# The Peril and Promise of Risk Assessment

## Richard B. Belzer

Unfortunately, the practice of risk assessment by the federal government routinely departs from the academic ideal. Federal risk assessments continue to rely on conservative models and assumptions that effectively intermingle important policy judgments with science. This often makes it difficult to discern serious hazards from trivial ones, and it distorts the ordering of the government's regulatory priorities. These distortions typically lead to disproportionate investments in reducing very small threats to health and life. In some cases these distortions may actually increase net health and safety risks.

Widely acknowledged problems that continue to plague the practice of risk assessment in the federal government were described in the 1990 edition of the *Regulatory Program of the United States*, an annual publication of the Office of Management and Budget. The issues were not new, nor was the forum original inasmuch as previous editions of the *Regulatory Program* had raised similar concerns. But the unusual candor of the 1990 edition provoked a storm of controversy within federal regulatory agencies. The policy issues kindled by risk assessment, which for years had been relegated to obscure scientific journals, had finally become visible to

*Richard B. Belzer is a staff economist in the Office of Information and Regulatory Affairs of the Office of Management and Budget.*

the highest levels of the federal government.

The 1990 *Regulatory Program* highlighted three concerns. First, the continued reliance on "reasonable worst-case" assumptions distorts risk assessment and yields estimates that may overstate the expected level of risk by several orders of magnitude. Second, the assumptions embedded in risk assessments impart arbitrary "margins of safety" for which there is no scientific basis. The choice of an appropriate margin of safety is a value judgment that should remain the province of responsible risk management officials, and it is inappropriate to conceal it within ostensibly scientific risk assessments. Third, current risk assessment procedures distort the regulatory priorities of the federal government and direct scarce resources toward reducing trivial carcinogenic risks while failing to address more substantial threats to life and health.

Cancer risk assessment has become extraordinarily controversial over the past few years. It has been subjected to a crescendo of criticism by prominent scientists, risk assessment professionals, and policy analysts. Defenders of the faith have responded in kind by challenging the arguments of the accusers with gusto and occasional vitriol. It remains an open question whether risk assessment can survive this internecine warfare.

Despite these battles over its underlying validity, quantitative risk assessment plays an increasingly important role in the federal government's management of risks. Public confidence in the government's

scientific objectivity never has been so important. Policymakers and risk management officials need high-quality risk assessment to assure an effective ordering of regulatory priorities and to maintain (or perhaps to restore) public confidence in the risk management process. As former EPA Administrator William D. Ruckelshaus noted in 1983, "risk assessment . . . must be based on scientific evidence and scientific consensus *only.* Nothing will erode public confidence faster than the suspicion that policy considerations have been allowed to influence the assessment of risk."

## Current Risk Assessment Procedures

Risk assessments of chemical substances in general (and possible carcinogens in particular) consist of a mixture of facts, models, and assumptions. Facts are beyond dispute, of course, but there is considerable debate concerning the scientific merits of the models and assumptions commonly used in risk assessment. In some cases a scientific consensus has developed to support a particular model or assumption, but in many other instances certain models and assumptions are relied on simply because they reflect past practices. Put simply, no scientific basis exists for some of the most critical models and assumptions used to assess cancer risk.

These models and assumptions generally lead to a substantial overestimate of risks. That is, they lead to estimates of a "reasonable worst case" rather provide information about the typical or average level of risk. This bias arises within the procedures used to estimate both *hazard* and *exposure.* In fact, additional biases are embedded in so many steps that in the final result risk assessments often exceed by orders of magnitude the risk posed to the average exposed individual.

Several procedures generally used to extrapolate the results from animal tests to human risk are explicitly and intentionally biased. Therefore, risk assessors often characterize estimates as "upper-bound excess lifetime cancer risks." The term *upper-bound* means that there is a small (but known) probability that the true (but unknown) risk actually exceeds the value specified. Of course, the true risk is just as likely to be as small as a corresponding lower bound, which may be zero. Similarly, the caveat "lifetime" is added to reflect the assumption that exposure to the substance in question occurs continuously for seventy years.

It is also important to recognize that these estimates refer to "excess" cancer risks. The average

American's lifetime risk of cancer is approximately one in four. One-third of this risk is attributable to smoking; another one-sixth is related to diet. All other causes, including environmental, occupational, and dietary exposures to carcinogens and aging, thus pose an average lifetime cancer risk of one in eight. When a risk assessment is published that suggests that a particular substance poses an "excess" cancer risk of one in 10,000, this means that the lifetime risk of cancer faced by the average non-smoking American exposed to this substance may be increased by as much as one tenth of one percent.

**Choosing between Animal Tests and Epidemiology.** Animal testing enables scientists to estimate risks before human health effects become evident. Animal tests can also be conducted under tightly controlled laboratory conditions that allow exposure to be carefully calibrated. In contrast, epidemiological studies must rely on less accurate exposure

---

**Federal risk assessments continue to rely on conservative models and assumptions that intermingle important policy judgments with science. This often makes it difficult to discern serious hazards from trivial ones, and it distorts the ordering of the government's regulatory priorities.**

---

measures, some of which (such as recall) are inherently biased. It is also easier to control for confounding factors that would systematically alter risk estimates with laboratory animal tests than with epidemiological studies.

For these reasons, combined with an ethical aversion to delaying action until human "body counts" are available, animal studies are the dominant source of risk assessment data. Unfortunately, animal testing also suffers from serious limitations. Laboratory controls are by no means complete or sufficient. They generally fail to control for total caloric intake, for example, which has been associated with an increased incidence of tumors independent of exposure to possible toxins. Even more important, there is no generally accepted scientific basis for extrapolating low-dose human cancer risks from high-dose rodent bioassays. Current practice reflects a collection of scientific conventions for which there is little more scientific support today

than there was over a decade ago when the procedures were first developed.

Despite these problems, properly conducted animal tests and epidemiological studies both have useful roles to play in quantitative risk assessment. Indeed, they are complementary. The usual weaknesses of epidemiological investigations—unreliable exposure data, confounding effects—are readily avoided in laboratory tests on animals. Conversely, the weaknesses of animal tests—problematic extrapolation from high to low doses, arbitrary conversion of animal exposure to human equivalents—do not arise in epidemiological studies. Careful risk assessments incorporate both kinds of analysis to ensure that the emerging pictures are themselves internally consistent.

Current practice among federal regulatory agencies departs significantly from this ideal. Animal tests are often preferred to epidemiological studies when the former suggest higher risks. In a recently proposed regulation concerning cadmium, for example, the Occupational Safety and Health Administration (OSHA) proposed a new permissible exposure limit based on a risk assessment derived from an animal test rather than from a high-quality epidemiological investigation. OSHA rationalized its preference by pointing to the animal study's superior control of exposure and its capacity to predict tumors at multiple sites. Animal tests inherently have these advantages over epidemiological studies, however, so the conditions under which OSHA would rely on human rather than animal data are unclear. But the more important question is whether OSHA was also influenced by the fact that the data from the animal test predicted low-dose cancer risks ten times greater than the data obtained from the epidemiological study.

**Biases Embedded in Cancer Risk Assessment.** In many important ways the judgments that enter into animal-based risk assessments are intended to amplify the resulting estimate of risk.

● *Sensitive Test Animals.* Animal bioassays rely on homogeneous, genetically sensitive strains of rats and mice. This enhances the power of the test to detect abnormalities such as cancer. Certain animal strains have high rates of spontaneous tumor formation, however, and some scientists question whether observing elevated tumor rates in such animals provides useful information in estimating human cancer risk. Despite these concerns, cancer risk assessments often proceed on the assumption that elevated tumor rates found in sensitive animals are

sufficient to conclude that a substance is likely to be a human carcinogen.

The use of sensitive animal strains is not suggestive of bias per se, however. Rather, the bias arises because federal risk analysts often select the combination of species, strain, and gender that yielded the most significant tumorigenic response, and disregard all other results. Because there is no scientific basis for making such determinations, this practice cannot avoid imparting bias to federal agency risk assessments.

● *Severe Testing Conditions.* Current risk assessment protocols require the use of very high doses in animal tests. One group of animals is exposed to the highest dose that can be administered without inducing chronic excessive morbidity or mortality—the so-called maximally tolerated dose. A second group is exposed to one-half of this dose, and a third group

---

Some scientists have concluded that it is not scientifically credible to use the results from rodent tests performed at the maximally tolerated dose to estimate human health risks arising from exposure to low doses.

---

(if there is one) is exposed to one-fourth of the dose. Typically, all of these doses greatly exceed the level of exposure encountered by human populations.

Unfortunately, high doses may induce cancer for reasons unrelated to biological mechanisms that operate at low doses. At the maximally tolerated dose substances often cause severe inflammation and chronic cell killing. These doses may induce cancer simply because of chronic toxicity. For example, formaldehyde administered at the maximally tolerated dose causes nasal tumors in rats. These tumors appear to result from the inflammation of the nasal passage tissues. It is unclear whether the observed response is due to high-dose toxicity or perhaps to some other characteristic of the test species since the observed tumor rates exceed by a factor of twelve the rates found in the next-most-sensitive species tested.

Some scientists have concluded that it is not scientifically credible to use the results from rodent tests performed at the maximally tolerated dose to estimate human health risks arising from exposure to low doses. By one estimate, about half of all chemicals tested at the maximally tolerated dose cause

tumors in animal tests, and this ratio appears to be the same whether the chemical in question is natural or synthetic. Two-thirds of these positive results drop out at a dose equal to one-half the maximally tolerated dose, however. This leads some scientists to ask whether other factors besides mutation (cell proliferation, for example) may be the underlying mechanism behind high-dose carcinogenesis. Such questions have led to considerable pressure within the scientific community to reconsider whether maximally tolerated dose administration is appropriate for estimating human cancer risks.

● *Conversion from Animals to Humans.* When relying on animal tests to estimate human cancer risks, scientists must convert exposures in the test animal to human dose-equivalents. The two most common conversion formulas involve body weight and surface area, and there are scientific reasons for choosing either approach in individual cases. The surface area approach leads to estimates of risk that are between seven and twelve times greater than those that derive from the body-weight methods, however, and despite the ambiguity of the underlying science, EPA guidelines require the use of the surface-area method except in extraordinary cases.

Federal risk analysts have been working for some time to resolve the dispute concerning the appropriate conversion factor. This is both a welcome development and a potential problem. Although it is indisputable that scientific consensus is desirable on this issue, the anticipated resolution—using body weight raised to the two-thirds power—appears to be more of a political compromise than a scientific consensus. A uniform assumption based on non-scientific concerns may bury this legitimate scientific

---

**In cancer risk assessment, once a statistically significant positive result has been obtained in one test species, strain, or gender, the statistical burden of proof shifts to the no-effect hypothesis, which is logically impossible to prove.**

---

dispute within the risk assessment process and leave risk management officials and the public unaware of one more significant area of scientific uncertainty.

● *Selective Use of Alternative Studies.* Federal risk assessment guidelines recommend that relevant animal studies be considered irrespective of whether



" 'BE CAREFUL' ! ALL YOU CAN TELL ME IS 'BE CAREFUL' ? "

they reveal a positive relationship. These guidelines give appropriately greater credence to studies that show a positive response than to studies that are ambiguous or negative. In practice, however, a single positive study may overwhelm a host of negative studies.

A recent example of the selective use of alternative studies is the EPA's decision to ban the plant growth regulator daminozide (Alar). The scientific basis for this decision was a single positive animal bioassay. According to the EPA's cancer risk assessment guidelines, overcoming such a classification requires, at a minimum, two "essentially identical" studies showing no positive relationship. In the case of Alar, however, a more stringent test appears to have been applied. Three high-quality studies failed to show significant effects, but they received little or no apparent weight in the classification decision. In cancer risk assessment, once a statistically significant positive result has been obtained in one test species, strain, or gender, the statistical burden of proof shifts to the no-effect hypothesis. Because it is logically impossible to prove a negative, however, these procedures establish a virtually irrebuttable presumption in favor of the carcinogenesis hypothesis.

A more defensible approach is to assign weights explicitly to each relevant study that meets the minimum standards of scientific quality. Such a procedure would actively seek to incorporate in a scientifically appropriate manner all the information

available at the time a decision must be made. Risk analysts shy away from such a process because they consider any weights to be subjective emendations lacking scientific basis. Although this concern is certainly valid, the absence of an *explicit* weighting system leads to an equally subjective but hidden *implicit* weighting scheme. A weight-of-evidence procedure with documented weights would reflect the informed judgment of respected scientists, whereas the existing procedure is both undocumented and politically unaccountable.

---

**No single mathematical model is accepted as generally superior for extrapolating from high to low doses. Rather than be a scientific footnote to the risk assessment process, however, the choice of model becomes an important policy issue.**

---

● *The Choice of Dose-Response Model.* Having selected a single data set from among the laboratory animal tests, risk analysts must then extrapolate low-dose human risks from the data generated by high-dose animal tests. They use mathematical models to do this.

No single mathematical model is accepted as generally superior for extrapolating from high to low doses. Rather than be a scientific footnote to the risk assessment process, however, the choice of model thus becomes an important policy issue. For example, when OSHA used five different dose-response models to estimate cancer risks from cadmium, risks at moderate doses varied by a factor of 100. At doses in the range of the proposed exposure limit, two of the five models yielded excess lifetime cancer risk estimates on the order of one in 1,000, a level often regarded by policymakers as unacceptable. Two other models predicted essentially zero risk, however. Since none of the five models enjoys a biologically superior basis for estimating low-dose risks, the choice of dose-response model became a critical policy decision.

The preferred procedure under such circumstances would be to explicitly develop a subjectively derived "best" estimate or risk distribution while fully informing both political officials and the general public as to the uncertainties involved. In the case of OSHA's cadmium proposal, however, this practice was not followed. Agency staff used a

multistage model to determine whether low-dose exposures constituted a significant risk and estimated both the baseline risks and the benefits from regulation solely on the basis of this embedded policy choice.

The multistage model is the most commonly used method for estimating low-dose risks. Various features of the model typically cause it to produce high risk estimates even when the data are poor or inconsistent. Moreover, it yields higher risk estimates than many other models that have equal scientific plausibility. The linearized multistage model, a special version of the multistage model, is much more inherently conservative than the multistage model because it is explicitly and intentionally biased. Some agencies routinely use the linearized multistage model despite (or perhaps because of) its additional inherent bias. This practice lacks any basis in either biology or statistics. Ironically, the degree of hidden bias is greatest where the true risk is the lowest.

Advocates of the linearized multistage model argue that it offers important advantages over alternatives. For example, they say that it is more "stable" than alternative models and that this stability is a desirable trait in the face of uncertainty. In addition, proponents contend that using the same model across a variety of chemicals provides a "yardstick" for comparing relative potencies and thus for ranking relative risks. Finally, advocates of this model argue that it is prudent risk assessment practice to err on the side of caution when dealing with potentially carcinogenic substances. None of these arguments has any merit.

The observed statistical "stability" in the linearized multistage model arises because the model is insensitive to the data it is supposed to fit. Stability arises from an intentional specification error, not from any desirable characteristic of the model. By constraining the data to fit the model, risk analysts implicitly display greater scientific confidence in the model than in the underlying data.

The yardstick argument in favor of the linearized multistage model fails because it institutionalizes these systematic biases. Any rank-ordering of chemical hazards based on this model will be biased in theory as well as in practice. An especially pernicious use of the yardstick argument is the assertion that it enables government agencies to set regulatory priorities. Besides the structural bias implied in the model, further bias occurs because the model fails to take account of human exposure. This failure

virtually guarantees that regulatory priorities will be misordered. For example, in the air toxics title of the recently enacted Clean Air Act amendments, Congress gave special consideration to the cancer risks said to be associated with dioxins. It is reasonable to believe that the congressional concern about dioxin was motivated substantially by the very high potency estimate for one of those chemicals—an estimate that is widely believed by scientists to be a gross overestimate of the true risk.

Proper model specification is the foundation of modern statistical methods, so challenges to the multistage model should be expected and encouraged as better data and improved models become available. Indeed, change is a hallmark of scientific inquiry; policies that institutionalize any particular model specification effectively stifle scientific advancement.

In practice, however, use of other models is generally discouraged. For a risk assessment to be based on an alternative model, there must be substantial scientific evidence supporting the alternative. Instead of incorporating the latest scientific information and statistical procedures, current federal agency practices discourage such advancements by communicating a generic mistrust of alternatives. The resulting value judgments embedded in the multistage models were never explicitly approved by risk management officials. In many cases government officials charged with making difficult regulatory decisions are never even aware of the implicit policy judgments of staff risk analysts.

**Biases Embedded in Human Exposure Estimates.** It is a generally accepted principle of exposure assessment that estimates should be based on realistic scenarios, with appropriate consideration of uncertainty. Nevertheless, regulatory agencies often rely heavily on "reasonable worst-case" environmental conditions, base human health assessments on the so-called maximum exposed individual, and assume that exposure occurs constantly over an entire lifetime, even when it is intermittent or short-lived. Each of these assumptions tends to overstate the estimate of average human risk. In combination these biases are multiplied so that the final result is a cascade of biases that may mislead policymakers and create undue public alarm. Most disturbing, perhaps, is that excessive bias in risk assessment encourages regulatory initiatives that promise more protection from the ravages of cancer than policymakers can possibly deliver.

● *"Reasonable Worst-Case" Exposure Conditions.* When exposure data are available, they often relate to unusually sensitive environments or highly contaminated conditions. But agencies frequently use these data to estimate regional or nationwide environmental exposures under the false assumption that unusual localized circumstances apply rather generally.

In a recently proposed rule governing the allowable level of synthetic organic chemicals in drinking water, the EPA estimated the level of existing contamination by using a handful of state studies. These studies had been undertaken to measure contamination levels at previously identified "hot spots," not to characterize nationwide exposures. Nevertheless, data from these studies were extrapolated nationwide. After combining modelling assumptions, hot-spot data, and conservative potency estimates derived from the hazard assessment process described earlier, the EPA estimated a baseline cancer incidence of seventy-four cases per year. But the true incidence is very likely to be much lower simply because of the extreme environmental conditions on which nationwide exposure estimates were based.

● *The "Maximum Exposed Individual."* Risk analyses must also consider the conditions under which humans may be exposed. Actual exposure varies considerably depending on location, population mobility, and a host of other factors. But exposure estimates are often based on the "maximum exposed individual," a hypothetical person whose exposure represents the "reasonable worst case." Exposures to

---

**Excessive bias in risk assessment encourages regulatory initiatives that promise more protection from the ravages of cancer than policymakers can possibly deliver.**

---

environmental contaminants are generally assumed to occur twenty-four hours each day for seventy years. Occupational cancer risks are based on an analogous construct—a hypothetical worker who is exposed at the permissible exposure limit eight hours per day, five days per week, fifty weeks per year over a forty-five-year working lifetime. Risks to the entire exposed population are often estimated by assuming that all are exposed at levels equivalent

to the maximum exposure—a statistical absurdity that imparts a substantial and quantifiable bias.

Risk assessments focused on the drinking water pathway offer another example of exposure bias. First, adults are assumed to drink two liters of tap water per day, but the average adult consumes only 1.4 liters of all beverages per day, less than half of which is drinking water. Second, the full daily consumption of drinking water is assumed to come from the same contaminated source, but the average adult spends more than one-half of all waking hours away from home. Finally, exposure is assumed to occur for seventy years, but the average person spends just nine years at any one residence. Each of these assumptions may be plausible for a small subset of the exposed population, but the likelihood that anyone is accurately characterized by *all* three is extremely remote. Indeed, these three assumptions lead to estimates that exceed the average level of exposure by a factor of more than fifty.

The design of cleanup plans for hazardous waste sites offers another example in which biased assumptions are used to estimate human exposure. The procedures give special weight to unusually sensitive subpopulations, such as children, pregnant women, the elderly, and those with chronic illnesses. Children's exposure is generally estimated by assuming that half of nearby households include children and that one child from each household plays at the hazardous waste site. Soil ingestion exposures are based on children who intentionally eat dirt. For air exposures, all nearby residents are assumed to spend the entire day within the contaminated zone. Dermal exposures are similarly calculated on the basis of worst-case conditions and assumptions.

A common defense for these biased exposure assumptions is that risk assessments often fail to measure risks from all relevant pathways. Risk assessors thus account for what they cannot estimate by intentionally exaggerating what they can. This was the case for many years because analytic methods for some pathways were considered excessively primitive. More recently, however, federal risk analysts have worked diligently to capture multiple pathways. It is now quite common to see risk assessments that estimate risks from inhalation, dermal absorption, and ingestion through drinking water, meat, milk, home-grown vegetables, and locally caught fish. These efforts to analyze pathways comprehensively have not diminished the use of conservative exposure assumptions, however. These assumptions are simply extended to the additional

pathways. The resulting exposure scenario combines the reasonable worst case from each pathway into a mega-worst case.

● *Assumptions versus Real-World Exposure Data.* These exposure assumptions are typically used in lieu of real-world data, even when such data exist. Risk estimates are only as good as the data and assumptions used to create them, and even small biases in assumed exposure levels can result in substantial overestimates of average risk.

For example, regulatory agencies may not have statistically reliable real-world data on pesticide residues in agricultural products, and they also may not know the proportion of a given crop that has been treated with a particular pesticide. A common resolution of these uncertainties is to assume that residues are equal to the regulatory "tolerance" (the maximum level allowed to be present in food sold in interstate commerce) and that 100 percent of the relevant crop has been treated. Both assumptions are likely to overstate actual exposure, but they are encouraged by agency guidance as mechanisms intended to produce inflated estimates of risk.

When data are available, the extent of this bias becomes evident. In a recent pesticide review the EPA reduced its earlier upper-bound excess lifetime cancer risk estimate by a factor of 100 when its exposure assumptions were replaced with real-world data. The EPA then still acknowledged that upper-bound risks were probably overstated because field tests were performed on the basis of applications at the maximum legal rate and as close to harvest as

---

**Risk estimates are only as good as the data and assumptions used to create them, and even small biases in assumed exposure levels can result in substantial overestimates of average risk.**

---

the label permits. Similarly, feeding studies assumed that animal diets were dominated by feedstuffs containing relatively high residues, such as almond hulls and raisin waste. As the EPA noted, even if these assumptions accurately reflected typical animal diets, they would do so only for portions of California where almonds and raisins are grown. Nationwide extrapolations based on these unusual diets significantly overstate average exposure.

## Implications of Biased Risk Assessment for Regulatory Decisionmaking

The primary purpose of risk assessment is to provide data and analysis that can serve as the foundation for making risk management decisions. This requires the synthesis of information concerning risks and exposure levels into a coherent package that can be used to develop regulatory options. Decisionmakers can then use risk estimates as inputs in their regulatory analysis.

Unfortunately, risk information tends to be presented in ways that frustrate regulatory analysis

---

**Regulatory proposals that utilize risk assessment rarely provide information that quantifies uncertainties for decisionmakers, nor do these proposals analyze the implications of uncertainty.**

---

and mislead decisionmakers. First, the substantial uncertainties underlying risk estimation are generally discarded in favor of reporting only point estimates. Decisionmakers are thus led to believe that scientists have determined the actual level of human cancer risk. Second, the point estimates provided do not represent the expected values of the underlying risk distributions. Instead, they are laden with biases. Both of these factors imply that regulatory choices may differ systematically from what they would have been if decisionmakers had been fully and accurately informed.

**Failure to Quantify Uncertainty.** Scientists agree that uncertainties should be quantified and presented to decisionmakers as part of the risk assessment package. In practice, regulatory proposals that utilize risk assessment rarely provide this information, nor do they analyze the implications of uncertainty. Virtually all risk assessments prepared in support of regulatory decisionmaking identify only the upper-bound risk estimates.

The difference between upper-bound and expected-value estimates may be considerable. The EPA's current upper-bound risk estimate for dioxin may be 5,000 times greater than the expected-value estimate. The upper-bound risk estimate for perchloroethylene (the primary solvent in dry cleaning) exceeds the expected value estimate by a factor of

about 35,000. The significance of these distortions becomes evident only when agencies strive to avoid them. For example, the EPA's recent decision to ban asbestos relied on epidemiological data rather than animal studies and on the geometric mean from a collection of studies rather than the highest risk estimate available. These simple improvements in risk assessment combined to reduce the estimated risk of lung cancer by a factor of ten and the estimated risk of mesothelioma by a factor of twenty.

In many instances decisionmakers are not informed that risk estimates differ because of underlying methodological and policy choices. In the EPA's draft proposed rule limiting emissions from coke ovens, for example, cancer risks were estimated on the basis of the linearized multistage model described above. In previous rules involving similar types of risks, however, the EPA had used a maximum likelihood procedure designed to identify the expected value of the dose-response mode. Unsurprisingly, the linearized multistage model projected higher risks. To the extent that decisionmakers were not informed that the higher risk estimate was largely due to the use of a different extrapolation procedure rather than to any fundamental change in scientific knowledge, choices based on this risk assessment were likely to reflect misunderstanding rather than science.

Some risk estimates are so large as to defy all reason and common sense. In a recent decision to list spent wood-preserving chemicals as hazardous wastes, the EPA provided a table listing all of the contaminants in the waste stream, the levels of these constituent chemicals, and the calculated groundwater risks based on specified but arbitrary dilution and attenuation factors. When the risks posed by these individual contaminants are summed, they yield an estimated upper-bound excess lifetime cancer risk of forty-two. This implies that an individual exposed to the diluted form of this waste stream could expect to die from cancer every two years for seventy years.

**Misordered Priorities, Perverse Outcomes.** Logically, one would expect that routine exaggeration of likely risks would lead to inefficient regulatory choices. Decisionmakers, convinced that a certain substance or activity poses a significant threat to public health, may well take actions that they would otherwise resist. Nevertheless, decisionmakers would still be able to establish sensible priorities as long as all risk estimates were equally exaggerated.

Federal risk analysts are not consistent in their assessments of different risks, however. This makes it difficult to determine which activities pose the greater risks or to establish reasonable priorities for regulatory action. The bias in risk assessment is especially severe with respect to carcinogens. It is thus reasonable to expect that other health and safety risks tend to receive relatively less attention and weight than they would if different types of risk were measured more consistently. Society

---

**Risk assessment remains a seamless web of science and value judgment that is impenetrable by the average citizen and wholly lacking in public accountability. Confidence in government as a risk management institution cannot improve until risk assessment is demonstrated to be a scientific enterprise.**

---

implicitly bears greater total risks because the bias in cancer risk assessment has misordered regulatory and budgetary priorities.

Conservative risk assessments can lead to truly bizarre regulatory decisions. When the EPA established its new "toxicity characteristic for hazardous waste," the agency also identified twenty-five organic chemicals that, if detected above specified thresholds, would render a waste stream "hazardous." This designation is significant because it triggers expensive treatment and disposal requirements. Biased risk assessment procedures dictated very low thresholds for these organics.

Several months after promulgating the regulation, the EPA learned that common chlorofluorocarbons (CFCs) may contain trace levels of carbon tetrachloride and chloroform—two of the twenty-five organic chemicals listed. Further, under previously established EPA rules the act of removing these CFCs from refrigeration units for recycling made them "solid wastes." Thus, anyone seeking to reclaim CFCs rather than to vent them to the atmosphere faced a rather difficult decision. The required testing of these "solid wastes" would trigger a "hazardous waste" designation and the full weight of expensive regulation under the Resource Conservation and Recovery Act. These burdens could be avoided only by doing the wrong thing—venting the refrigerants to the atmosphere.

After discovering this problem, the EPA moved quickly to suspend the application of the toxicity

characteristic to CFCs, but the event symbolizes the perversities that can result from conservative risk assessment. As it happens, the same CFC compounds that would have been hazardous wastes if reclaimed from refrigeration units are also used as inert propellants in a variety of pharmaceuticals—including the inhalers that asthmatics rely on to breathe freely.

Finally, the use of biased risk estimates may actually increase individual risk, even in situations in which cancer is the only concern. Regulatory actions taken to address what are in fact insignificant threats may implicitly tolerate or ignore risks that are far more serious. For example, before it was banned, ethylene dibromide (EDB) was used as a grain and soil fumigant to combat vermin and molds. Vermin transmit disease, and molds harbor the natural and potent carcinogen aflatoxin B. The estimated human cancer risk from the aflatoxin contained in one peanut butter sandwich is about seventy-five times greater than a full day's dietary risk from EDB exposure. By eliminating the relatively small hazard from EDB, federal officials may have intensified the relatively potent threat of aflatoxin.

## Strategies for Improving Risk Assessment

The practice of risk assessment is extremely complex and fraught with controversy. The underlying problem is inherently difficult to analyze, and the stakes involved are enormous. Seemingly innocuous choices made in assessing risk often have huge consequences.

The problems identified here do not imply that risk assessment should be abandoned, although increasing dissatisfction with the process has intensified the pressures to do so. For risk assessment to survive as a useful component of regulatory analysis and decisionmaking, dramatic changes must occur that will restore its credibility and relevance.

**Renewed Commitment to Separating Science from Policy.** First, heroic efforts must be made to separate science from policy. Criticisms leveled more than a decade ago by the National Academy of Sciences are still unanswered. Risk assessment remains a seamless web of science and value judgment that is impenetrable by the average citizen and wholly lacking in public accountability. Confidence in government as a risk management institution cannot improve until the credibility of risk assessment as a scientific enterprise is restored.

Regulatory agencies tend to be institutionally resistant to change. Scientific advancements in risk assessment methodology that implicitly cast doubt upon earlier decisions are particularly distressing. Although this phenomenon characterizes many institutions, it appears to be especially pernicious with regard to regulatory agencies and risk assessment. Thus, a formulaic approach to risk assessment has evolved in which departures from the accepted pattern are inherently controversial simply because they are different. The process needs to be reopened to admit a wider variety of new ideas, hypotheses, and results.

**Develop Risk Distributions in Lieu of Point Estimates.** Perhaps the single most important reform needed is the replacement of upper-bound estimates with risk distributions. There are a variety of analytic methods available for estimating distributions and retaining the uncertainties of risk analysis. While these methods were computationally quite difficult a decade ago, contemporary computer technology is more than adequate for the task.

Besides enabling risk analysts to communicate uncertainty, risk distributions are compatible with efforts to incorporate all the available information. Risk assessments would be far less sensitive to individual assumptions, model choices, and data, and they would reflect scientific and statistical advancements more quickly.

The role played by decisionmakers would be enhanced in such a setting. If decisionmakers wanted to choose a very cautious strategy, they could do so and explicitly apply a margin of safety in the final decision. The public and affected parties would also benefit from knowing the full risk distribution and its expected value, rather than learning only an alarming estimate implicitly derived from the distribution's upper tail.

**Sensitivity Analysis of Major Parameters and Assumptions.** In the short run, risk assessments would be substantially improved if analysts performed sensitivity analyses on those parameters and assumptions that are believed to dominate the outcome. This is the conventional practice in benefit-cost analysis where both sides of the economic ledger are often characterized by considerable uncertainty. There is no reason why federal risk assessments should not be so rigorous as the economic analyses that agencies perform in support of regulatory decisionmaking.

## Conclusion

Risk assessment lies at a crucial stage in its evolution. Whether it will survive as a useful policymaking instrument will ultimately depend on whether the risk assessment profession responds to long-standing concerns such as those discussed here. An objective

---

**Whether risk assessment ultimately survives will depend on whether the methodology (and its practitioners) can adapt to the changing needs of policy officials and decisionmakers and can incorporate the latest advances in science.**

---

observer could well interpret the pattern of bias—as extensive, pernicious, and resistant to reform as it appears to be—as a malignant invasion of such magnitude that the organism cannot be saved. Whether risk assessment ultimately survives will depend on whether the methodology (and its practitioners) can adapt to the changing needs of policy officials and decisionmakers and can incorporate the latest advances in science.