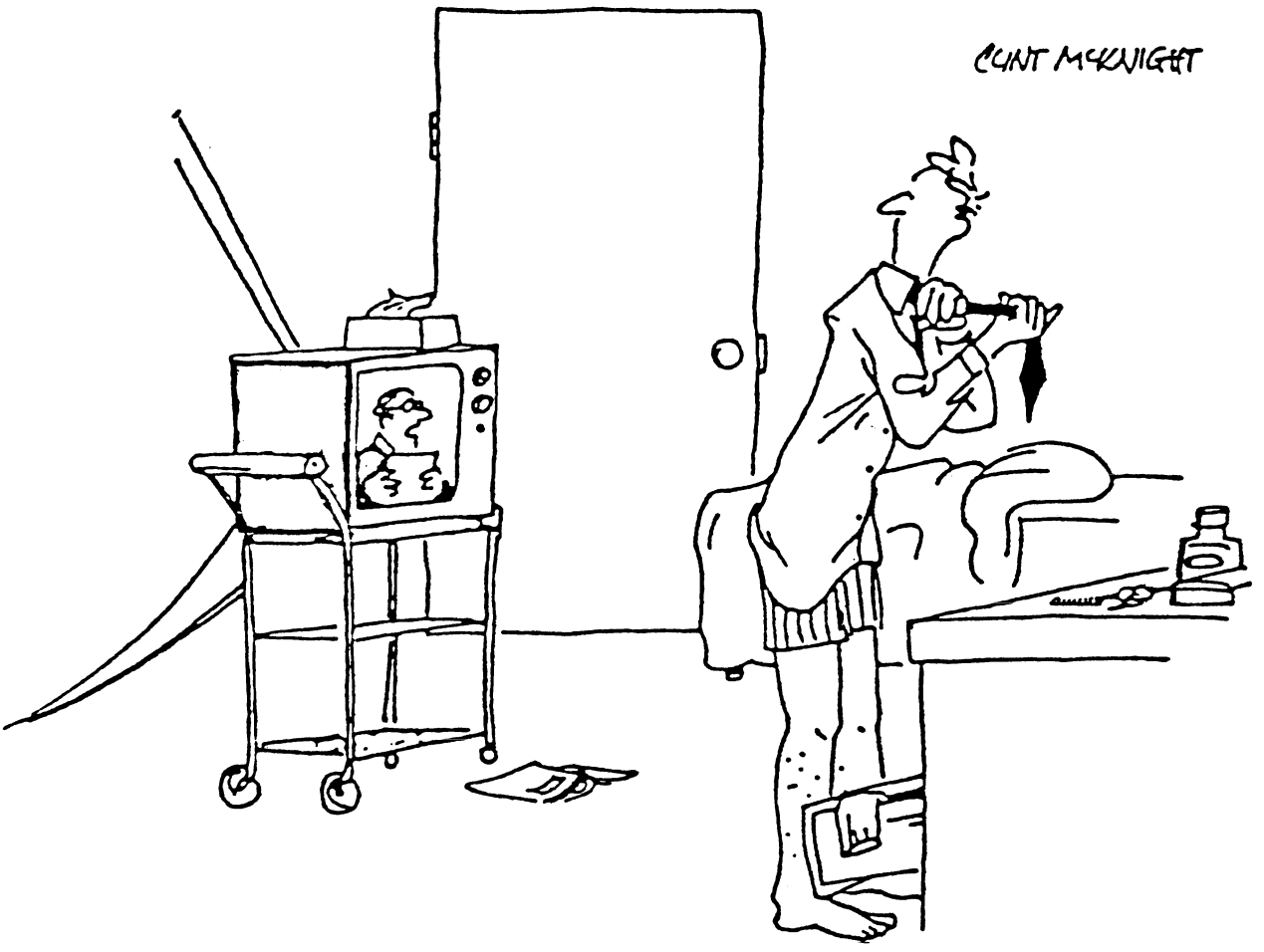
"Risk aversion" is a technical term used to describe the attitude of most people toward risk. Used correctly, it refers to our willingness to pay a premium above expected value to avoid a gamble. Many of us, for example, would pay $200 to insure against losing a $100,000 house to fire, even if the risk of loss was only 1 in 1,000 and the expected loss was, therefore, only $100. Similarly, we generally need to be compensated to accept risk, which is why, for example, stocks offer a higher expected return than more stable government bonds. But rational individuals make such choices based on their preferences regarding different outcomes and their assessments of the likelihood of those outcomes—not by exaggerating the risks. We do not decide whether to purchase fire insurance or equity securities by assuming that our children will surely play with matches or that the stock market crash of 1929 is sure to happen next year. The same principle applies to decisions involving health rather than money, or involving tradeoffs between the two. Tradeoffs between risk and dollars are questions of preference, not science, and should be made in light of the most accurate information about risk—the expected value.

This is an extremely important point, worth elaborating by example. Suppose an individual must choose between exposure to two chemicals. Chemical A poses a risk of cancer of 1 in 10,000 for each year of exposure. The risk from chemical B is uncertain; there is a 95-percent chance that it is "safe" (zero risk) and a 5-percent chance that it poses a risk of 1 in 1,000. Which one will a rational individual choose? Intuition may suggest that he would prefer A because with B he runs the chance of exposing himself to a much higher risk. In fact, however, he would be much better off exposed to B. His annual risk of contracting cancer from B is half of that for A (since 5 percent of 1 in 1,000 is 1 in 20,000), so choosing A would not "err on the side of safety." Whether or not the individual is risk averse is irrelevant because the possible outcomes are the same with both chemicals—he either gets cancer or he does not. All that matters is the probability of contracting cancer. This situation is quite different from our fire insurance example where there is a choice among different possible *outcomes* (having a fire and receiving compensation versus having a fire and receiving no compensation); in such cases, risk aversion appropriately plays a role, and the more risk averse the individual is, the more he will be willing to pay for insurance.

We do not decide whether to purchase fire insurance or equity securities by assuming that our children will surely play with matches or that the stock market crash of 1929 is sure to happen next year.

Because resources are limited and all risks cannot be eliminated, tradeoffs inevitably must be made in risk-reduction efforts. While overstating one risk may cause more resources to be devoted to its reduction, it also is likely to divert resources from other problems, some of which may be more serious. Any decision based on upper-bound estimates risks making the "chemical A mistake."

Risk Aversion and Risks to Society. A somewhat more sophisticated objection to the expected-value approach disputes its applicability to public decisions in which the health of many people is at stake. Suppose, for example, that our two chemicals are alternative pesticides;



"The Food and Drug Administration today banned all forms of physical activity when five laboratory rats became, quote, 'super tired' after running on their wheel for a few hours."

that the EPA is deciding which one to allow on the market; and that in either case 1 million people will be exposed. If B is banned, A will cause about 10 extra cancers per year. If A is banned, there is a 5 percent chance B will cause 100 cancers per year, but a 95 percent chance that it will cause no cancers. As pesticide B offers half the expected number of cancers, any rational individual would prefer to be exposed to B. Nevertheless, some believe that society should choose A on the grounds that, ignoring the probabilities, fewer people would die if the worst case were to result.

This argument is made most often in the context of nuclear power plants, which pose the risk of a major catastrophe albeit with extremely low probability; it is implicit, however, in all risk assessment procedures that rely on upper-bound estimates. The rationale is that the simultaneous death of 1,000 people in the same incident is somehow worse than the isolated deaths of 1,000 otherwise identical people in separate incidents. Indeed some observers have argued that the perceived loss rises with the square of the number of people killed in a single incident implying that society should do no more to prevent a million individual deaths than it does to prevent an accident that would kill 1,000 people simultaneously.

We are extremely skeptical of such views. Although it is clear that a single large accident attracts more public attention and concern than the same number of fatalities reaped one or a few at a time, it is far from obvious that the total loss is greater. But even if it is, this is no argument for inflating risk estimates themselves—which can only increase (unnecessarily) public anxiety over risks of catastrophes.

Attitudes toward large-scale catastrophes are, in any event, largely irrelevant to the question of how the risks from environmental carcinogens (and most other environmental health threats) should be assessed. Most environmental carcinogens result in relatively few cases of can-

cer, which tend to be so scattered that they cannot be distinguished from normal, random variations in cancer incidence. This is so even for worst-case risk estimates. For example, a recent EPA study of airborne toxic substances estimated that the total cancer risk for several dozen substances, based on all the standard conservative assumptions, is less than 2,000 cases of cancer per year, or less than one-quarter of 1 percent of the 850,000 cases of cancer Americans contract each year. Moreover, since many of these 2,000 cases are attributed to "products of incomplete combustion," a catchall category including millions of individual sources (cars, wood stoves, and virtually every other activity that involves the burning of fuel), most regulatory decisions would be likely to involve only a handful of cases.

Although risk aversion in its technical sense is largely irrelevant to assessing and managing carcinogenic risks, the expected-value approach is perfectly compatible with the informal use of the phrase to mean simply that people wish to avoid risks to their health. If society values risk reduction highly, that should be reflected in the decisions of policy makers, not in exaggerated risk assessments that mislead the public about the real tradeoffs between risk and costs, and distort priorities across different types of risks.

Toward Reform

The expected-value approach, applied thoroughly to all aspects of risk assessment, is a goal for long-range reform rather than a method that could be applied right away. At several critical steps in risk assessment, such as extrapolating from high-dose to low-dose risks and from animals to humans, our knowledge is so meager that there is no way to judge objectively the probability that alternative risk estimates are correct. Until our understanding of underlying biological processes improves, the only recourse is to obtain subjective probability estimates from a number of experts.

There are, however, a number of straightforward changes that could be made immediately and that would have a very considerable effect in making risk assessments more accurate and scientifically neutral. Simple and reasonable changes in the method used to estimate the parameters of the dose-response model from experimental data could reduce by half the overestimation in cancer risk assessments. Better exposure assessments—based on realistic assumptions about pollution dispersion and human activity—offer similar opportunities for reducing exaggeration and making estimates more consistent across substances. A slightly more ambitious reform would be to develop procedures to incorporate the full range of available animal test results rather than only the one yielding the highest risk estimate. While each reform might have only a "small" effect (e.g., reducing estimated risks by half), the cumulative effect could be large.

In addition, as mentioned earlier, scientists should begin to present policy officials with a broader range of risk estimates. While policy makers may continue to take refuge in upper-bound estimates, at least the uncertainties involved, and the degree of exaggeration built into conventional estimates, would be clear to them and to the public. Moderation in regulatory decisions would be a likely result.

The goal, in any event, should be clear: Risk assessments should be as close to expected values, rather than overstated (or for that matter understated) values, as the state of scientific knowledge permits. Perhaps the most important step toward this goal is simply wider acknowledgment and understanding of the degree of deliberate overstatement in current risk estimates. Without this, efforts to reduce the degree of "prudence" may be perceived as attempts to politicize risk assessment and reduce regulatory protections. In fact their entire purpose is to depoliticize risk assessment and improve regulatory decisions concerning health and safety. ■

Selected Readings

Huber, Peter. "Exorcists vs. Gatekeepers in Risk Regulation," *Regulation*, Vol. 7, No. 6 (1983).

National Academy of Sciences. *Risk Assessment in the Federal Government: Managing the Process*, Washington, D.C.: Report of the Committee on the Institutional Means for Assessment of Risks to Public Health, 1983.

Nichols, Albert L. "The Regulation of Airborne Benzene," in T.C. Schelling, ed., *Incentives for Environmental Protection*. Cambridge, MA: MIT Press, 1983.

U.S. Environmental Protection Agency. *Costs and Benefits of Reducing Lead in Gasoline: Final Regulatory Impact Analysis*, Washington, D.C.: Economic Analysis Division (February 1985).