

Who's **COMMENTARY** Exaggerating?

BY ADAM M. FINKEL

IT OUGHT TO BE THE BEST OF TIMES for someone like me. I'm a professional risk assessor—the director of health standards at the U.S. Occupational Safety and Health Administration (OSHA)—and risk assessors are much in demand these days. One reason is that the nature of the health and environmental hazards we face has changed. When rivers were bursting into flame, as Cleveland's Cuyahoga River did in 1969, or when smog was so bad it made your eyes tear, we didn't need elaborate risk assessments to tell us something had to be done. The newer generation of environmental threats are different: they're often no less serious, but they're harder to measure and harder to eliminate. When toxic chemicals leak

spectrum. For a long time the notion that the value of saving human lives could be quantified and compared with the costs of regulations was anathema to many public-interest advocates. In recent years that has changed; environmental groups have begun to realize that setting such values is not immoral (although setting them too low might be).

But the most striking recent development has been the embrace—actually more like a bear hug—that QRA has received from the right. “Regulatory reform” was one of the key provisions of the Republican Contract With America, and legislation passed by the House of Representatives last year and now pending in the Senate would require risk assessments to be carried out on an unprecedented scale. In effect, every new health and

Overcautious risk experts these days are hobbling American industry and scaring the pants off the rest of us—or so say congressional reformers. Is it really true?

onto an open field for decades, and a neighborhood grows up around that field, and some of the chemicals seem to be reaching groundwater used for drinking, should the field be cleaned up, and if so, to what extent? When the air in a factory contains potential carcinogens, should the owner be forced to retool, possibly at great expense? Quantitative risk assessment, or QRA, tries to give policymakers the information they need to answer such questions in a rational way. It tries to determine how many people are likely to get sick or die as a result of a particular hazard, and how much it would cost to save at least some of them.

QRA is a young discipline. I've been involved with it for 15 years, which is practically since its beginning. In the early years, we risk assessors took a lot of criticism from the left side of the political

environmental regulation would have to be based on a QRA certifying that the benefits of the regulation—the risks it reduced—justified its costs. I'm certainly in favor of creating more demand for risk assessors, assuming Congress also provides the resources to fund their work.

So then why do I think this is in fact the worst of times to be a risk assessor? Because for one thing, Congress is actually trying to gut the budgets of the agencies that do QRA—agencies like OSHA and the Environmental Protection Agency (EPA). But more important, the version of QRA that many in Congress, academia, and the media have embraced is a repudiation of much of what has gone before in this field. It is reform premised on a myth—namely, the myth that current assessment methods routinely exaggerate risk, at a huge cost

ILLUSTRATIONS BY FRANCES JETTER

to society. To my mind, risk assessment is in danger of being subverted just as it is coming into its own as a scientific discipline. Sometimes I feel like a chemist who thought his field was finally about to take off, only to discover the government was poised to mandate alchemy as the official state science.

THE IDEA THAT RISK ASSESSMENTS are wildly overconservative first appeared a decade ago in academic articles sporting titles like "The Perils of Prudence: How Conservative Risk Estimates Distort Regulation." It has since been adopted by any number of influential public figures. In his 1993 book *Breaking the Vicious Circle*, Stephen Breyer, now a Supreme Court justice, argued that the public irrationally desires to fear the worst and that risk assessors feed that fear. He concluded that many risks the government deems large enough to regulate are actually laughably small. In one passage, for instance, Breyer told the story of a New Hampshire company that had been ordered, as a result of an EPA action, to pay \$9 million extra to clean up a dump site so that kids could safely eat dirt there 245 days a year—even though the site was a swamp.

Many in Congress clearly believe that anecdotes like these are both accurate and representative. In speaking up for the regulatory reform bill that he and Senator Bob Dole have sponsored, Senator J. Bennett Johnston (Democrat-Louisiana) claimed that "federal agencies today . . . are not using good science, and their regulations are a disaster." The Dole-Johnston bill, like the one passed earlier by the House, would solve this purported problem by telling risk assessors how to assess risks. The allegedly "conservative" procedures we now use would be replaced with ones designed to yield the "best" estimate. The bills do not precisely define this alluringly upbeat term, but as far as I can tell, "best" means "average": it means that regulators should focus on the risk faced by the average person. And where different scientific theories lead to different assessments, we should average together the results—although the bill also instructs us to pick the single "most plau-

sible" theory, even if it yields a below-average estimate of risk.

Producing either "best" or "most plausible" estimates of risk sounds like a commonsense goal. But is it really? To decide, you need to answer two questions that critics of current QRA tend to duck. First, what is the evidence, apart from the occasional anecdote like Breyer's, that risk estimates today are routinely skewed in an overly conservative way? Second, if risk estimates really are skewed, is that a serious social problem and one for which "best" estimates are the cure?

Tackling the second question first, let's think about what it would mean to protect the "average person" from the "average risk." Most individuals are not average as far as risks are concerned; they vary greatly in their exposure and susceptibility to pollution, just as they vary



in, say, body weight. Suppose the government set a standard for wooden ladders. Would you say that a ladder that could support only a 140-pound person was the best, while one that could also support a 200-pound person was a wasteful example of an unjustified bias toward being "safe rather than sorry"? Probably not. Yet when OSHA calculates risk for a worker who works in the same industry and breathes the same pollutants for 40 years—perhaps twice the national average—it is accused of practicing "bad science." So is the EPA when it assesses the risk of pesticide residues in apple juice on the assumption that children, who may be more vulnerable than adults, may

drink three glasses a day—even though the population average is less than one. If we design our regulations to protect the average person, risk assessors reason, we may fail to adequately protect large segments of the population. Would that really be "good science"?

Then there is the question of what it would mean to estimate an "average risk." All estimates of risk involve uncertainty. For example, when you try to estimate, from laboratory and epidemiological studies, just how potent a certain carcinogen is, you can never get a single definite number; if you're honest with yourself, you'll get a range of answers. Picking the average value of that range is no more scientific than any other choice; all choices are value judgments in that they strike some balance between the health and economic costs of underestimating the risk and the costs of overestimating it. Choosing the average, as "unbiased" as that may sound, merely implies that those costs are exactly equal—which is a strong bias indeed.

Let me explain with another analogy. Suppose you are told that the average amount of time you need to get to the airport from your house is 20 minutes, but that the drive could take as little as 5 minutes or as long as 80 minutes. If you would rather be 4 minutes late for your plane than 5 minutes early, then the average estimate is the one for you. But for those of us who regard missing the plane (or allowing pollution to cause some unnecessary deaths) as more dire than having to wait a few minutes (or wasting some money on pollution controls that turn out to be overly stringent), a more prudent estimate is called for. Risk assessors traditionally set their sights on the "reasonable worst case"—that is, they try to give themselves more than an even chance of overestimating the risk in order to be reasonably sure they won't underestimate it. That's like allowing, for instance, 40 minutes to drive to the airport.

If instead of averaging the various risk estimates we simply pick the "most plausible" among them, we may be even worse off. How do we decide which of two or more plausible but competing scientific theories is most plausible? By some kind of majority vote among experts? Suppose—if I can be permitted

one last analogy—a hurricane is brewing off the coast of Florida, and two theories of hurricane behavior are at odds. Forty percent of recognized experts believe the storm will turn landward and hit Miami, while 60 percent believe it will turn harmlessly out to sea. Should we decide the latter theory is most plausible and fail to warn Miami?

Seen in that light, risk assessments that reflect more than just the average risk to the average person, and that are both plausible and conservative, begin to seem like real common sense. In fact, it's the very kind of common sense our society has applied in many other arenas. We deal with uncertain threats by being prudent; this is why there was no self-righteous clamor for "best" estimates of the exact probability of Soviet aggression during the cold war. Instead we acted on

searchers considered all 23 known carcinogens, including benzene, vinyl chloride, and asbestos, for which quantitative comparisons could be made. They found that, on average, the number of human deaths predicted by the animal studies only slightly exceeded the actual death tolls. More important, about as many of the 23 animal studies underestimated human harm as overestimated it.

Recently a twenty-fourth case study has come to light that dramatically illustrates the folly of disregarding animal test results. In 1990, OSHA was prepared to regulate 1,3-butadiene (BD), a toxic gas released in the production of synthetic rubber. Animal studies had predicted that humans exposed to just one part per million of BD in the air would have about 8 chances in 1,000 of developing cancer as a result. OSHA wanted to reduce the al-

biologists routinely sacrifice their laboratory mice and rats at 24 months, an age roughly comparable to age 70 in humans. British statistician Richard Peto has shown that if they waited for the animals to die naturally before tallying all the tumors produced by test substances, their estimates of carcinogenicity might rise sevenfold.

All in all, no one has yet succeeded in finding a systematic bias in current risk assessment procedures, apart from the desire—not always fulfilled—to protect nonaverage people. Some of the lurid tales of absurdly exaggerated risk are no doubt true; risk assessment is hard, and even well-meaning professionals will be taken by surprise sometimes. But some of the stories do not even hold up as cautionary anecdotes. Take the one about Alar, the growth regulator used by apple

We should not pretend we are promoting "good science" when we are really pushing an ideology—one that says less regulation is always better than more.

the reasonable possibility that the threat was serious enough to merit a very expensive response. And we acknowledge every day that individuals vary from the average, by building doorways high enough for tall people and ladders strong enough for stout people—although here too, as in the kind of risk assessment I do, we don't go to extremes: we don't build for the rare eight-foot, 400-pounder.

SO MUCH FOR THE IDEA that "best" estimates would naturally lead to "objective" or commonsense regulations. But the great irony of the current debate is the surprising lack of credible evidence that today's risk assessments do in fact tend to be overly conservative. It has become commonplace, for instance, to ridicule the animal studies that risk assessors use to evaluate suspected carcinogens. Pump high doses of a chemical into rats, the critics say, and of course you will overestimate how many cancers will occur in humans exposed to smaller amounts.

But is that really so? In 1988 a team of scientists in Louisiana tried to find out, by systematically comparing the results of animal studies with those of the best available cancer epidemiology studies, most of which focused on exposure to carcinogens in the workplace. The re-

lowable limit for BD from 1,000 ppm to 2 ppm. Then in the early 1990s a series of journal articles and editorials appeared denouncing that plan as regulatory overkill, on the grounds that the results in mice were irrelevant for humans. To its credit, the rubber industry continued an epidemiological study of its BD-exposed workers. Several months ago the preliminary results of that study were released—and they seem to suggest that the workers have developed as many cases of cancer as one would have predicted from the animal data. Now the rubber industry and the labor unions are urging OSHA to issue a regulation immediately and to drop our allowable level of BD to one ppm, half of what we had originally proposed. If we had simply heeded the animal studies six years ago, we might have prevented some cancers.

To be sure, risk assessors do make assumptions that tend to exaggerate risk. For example, they usually assume that subjects are exposed to a suspected carcinogen 24 hours a day instead of, say, 8. The pending legislation would allow court challenges to any regulation that uses that assumption, on the grounds that it typically introduces a threefold overestimate of exposure. But dozens of assumptions go into a risk assessment, and critics tend to ignore the ones that cut in the opposite direction. For example, tox-

growers, which was withdrawn from the market in 1989 when an EPA ban loomed. That case has become notorious as an example of overzealous regulation. But in 1991 an animal study sponsored by Alar's manufacturer found that it caused as many tumors as the EPA had assumed—and at lower doses.

Or consider Justice Breyer's fable of the "nonexistent dirt-eating children" in New Hampshire. In fact the waste site was not in a swamp; it was on undeveloped land, but land that was zoned for residential development. Nor was the EPA assuming, as Breyer implied, that children would be spooning dirt into their mouths 245 days a year when it ordered additional cleanup work. It was assuming only that trace quantities of contaminants would inevitably find their way into the systems of any children who might one day play in the contaminated dirt. Toxicity studies suggested that those trace quantities would be enough to put such children at unacceptably high risk. Was the \$9 million cleanup worth it, given how little money is available these days to address all sorts of other social problems? I'm not sure. But the fault, if there is one, lies with the response to the risk assessment, not with the science itself.

Of course, if the cleanup had cost only \$9,000, Justice Breyer would presumably not have gotten exercised about it—and

if it had looked like it would cost \$9 billion, the EPA would have had to move on to other things. To make sensible cost-benefit decisions, obviously, you need to know the costs as well as the benefits of the action you're contemplating. The legislation Congress is now considering prescribes dozens of rules for risk assessment without mentioning problems of cost analysis at all. But figuring out the true cost to the economy of health and environmental regulations is as hard and as fraught with uncertainty as risk assessment itself. And the evidence that costs are routinely exaggerated is in fact far stronger than the evidence that risks are.

MEASURING THE DIRECT cost to a company of complying with a regulation is the easy part of cost assessment, yet often the assessors even get that wrong. Usually they are forced to estimate the price of compliance technology before there is any demand for it—that is, before vendors of the technology have faced any incentive to reduce its price, before users have learned to use it efficiently, and before either group has had a chance to develop entirely new ways to comply. Worse, regulatory agencies must often rely in part on engineers at the regulated companies themselves for cost estimates—and those employees, obviously, are averse to underestimating how much complying will cost their companies. Interestingly enough, the regulatory agencies have the same aversion: if they overestimate costs, they will be less vulnerable to court challenges for having incorrectly declared a regulation economically feasible.

For all these reasons, the direct costs of regulations tend to be overestimated when they are first imposed. For example, one of the most costly EPA programs involves controls on nitrogen oxide emissions from factories and power plants. The manufacturers of the emission control equipment recently reported that it now costs between a fifth and a half what regulators had initially estimated. And in a comprehensive study last fall, the now-defunct congressional Office of Technology Assessment examined the costs of seven major regulatory programs mandated by OSHA. In no cases had regulated companies spent much more than OSHA had predicted—and in five of the seven cases they had spent substantially less.

Moreover, the direct cost of comply-

ing with regulations is only the beginning of the story. When OSHA regulated vinyl chloride, a carcinogen used in plastic production, the controls were costly, but they paid for themselves by increasing the amount of valuable product recovered. When it regulated cotton dust, controls that saved hundreds of thousands of workers from brown lung disease also helped modernize and revitalize the declining U.S. textile industry. The Alar case is another example of misreported cost estimates: while many growers of the two apple varieties most dependent on Alar, Red Delicious and McIntosh, did suffer economic hardship for several years after the withdrawal, growers of other varieties enjoyed a boom in sales. Overall, consumer demand for apples and the apple industry's profitability have nearly doubled in the years since the



carcinogenic chemical was abandoned. The losses suffered by some growers should not be dismissed, but they are not the whole story. That some parts of a regulated industry may actually benefit from regulations—as do other industries altogether, such as manufacturers of pollution-control equipment—is not usually captured by the methods economists have for analyzing costs.

Economists have had more than two centuries to develop those methods, yet they are still not very good. Why then are we so impatient with risk assessors? The most recent major study of risk assessment—a three-year effort by 24 experts (including me) chosen by the Na-

tional Academy of Sciences—concluded that “risk assessment methods are fundamentally sound, despite often-heard criticisms.” If the members of Congress who are now trying to rewrite those methods by political fiat were aware of that study, they were apparently not impressed.

Risk assessment, as I said at the outset, is a young field. It is so young that it is still severely constrained by a dearth of qualified practitioners. To do risk assessment well, you need to know something about toxicology, environmental chemistry, statistics, physiology, and several other fields. Specialists abound in most of these individual disciplines, but they have trouble talking to one another, and generalists are scarce indeed. In the entire United States, for example, there is currently only one individual with a Ph.D. in risk analysis—an undergraduate classmate of mine. We need more like him. On that point I agree with Justice Breyer: he suggested creating a cadre of risk experts to rotate through regulatory agencies, the judiciary, and congressional offices, helping to improve the whole field as they improve individual regulations. Only if we begin working on the supply of those experts now can risk assessment ever hope to catch up with society's growing expectations for it.

In the meantime we should not ask of risk assessment more quantity or quality than it can yet deliver. Above all, we should not pretend we are promoting “good science” when we are really pushing a political ideology—one that says less government regulation, at least where health and the environment are concerned, is always better than more. It is not that there is anything wrong with value judgments; risk assessment cannot be done without them. It is just that those value judgments should be made explicit and not be allowed to masquerade as objectivity. Here are my values, the ones that got me into this business in the first place: I believe that risk assessment, as it is now practiced and as it is steadily being improved, can help us protect health and the environment more cheaply and efficiently and prevent unnecessary injuries, illnesses, and deaths. And I believe that whatever society decides about how much it is worth to save a life or protect an ecosystem, the real “best” risk assessment is one that encourages decision makers to be honest about uncertainty and to make smart—and humane—responses to it. □