Environmental and Occupational Medicine,
Fourth Edition, edited by William N. Rom, MD, MPH
Lippincott–Raven Publishers, Philadelphia © 2006.

CHAPTER 112

Risk Assessment and Economic Evaluation

James K. Hammitt

Risks to health and life are ubiquitous and diverse, ranging from the risks of being struck by a car while crossing a street; developing cancer from exposure to contaminants in air (e.g., radon, tobacco smoke), water (e.g., arsenic), or food (e.g., some pesticides, some naturally occurring food components); natural hazards (e.g., tornado, hurricane, tsunami, earthquake); fires in buildings or forests; infectious diseases (e.g., AIDS, SARS, tuberculosis, malaria); global climate change; failure of engineered systems (e.g., collapse of dams, bridges, escaping radiation from a nuclear power plant); and even the risk that a comet or meteor may strike the earth.

Risk analysis is a form of structured thinking or systems analysis developed to help grapple with these risks. The basic concept is to quantify risk—both the probability and the severity of potential harm (or benefit)—and to use this information to help (a) evaluate alternative actions that may be taken to mitigate the risk and (b) determine if the benefits of an action (in reducing risks) justify its costs. Decisions about government regulations and other social policies toward environmental and occupational risks are often informed by using economic evaluation (either benefit-cost or cost-effectiveness analysis) to account for and compare the benefits and costs of risk-reducing actions. Risk analysis can be divided into risk assessment (quantifying risk), risk management (evaluating and choosing actions to mitigate risk), and risk communication (eliciting information and public values about risks and describing the results of risk assessment and risk management in ways that are sensitive to how humans perceive risks). The concepts and methods of risk assessment are influenced by the need to provide information for risk management, economic evaluation, and risk communication. This chapter provides an overview of risk assessment and economic evaluation.

Characterizing risk by the probability and severity of potential outcomes is consistent with the methods of decision analysis—a field of decision or management science concerned with making choices when the consequences are uncertain (1-4). Decision analysis views the consequence of a decision as depending on two factors: the choice that is made (i.e., the "act" selected by the decision maker from among the possible alternatives) and factors beyond the decision maker's control (i.e., the "state of nature"). In general, the decision maker cannot choose consequences but can only choose from among alternative probability distributions ("lotteries") on possible consequences. Moreover, she cannot avoid the decision, since declining to make an explicit choice also leads to a lottery on consequences—the one associated with the default choice (e.g., continuing the *status quo* policy).

In making her choice, the decision maker should consider both the probabilities and the severity of the consequences—neither factor should be excluded. Consider the choice of whether to face a risk of harm in order to obtain some sure benefit. No matter how severe the worst consequence, it may be worth taking the risk if the probability of the worst consequence is small enough. For example, even though jogging entails the risks of being hit by a car or suffering a heart attack, taking those risks (if their probabilities are sufficiently small) may be justified by the health benefits of exercise. Similarly, even if the probability of harm is large, it may be worth taking the risk if the severity of the harm is sufficiently mild (such as getting blisters or muscle strains from jogging). Decision analysis shows that a decision maker can maximize her well-being by choosing the lottery that has the highest expected utility, i.e., the choice for which the sum over all possible outcomes of the utility of the outcome, multiplied by the probability of its occurrence, is at least as large as the corresponding sum for the alternative lotteries. (The utility of an outcome is a numerical measure of its desirability to the decision maker relative to the other possible outcomes.)

Efforts to simplify decision making by considering only the probability or severity of harm are often proposed but are generally inadequate. Examples include the concept of an acceptable or "*de minimus*" risk level and the maxi-min decision criterion. The notion of a *de minimus* risk level is that some probabilities of harm are so small there is no need to consider them, regardless of the severity of impact and regardless of the benefit of facing the risk (equivalently, the cost of reducing it). For example, a one-in-a-million increase in the lifetime risk of developing cancer is sometimes suggested as a *de minimus* risk. But if there is no benefit

to facing this risk (i.e., if it could be eliminated without sacrifice), then even such a small risk should be viewed as unacceptable. If, however, the very small probability means that the benefits of mitigation are likely to be exceeded by the initial costs of even evaluating the risk and deciding how to mitigate it, then the risk would be acceptable. The key here is that the reason for accepting the one-in-a-million risk would be that the costs of addressing it exceed the benefits of doing so—not that the low probability by itself warrants inattention. The severity of the potential outcomes matters as well: clearly, a one-in-a-million probability could be quite important and worth addressing, such as a one-in-a-million chance of cancer to which billions of people are exposed, or a one-in-a-million chance of a large asteroid striking the earth in the near future.

Other approaches focus on severity of outcome and neglect probability. The maxi-min decision criterion proposes that one should choose whichever action has the best outcome in the worst case, i.e., minimize the possible harm regardless of its probability. This criterion implies one would never choose an action offering a very small probability of a slightly worse outcome even if it offered a large probability of a much better outcome than an alternative choice. For example, one would never take an airplane flight to another city, because the severe impact of the airplane crashing (despite its tiny probability) would outweigh the more modest gain of the pleasant experiences or new job one could enjoy in the new city (despite their much higher probability). Consistently following a maxi-min rule is likely to lead to accepting lotteries in which the probability of the worst outcome is relatively large, the potential benefits are comparatively modest, and hence to rather mediocre consequences on average—"nothing ventured, nothing gained."

Risk assessment is the component of risk analysis concerned with characterizing possible harms (or benefits) and their probabilities. Risk analysis can be applied to risks to individuals or populations. It is often used in support of government policy decisions, where decision makers are interested in both the average risks in a population and the distribution or variability of risk, including in particular the characterization of risks to the most highly exposed and most sensitive individuals. Population risks are often described in terms of the expected number of cases (e.g., fatalities or cancer cases), but this is a potentially misleading shorthand form of expression. Risk is fundamentally an issue of uncertainty and probability. *Ex ante*, and for many environmental and occupational risks *ex post*, the identities of the individuals who suffer the adverse outcomes are unknowable.

The following sections describe some of the concepts and methods used to estimate probabilities of harm, focusing on risks associated with exposure to chemicals and radiation which are often of concern for environmental and occupational health, and to value changes in health risks so that fatal and non-fatal risks can be aggregated and compared with mitigation costs (5-8). They are followed by a description of the principles underlying economic evaluation and an introduction to the methods used to estimate benefits and costs.

Risk Assessment

The process of characterizing health risks associated with exposure to environmental agents (e.g., chemicals, radiation, physical forces) is conventionally divided into four components proposed by the National Research Council (9):

- *Hazard identification*: the determination of whether a particular agent or activity is or is not causally linked to particular health effects.
- *Dose-response assessment*: the determination of the relation between the magnitude of exposure and the probability of occurrence of the health effects in question.
- *Exposure assessment*: the determination of the extent of population exposure to the hazard, before and after application of regulatory controls.
- *Risk characterization*: the description of the nature and often the magnitude (probability) of risk, including attendant uncertainty.

Hazard identification involves determining what health effects can be produced by the agent, under what conditions. For example, an agent may cause cancer or other disease, may be acutely lethal, or may cause some form of nonfatal illness or reproductive anomaly. Agents can be characterized according to the effects they can produce, e.g., carcinogens (cancer), teratogens (birth defects), mutagens (genetic mutations). Hazard identification for chemicals often relies on short-term tests of laboratory animals or micro organisms, such as the Ames test for mutagenicity. In general, the results of these tests and their implications for risk to humans are not definitive and so it is necessary to consider the weight of evidence provided by different experiments and theories. The result of a hazard assessment is typically in the form of a qualitative statement about the possible health consequences of exposure to an agent and the conditions under which these effects may occur.

Dose-response assessment involves determining how the type or severity of the response (i.e., the health effect) and its probability depend on the degree of exposure of the individual to

the agent. Dose-response assessment is critical to characterizing risk for, as the 16th century physician Paracelsus recognized, “All substances are poisons; there is none which is not a poison. The right dose differentiates a poison and a remedy.” (7).

For carcinogens, the type and severity of the cancer that may be caused are generally assumed to be independent of dose, but the probability of developing cancer is usually assumed to be proportional to dose, which implies a positive probability at all exposure levels above zero. This “linear no-threshold” model was initially motivated by the notion that a single genetic mutation can initiate a process that leads to development of cancer, and even one molecule of a carcinogen (or one quantum of radiation) could cause the critical mutation. In other cases, there may be a threshold dose below which the probability of harm is zero and above which the probability or severity of harm increases with increasing dose. The threshold model represents a situation in which the body’s immunological or other defense mechanisms can effectively protect against low doses but are overwhelmed by higher doses of the agent. A third possibility is a “hormetic” dose-response function, in which a low dose of the agent causes (or has some probability of causing) a beneficial effect but larger doses cause harmful effects with increasing severity or probability as the dose increases (for example, drinking alcoholic beverages may be beneficial at low doses but harmful at high doses).

In evaluating dose-response relationships, it is important to distinguish the incremental dose associated with a particular environmental or occupational exposure from the absolute or total dose including background exposures. Whenever the change in dose is small compared with the absolute dose the assumption of a linear response is generally adequate for estimating the change in risk (10, 11).

In some cases, risk assessment purports to identify a “safe” dose or exposure, i.e., an exposure or dose below which the probability of an adverse effect is negligible. For example, the EPA “reference dose” (RfD) is “an estimate (with uncertainty spanning perhaps an order of magnitude) of daily exposure to the human population (including sensitive subgroups) that is likely to be without an appreciable risk of deleterious effects during a lifetime” (12). In cases where the dose-response relationship has a threshold, the threshold dose satisfies this criterion but in cases with no threshold, where the probability or severity of harm are positive for all dose levels, identification of a “safe” dose depends on the definition of “safe.” In selecting an RfD, the interpretations of “likely to be,” “appreciable risk,” and “deleterious effects” are critical. As

discussed above, it does not appear to be useful to attempt to define safety or acceptable risk without considering the potential benefits of incurring the risk (i.e., the costs of mitigating it); this is a question of risk management. An additional difficulty with using some estimate of a "safe" dose rather than a dose-response function is that by drawing a dichotomy between safe and not safe, it provides no information about how the magnitude of the risk depends on changes in exposure, either from levels above the "safe" dose to levels below it, from one level above the "safe" dose to another level that also exceeds it, or from a level that is below the "safe" level, to an even smaller level.

Exposure assessment involves determining the quantities of the agent to which individuals are exposed, together with any other information about the route and time pattern of exposure that influence the probability that the exposure will cause the identified harms. For some agents and hazards (e.g., carcinogens and development of cancer), the probability of the adverse effect is usually treated as a function of cumulative exposure over an individual's lifetime. For others (e.g., carbon monoxide and mortality), the maximum exposure over a short period of time may be more relevant. The effect may also depend on the exposure route, i.e., whether an individual is exposed to the agent through the air she breathes, food and water she consumes, or through dermal contact. In principle, it may be necessary to assess a different form of exposure for each health effect.

The exposure analysis and dose-response analysis must be linked so that one can determine the probability and severity of harm for each exposure. This requirement has implications for the definition and measurement of exposure and dose. In some cases (e.g., ingestion of a food-borne contaminant), the concepts of dose and exposure may be treated as synonyms and the link is direct. If health effects depend on the total exposure over a long time period, exposure and dose might be measured by the average daily ingestion rate, e.g., μg of the agent per day. If exposure over a short period is of concern, exposure and dose might be measured by the mass of the agent ingested in a single short-term exposure, e.g, μg of the agent.

When the concepts are distinguished, exposure generally refers to the quantity (and time pattern) of the agent in environmental media where it comes into contact with an individual (e.g., the concentration of the agent in drinking water or in ambient air) and dose refers to the quantity that enters the individual organism or reaches a target internal organ (e.g., the mass of the water contaminant that is consumed or the mass of the air pollutant that penetrates to the lungs). Dose

is frequently estimated from exposure by assuming a standard intake rate for the exposure medium, e.g., a breathing rate of 20 m^3/day of air and a consumption rate of 2 l/day of water. The US EPA provides a large number of factors useful for calculating exposure and dose (13).

In some cases, dose is not explicitly considered and risk assessment substitutes an exposure-response function for a dose-response function. For example, assessments of the risk from exposure to air pollutants such as fine particulate matter and ozone often define exposure as the average concentration of the agent in ambient air and employ an exposure-response function which characterizes the probability of death or other health effects as a function of either the average daily concentration or the peak concentration of the pollutant over a few hours. This approach ignores differences in the quantity of the pollutant taken into the body that result from interpersonal differences in physiology or behavior, such as differences in breathing rate. When exposure and dose are distinguished, it is necessary to have a method to quantify the dose that results from each pattern of exposure.

Risk characterization involves aggregating the results of the other three components to describe the risk, i.e., the potential health effects and their probabilities of occurrence, including uncertainty about the effects and probabilities. In characterizing risk, it is often useful to distinguish between variability and uncertainty. Variability is a characteristic of the risk and affected population which refers to differences in risk among people, over time, or by specific characteristics (e.g., age, sex, smoking status). Uncertainty is a characteristic of one's state of knowledge which refers to limitations of the theories and data available for assessing the risk. In principle, uncertainty can be reduced and possibly eliminated through research, but variability is not affected by changes in information (the variability of risk in a population can be uncertain, of course).

Risk characterization can include valuing different health effects by putting them on a common scale which characterizes their severity or the importance to people of avoiding them (see the subsection on valuation below). In some cases, risk characterization may be solely qualitative, but quantitative estimates are more useful for determining how important a risk may be and the merits of alternative mitigation measures.

Data and Methods for Characterizing Risk

The data and methods used to characterize the dose-response function and exposure pattern depend on the agent and health effects of concern. This section describes methods for dose-response and exposure assessment.

Dose-Response Assessment

Estimates of the dose-response function may be based on data from humans, either experimental or observational, or on data from experiments on animals. In many cases, it may be necessary to extrapolate from the dose levels and other circumstances in which data are obtained to the circumstances that are relevant for assessing the risk of concern.

Consider three fatal risks and the types of data and extrapolation required to estimate the dose-response function: traffic crashes, fine particulate air pollution, and a carcinogenic chemical. In the case of traffic crashes, the probability and the number of people exposed to risk are sufficiently large that numerous data are available from which one can estimate the average fatality risk in a population (e.g., among the United States population of 300 million there are about 40,000 traffic fatalities per year, including 5,000 pedestrians and cyclists, yielding an average annual risk of slightly more than one in 10,000). The annual risk to an individual obviously varies with many factors, such as how much time he or she spends in a vehicle or as a pedestrian and the conditions in which he or she is exposed (e.g., type of road, traffic volume, time of day, seatbelt use). It also varies over time, e.g., with weather conditions and driver's blood-alcohol level. For traffic fatalities, the data are sufficiently rich that statistical models can be applied to estimate how the risk depends on many factors, and the data are sufficiently relevant to the risk that little extrapolation is necessary.

Note however that, even in this case, one must extrapolate from data observed in the past to the risk incurred in the future, which requires an assumption that the future will be like the past, i.e., that unmeasured factors not included in the statistical model will change slowly, if at all. If one is trying to estimate the average risk of being killed in a traffic crash next year, factors that influence driving practice (e.g., changes in speed limits, enforcement of drunk-driving laws, public-awareness campaigns), the quantity of traffic (e.g., changes in fuel prices), the mix of vehicles on the road (e.g., substitution of sport utility vehicles for automobiles), or other factors could cause the future to differ from the historical experience. Note also that although one can

estimate the average risk for a subpopulation (e.g., 25 to 35 year old men living in Los Angeles), one cannot accurately estimate the risk for a specific individual because important information about his driving behavior and other factors may be unknown (e.g., he may be more cautious or reckless than the average for his group) or because factors like these and their effects on traffic risk cannot be accurately measured.

Particulate air pollution (often measured as $PM_{2.5}$, the concentration of fine particulate matter with aerodynamic diameter less than 2.5 microns) is estimated to cause perhaps 60,000 deaths per year in the United States (14). Unlike traffic crashes, however, it is impossible to know this number very accurately because one cannot count the deaths caused by particulate matter air pollution and so estimates are sensitive to uncertainty about the slope of the exposure-response function for observed exposure levels and to assumptions about the shape of the exposure-response function for exposure levels below the observed levels (e.g., whether or not there is a threshold exposure below which there is no risk). The proximate causes of death—heart attacks and other cardiovascular events—also occur for reasons unrelated to air pollution and these deaths cannot be distinguished from air-pollution-related deaths.

Estimates of the risk and the effects of reducing exposure are based on epidemiological studies that compare the frequency of deaths in populations exposed to different levels of particulate air pollution. Because these data are observational and many other factors are also at play, there is concern about whether an observed relationship between air pollution and frequency of deaths is causal: correlation does not imply causation. In addition, because the background fatality risk (i.e., the fatality risk in the absence of particulate air pollution) is positive, the estimate is of the “incremental” (above background) fatality risk associated with air pollution.

Although the possibility that an observed correlation is due to chance can be effectively ruled out with sufficient independent data, a more serious problem is that the correlation may be due to confounding variables that are correlated with air pollution and causally related to mortality rates. For example, if one attempts to estimate the effect of air pollution on mortality rate by comparing individuals living in different cities having different levels of pollution, it is important to identify and statistically control for individual factors that might be spatially correlated with air pollution levels, such as smoking behavior, stress, and obesity. Cohort studies, which collect data on individuals and follow them over time, provide greater ability to control for

individual characteristics than cross-sectional studies which compare population averages. Another type of epidemiological study compares the number of deaths that occur each day with the pollution levels observed on that day and several preceding days within a city. This time-series design has the advantage that the characteristics of the population do not change day by day and so the only possible confounding variables are those that vary in ways that are correlated with pollution levels, i.e., weather conditions and day of week.

Logically, one could estimate the dose-response function by conducting an experiment using human subjects but there are serious ethical and practical difficulties with this approach. Some experiments are conducted, but these are restricted to situations in which volunteers are exposed to elevated pollution levels for short periods of time and it is believed that effects are temporary and reversible. In other settings, such as development of new pharmaceuticals, experiments using human volunteers who are randomly assigned to treatment or control conditions are frequently employed (ethical concerns are ameliorated in part by the possibility that the human test subjects could themselves benefit from the experimental findings). These experiments are often conducted as double-blind trials, in which neither the subject nor his treating physician knows whether he has received the treatment or the control condition, and so knowledge of this factor cannot affect behavior in ways that would confound the results (e.g., if subjects in the control group were provided with supplementary treatments not available to those in the treatment group).

A serious drawback to relying on data on human responses, whether experimental or epidemiological, is that such data require that a substantial number of humans be exposed to the agent, and to doses large enough that the average probability of harm is measurable. For traffic crashes and particulate air pollution, these conditions are, regrettably, satisfied in the general population. In other cases, the doses experienced by the general population are too small to produce effects that are measurable against background mortality or disease rates, but subpopulations of workers or others may be exposed to sufficient doses to provide useful estimates. In these cases, it is necessary to extrapolate the risks from the relatively high levels at which these subpopulations are exposed to the lower levels at which the general population is exposed (issues associated with extrapolation from high to low dose are discussed below). For newly synthesized chemicals and other agents, human data are not available.

When human data are not available, dose-response functions can be estimated using data from animal experiments. Use of animal data has the advantage that it does not require exposing humans to potentially hazardous doses, but it obviously puts the experimental animals at risk and there is uncertainty about how best to extrapolate from the animals' responses to predict human responses.

Data for many chemicals come primarily from animal experiments. Dose-response assessment is typically conducted using different approaches for carcinogens and non-carcinogens. For carcinogens, a dose-response function is developed that relates excess (above-background) risk of developing cancer over a lifetime to chronic dose (i.e., a constant dose administered daily over the lifespan). For non-carcinogens it is conventional to estimate only a reference dose, below which the risk is believed to be negligible, and not a full dose-response function. This difference in approach results in part from theories and evidence which suggest that dose-response functions for cancer have no thresholds, while the functions for other health effects do. More recently, with improved understanding of carcinogenic mechanisms some carcinogens have been considered to have thresholds, and some non-cancer endpoints (like cardiovascular impacts from airborne particulate matter) are addressed with full dose-response functions, so this strict delineation does not always hold. In general, dose-response assessment is designed to be conservative (to err on the side of safety), i.e., to overestimate the risk associated with any dose and to underestimate the dose associated with any risk.

Dose-response functions for chemicals that may be human carcinogens come primarily from long-term bioassays such as those conducted for the US National Toxicology Program (15). In a long-term bioassay, the experimental animals (typically certain genetically homogenous strains of mice or rats) are exposed to a fixed quantity of the agent being tested every day of their adult lives. The agent may be administered through feed, drinking water, air, or gavage (a process in which a measured quantity is injected down the animal's throat, ensuring a known dose). Animals are sacrificed near the end of their natural lifespan (approximately two years), dissected, and the numbers of tumors in each tissue are counted. By comparing the numbers of animals with tumors (or the number of tumors per animal) between animals administered different doses, the relationship between dose and probability of developing tumors can be estimated.

Obviously, mice and rats differ from humans in a number of characteristics, including size, longevity, metabolic rate, and morphology. For example, mice and rats have Zymbal glands in which tumors can develop. Because humans do not have these tissues, it is unclear whether humans are safe from chemicals that cause only Zymbal gland tumors in rodents or whether humans would develop tumors at other sites if exposed to these chemicals. Moreover, the conditions under which the experimental animals live and are exposed to the agent (a constant amount every day) differ dramatically from the conditions in which humans are exposed (varying amounts of many different agents).

To maximize the chance of finding a statistically significant relationship between dose and probability of cancer (if one exists), experimental animals are exposed to large doses, limited only by the need to keep the dose small enough to avoid killing the animals or having other major effects on their health. Typical doses for the exposed groups are one-half and one-quarter of the "maximum tolerated dose," i.e., the largest dose that does not produce substantial immediate adverse effects. About half of all chemicals tested in chronic bioassays produce carcinogenic responses in some species. In part, this apparently high rate may reflect a selection effect, since chemicals that are unlikely to be carcinogenic are less often tested in chronic bioassays. In addition, it has been proposed that the rate may result from mechanisms associated with the high doses used in rodent bioassays that would not be important at lower environmental doses (16).

Use of animal bioassay data to estimate a dose-response relationship in humans requires two types of extrapolation—from animal to human, and from high to low dose. Both require a model of how the differences, in species or dose, affect the response.

Interspecies extrapolation is perhaps best addressed through toxicokinetic and toxicodynamic modeling. Toxicokinetic models characterize the process by which the agent to which an organism is exposed is absorbed, metabolized, distributed through the body to the organ(s) where its toxicity is expressed, and eliminated from the body. Toxicodynamic models characterize the relationship between the dose delivered to the target organs and the development of cancer or other effects. These models can account for non-linear relationships between exposure and the dose delivered to the target organ but require extensive data on how the agent acts within the organism. In the usual case, when reliable models are not available, interspecies extrapolation is accommodated through the choice of dose metric, i.e., the method used to

quantify the doses to the experimental animals and to humans that are expected to produce similar effects. For chronic oral exposure, the recommended approach is to characterize dose in terms of mass of the agent per day divided by body mass to the three-fourths power, e.g., mg/(kg$^{3/4}$-day). Scaling by the three-fourths power of body mass is consistent with the scaling of various physiologic processes and with empirical comparisons of toxicity across species (17). Previously, doses were scaled in proportion to body mass or body mass$^{2/3}$ (the latter was motivated as a method of comparing the ratio of dose to body or organ surface area).

Extrapolation between high and low doses is generally conducted by linear interpolation between the origin (zero dose, zero incremental effect) and a "point of departure" (POD) which is conceptually the smallest dose at which the data provide reliable estimates of risk without significant extrapolation. The POD is generally a point on a fitted statistical relationship between dose and probability of cancer (actually, in the interest of conservatism, the lower end of a confidence interval for such a point). In an animal bioassay, the POD might correspond to the estimated dose at which the incremental cancer risk is 1 percent or 10 percent (17). In accord with the preference for conservative estimates, when data from multiple species and both sexes are available, it is conventional to use whichever species/sex combination yields the smallest POD. Linear approximation is generally believed to provide a conservative estimate of risk because it neglects the possibility that defense mechanisms may protect against very small doses yet be overwhelmed by larger doses. As described above, linearity between the probability of cancer and dose is expected for carcinogens that can cause the genetic mutations that may lead to cancer, and also where the agent increases a background risk of cancer or other endpoints.

For non-carcinogens, conventional practice is to identify a reference dose (RfD), defined as a dose below which adverse effects are unlikely to occur, even among the most sensitive humans. An RfD can be estimated by identifying the largest dose at which there is no statistically and biologically significant increase in frequency or severity of adverse effects compared with the control group (the "no observed adverse effect level" or NOAEL) or the smallest dose at which such effects are observed (the "lowest observed adverse effect level" or LOAEL). More recently, the benchmark dose (BMD) has been developed as a substitute for the NOAEL and LOAEL with better statistical properties (e.g., it is less sensitive to the exact doses chosen for an experiment and to the random variation in response at a single dose). A BMD can be obtained by fitting a statistical model to the experimental dose-response data and using it to identify the

lower end of the confidence interval for the dose that is predicted to pose little or no risk. The BMD, NOAEL or LOAEL is converted to an equivalent human dose using an appropriate dose metric and then divided by one or more "safety factors" (typically equal to 10) to account for factors such as uncertainty about appropriate interspecies scaling, greater variability of sensitivity among humans than among the experimental animals (which are genetically homogenous and experience nearly identical living conditions), the use of a subchronic (less than lifetime) rather than chronic dose in the experiment, the use of a LOAEL rather than a NOAEL, and other factors (12). When data are available from experiments using different species or sexes, or multiple adverse effects are observed, it is conventional to choose the smallest of the corresponding BMDs, NOAELs or LOAELs, again in pursuit of conservatism.

Exposure Assessment

The goal of exposure assessment is to estimate or predict the levels of the agent to which humans may be exposed. Exposure assessment can be based on direct measurement, mathematical modeling, or a combination of these approaches (13, 18, 19).

Exposure to some agents can be estimated by directly measuring exposures to a sample of individuals, for a sample time period, and using these to estimate exposure to a larger population. For a food constituent or contaminant of concern, one could have subjects maintain a diary reporting the quantities of the relevant foods they eat and preserve samples of each serving that can be chemically analyzed to determine the concentration of the agent. For some airborne chemicals and radiation, personal exposure monitors have been developed that can be worn by subjects and provide a continuous or cumulative record of the quantities of the agent in the air around them. By continuously monitoring the air or radiation around a subject, these monitors provide a means of accounting for variation in exposure as the subject moves among different microenvironments (e.g., the home, workplace, outdoors along a street). Exposure to some agents can be estimated using biomarkers, which are measurements of the organism that provide a record of cumulative exposure, such as concentrations of heavy metals (e.g., lead, mercury) in blood, hair, or nails.

Other measurement approaches provide less detail on the time pattern and interpersonal differences in exposure. Exposure to air pollutants is often estimated using fixed-site monitors located in and around cities. Exposure to individuals living in a city can be estimated using the measured values at the monitors. A simple approach would estimate each individual's exposure

as equal to the concentration at his residence, which can be estimated as equal to the concentration at the nearest monitor or interpolated from several monitors. A limitation of this approach is that it does not account for differences in pollutant concentrations among locations, including differences between indoor and outdoor concentrations, and for differences in inhalation rates.

Accounting for time spent indoors and outdoors can be important because the differences in pollutant concentrations can be large, and either higher or lower indoors than outdoors depending on the pollutant, source, and other factors. For occupational exposures, exposure may occur only in the workplace (although exposure to the worker and members of his household could also occur if the agent is transported from the workplace, on clothing for example). In industrialized countries, most residents spend the great majority of their time indoors. Microenvironment modeling uses information on residents' time-patterns of location (e.g., residence, workplace, in motor vehicle, outdoors) to account for measured or estimated differences in concentrations among these locations and estimates cumulative exposure over a time period by weighting the concentration in each location (at the relevant time of day and day of week, if available) by the time spent there.

Conservative estimates of exposure are defined by considering the "maximum exposed individual" (MEI), a typically hypothetical person constructed using assumptions intended to estimate the greatest possible exposure. For example, the MEI may be assumed to breathe the highest concentration of pollutants in the air near an emissions source 24 hours per day for a 70 year lifetime, or the MEI may be assumed to be a small child who eats contaminated soil every day of the year.

Environmental fate and transport models are often used to estimate the effects of releasing an agent to the environment, or changing the rate of an ongoing release (e.g., to predict the effects of a regulation that reduces atmospheric emissions or discharge to surface waters). These models take the form of large and sophisticated computer programs that account for physical, chemical, and biological processes that affect the quantity of an agent in various environmental compartments or locations. For example, airborne fine particulate matter results in part from the exhaust of fossil-fuel combustion in electric-generating plants, motor vehicles, and other sources. One part of the fine particulate matter consists of particles released as part of the exhaust stream ("primary" particulate matter). Another part is the result of atmospheric processes

through which exhaust gases (sulfur oxides and nitrogen oxides) undergo chemical reactions and condense to form particulate matter ("secondary" particulate matter). Particulate matter and its gaseous precursors are transported by wind for hundreds to thousands of kilometers before depositing to land or surface waters either from gravity (dry deposition) or through rain and snow (wet deposition). Human exposure results when air containing particulate matter is inhaled.

Atmospheric emissions of mercury from coal-burning power plants are also transported by wind and deposit in wet and dry forms, with typical atmospheric residence times and transport distances depending on whether the mercury is released in elemental, reactive gaseous, or particle-bound states. Upon reaching surface waters, via deposition or runoff from land, micro-organisms in the water or sediments methylate some of the mercury forming methyl mercury. Methyl mercury accumulates up the food chain reaching its highest concentrations in top predator species such as tuna, swordfish, shark, and northern pike. Human exposure results from consumption of fish. The effects of changing emission levels on human exposure can be estimated using fate and transport models to estimate the relationship between mercury emissions and methyl mercury concentrations in fish, although the precision of these predictions is limited by uncertainty about quantitative aspects of many of the fate and transport processes, especially the conversion of mercury to methyl mercury in the environment. Changes in human exposure can be estimated by combining estimates of the change in methyl mercury concentrations in fish with estimates of human consumption patterns.

The relationship between release of a substance to the environment and eventual human exposure can be summarized as an "intake fraction." An intake fraction is a dimensionless quantity defined as "the integrated incremental intake of a pollutant released from a source … summed over all exposed individuals during a given exposure time, per unit of emitted pollutant" (20). Intake fractions provide a simple summary of population exposure to a substance from a particular source (or category of sources). The effect of a change in release on human exposure can be estimated by multiplying by an estimate of the corresponding intake fraction. The value of an intake fraction may depend on the time horizon over which exposure occurs, the exposure pathway (e.g., inhalation, ingestion, dermal uptake, or all pathways), the time or location of the release, and other factors. Intake fractions can be defined for subpopulations, routes of exposure, and other relevant characteristics. In cases where humans are exposed to a breakdown or other product of the substance released to the environment, intake fraction can be defined as exposure

to the product (e.g., methyl mercury) divided by release of the precursor (e.g., total mercury released by power plants).

Valuation

It is frequently useful to summarize the health effects of an environmental or occupational risk to characterize it, compare it with others, communicate its magnitude, or evaluate management actions to reduce it. Simply adding the expected numbers of disparate health effects (e.g., fatalities and minor illnesses) would be nonsensical, because it fails to recognize differences in severity of outcomes. Valuation is the process of characterizing health outcomes using a method that allows for aggregation and comparison. Because severity is determined in substantial part by people's judgments about how much they are affected by a health outcome, valuation is closely related to human perceptions and preferences.

If the risks of concern are all fatal risks, they can be relatively easily aggregated using the sum of the changes in the probability of death from each risk as the summary measure. However, this approach may be inadequate if there are substantial differences in the age at which the risks manifest, since fatal risks to children or young adults may be of greater concern than risks to middle-aged and elderly adults because they imply a larger loss in life expectancy. Differences in age may be incorporated by valuing fatality risks according to the number of years of life lost.

Aggregating changes in fatality risk obviously fails to include any information on changes in non-fatal illness or disease or in morbidity associated with a fatal disease. Individuals' concerns about health risks may also depend on factors such as perceived stigma or the extent to which the risk is particularly dreaded (21, 22).

There are two approaches to valuing health risks that are often used to value both fatal and non-fatal risks—monetary valuation and health-adjusted life years (23, 24). Monetary valuation measures the importance to an individual of a risk of fatality or of a specific disease by the amount of money that person views as equivalent to a specified change in the probability of suffering the adverse effect. For fatality risks, this value is described as the "value per statistical life," which is defined as the maximum amount he would pay per unit of probability reduction to reduce his chance of dying by a small amount in a specified time period, or the smallest amount of compensation he would accept per unit increase in the probability of dying in a specified period (in theory, these rates should be nearly equal for small changes in risk). The term "value per statistical life" can be motivated by noting that if each of a large number N people would pay

up to $V for an intervention that would reduce each person's chance of dying this year by 1/N, the population would value the intervention at $VN and the intervention would on average prevent one death this year.

The value per statistical life is not an estimate of how much an individual would pay to avoid certain death (which might equal his total wealth) nor an estimate of how much compensation he would demand to accept certain death (which might be infinite); it is only a measure of the rate at which he is willing to trade small changes in risk for money. Value per statistical life is estimated either by observing the choices people make that involve tradeoffs between money and safety (e.g., job choices where workers facing higher occupational risk are paid higher wages, purchase of residential smoke detectors and bicycle helmets) or by asking survey respondents what choices they would make in such situations.

The value per statistical life is not a constant but may vary with wealth and income, age, anticipated health status, total mortality risk, and other factors. It can also depend on qualitative aspects of the risk, such as the extent to which it is particularly dreaded. Analogously, the value per statistical case can be defined for diseases or injuries of interest. The monetary value of a risk change can be estimated by multiplying the change in probability of each endpoint by the corresponding average value per case (so long as the change in probability is small).

Health-adjusted life years, including "quality-adjusted life years" (QALYs) and "disability-adjusted life years" (DALYs), measure the value of a change in health risk by the expected value of the change in longevity where each time period is weighted by an index reflecting the quality of health during that period. The index is scaled so that perfect health is assigned a value of one and a health state no better or worse than dead is assigned a value of zero (DALYs measure losses from an idealized situation and are scaled in the opposite direction). Under this approach, the relative value of reducing fatality risks to younger people rather than older people is proportional to the difference in life expectancy, adjusted for differences in anticipated future health (in general, health declines with age as people develop chronic illnesses). Risks of non-fatal endpoints can be evaluated by estimating the change in the index of health quality associated with the health endpoint and weighting by the duration of the health effect. The index of health quality is estimated by asking survey respondents about hypothetical choices involving tradeoffs between health conditions and life expectancy. Health-adjusted life years provide a metric for combining the effects of changes in both fatal and non-fatal risks that

accounts for effects on health and longevity but typically does not account for other factors such as the source of the risk.

Economic Evaluation

Decisions about whether to take action to mitigate a risk and the choice of risk-mitigation measure require comparing the value of the risk reduction with the value of the benefits that could be achieved by using the resources for other purposes. While this comparison may be explicit or implicit, it seems logical that better decisions will be made if the beneficial and adverse effects of an action together with their probabilities are estimated and compared explicitly. Economic evaluation provides a framework for this comparison (4, 25-29).

Comparing the beneficial and adverse consequences of an action requires a method for valuing the impact of the target risk in a manner that can be compared with the value of the resources or potential benefits that would be sacrificed by actions to mitigate the risk. The comparison of benefits and costs should be comprehensive, including all the quantitatively significant effects of the risk-control measure. In particular, the comparison should include not only the risk targeted by the action, but also any significant reductions (ancillary benefits) and increases (countervailing risks) in other risks (30). These other risks also need to be assessed, using the methods described above. For example, a medication designed to treat one illness may be expensive and may cause both beneficial and adverse side effects.

The two most common forms of economic evaluation are benefit-cost analysis and cost-effectiveness analysis. Benefit-cost analysis (BCA) is conducted by measuring the consequences in monetary units (e.g., dollars) and estimating the expected value of the difference between the beneficial consequences (the benefits) and the adverse effects (the costs). Cost-effectiveness analysis (CEA) is distinguished from BCA in that some of the consequences are measured in non-monetary “effectiveness” units and the ratio of cost per unit effectiveness is estimated. In conducting a CEA of environmental and occupational health risks, the health effects could be measured in “lives saved” (i.e., the expected number of fatalities from a specific risk that are prevented), “life years saved” (the total gain in life expectancy associated with the risk reduction), cases of a specific disease avoided, or the expected increase in health-adjusted life years in the population. BCA allows one to rank alternative policies by the expected value of the net benefits (i.e., the expected difference between benefits and costs) to determine which

produces the greatest expected gain to society. CEA allows one to compare the efficiency with which different policies produce units of effectiveness, but the question of whether the beneficial effects justify the costs incurred requires an independent judgment about what value of the ratio of cost per unit effectiveness is acceptable, and what level is too high.

Principles of Economic Evaluation

A critical difference between individual choices and social choices (such as government rules that affect environmental quality and workplace safety) is that any social choice will typically benefit some people and harm others, at least compared with other alternatives that could be chosen. Hence a critical question in evaluating social choices is how to aggregate consequences across people, in particular how to determine whether harms imposed on some are outweighed by benefits conferred on others.

Economic welfare analysis begins with the assumption that it is not possible to accurately and reliably compare gains or losses in welfare to different people, i.e., if two people develop the same illness, the degree to which each suffers may depend on how much the illness restricts their respective abilities to engage in work or recreational activities, how much each enjoys those activities, how much each suffers from physical pain, and other factors. Given the assumption that it is not possible to make interpersonal comparisons of welfare, economics distinguishes two aspects of a change in social outcomes, efficiency and the distribution of well-being among people. A situation is defined to be Pareto or allocatively efficient if it is not possible to change matters in a way that benefits at least one person without harming anyone. Many situations may be efficient, but differ in how well-off each person is. For example, under the assumption that everyone at a dinner party prefers a larger to a smaller slice of pie, any division that leaves no crumbs is efficient, regardless of the sizes of the pieces. Evidently, whether a situation is efficient has no implication for the distribution of well-being in a population.

Any change in regulation of environmental or occupational risks that is a Pareto improvement (i.e., improves someone's situation without worsening anyone's) is efficiency-enhancing and there is a presumption that it should be adopted, although if it were to substantially increase inequality (e.g., by improving the well-being of only those who were already best off), some people would oppose the change (31). However, since many potential changes are neither Pareto superior nor Pareto inferior to the *status quo*, economists have developed the concept of a "potential Pareto improvement." A potential Pareto improvement is

one that satisfies the Kaldor-Hicks compensation test: those who benefit from the change could provide monetary compensation to those who are harmed such that everyone prefers the change with payment of compensation to the *status quo* (and, moreover, those harmed by the change cannot compensate those who would benefit to forgo the change such that everyone prefers the *status quo* with compensation to the change). The Kaldor-Hicks compensation test circumvents the problem of inability to compare welfare between people by using money as a measuring rod: a change for which the monetary value of the benefits to those who gain (defined by the maximum compensation they would pay to achieve these gains) exceeds the monetary value of the harm to those who lose (defined by the minimum compensation they would require to accept these harms).

But whether a change that satisfies the Kaldor-Hicks compensation test qualifies as a social improvement can be questioned. The test shows that it is possible for those who benefit from a change to adequately compensate those who are harmed, but in fact compensation need not be (and usually is not) paid, leaving some people worse off. The arguments in favor of adopting policies that satisfy the compensation test include the following: (1) Adopting policies that satisfy the test expands the "social pie." While it is not true that expanding the pie ensures that everyone gets a bigger slice, it is true that failing to expand the pie guarantees that some will have a smaller piece than otherwise. (2) If some redistribution of social resources is desired (e.g., to reduce inequality), it may be achieved at lower social cost through more direct transfer mechanisms, such as taxation and welfare policies, than by altering environmental and occupational safety regulations. (3) If society routinely adopts policies that satisfy the compensation test, the groups that benefit and are harmed will tend to differ from case to case, so that everyone may be better off from a set of policies each of which satisfy the criterion than if decisions were made on some other basis. This third argument is limited by the fact that monetary values of benefits and costs tend to be positively correlated with wealth and income and so BCA will tend to favor the interests of the wealthy over the poor. Yet as applied in practice, differences in values associated with income are usually ignored.

Benefit-cost analysis provides a method of determining whether a proposed change satisfies the Kaldor-Hicks compensation test. At the individual level, benefits are measured as the maximum amount a person would be willing to pay to obtain the changes he views as beneficial, and costs as the minimum amount of compensation he would require to accept the

changes he views as adverse. Summing the monetary values of benefits and costs across the population determines whether the net benefits (the total benefits less the total costs) are positive. If they are, the change satisfies the compensation test because those who gain would be able to compensate those who are harmed and still prefer the change. If net benefits are negative, there is no set of compensation payments that results in everyone preferring the change with compensation to the *status quo*.

Although it is conventional to define benefits as the reduction in harms due to reducing environmental or occupational risks and costs as the value of the resources used to achieve these gains (e.g., emission-control or safety equipment, use of more expensive but less hazardous materials), there is no fundamental difference between benefits and costs. A reduction in environmental health risk may be viewed as a benefit or as a reduction in cost (i.e., a reduction in the health risk associated with some process). Similarly, an increase in the cost of producing a product that results from substituting an environmentally superior input or safer working conditions may be viewed as a cost or a negative benefit. When costs are subtracted from benefits to calculate net benefits, the result is determined only by the sign of the terms, not by whether they are classified as benefits or costs. In contrast, the ratio of benefits to costs (or costs to benefits) is sensitive to how items are classified, and so ratios can be misleading and must be evaluated carefully.

Cost-effectiveness analysis can be contrasted with BCA in the way that effects are aggregated across the affected population. CEA differs from BCA in using a non-monetary measure of the changes in health risk, such as the expected increase in health-adjusted life years experienced by the population. Adding health-adjusted life years in a population circumvents the problem of non-comparability of changes in well-being among people by defining an improvement of one health-adjusted life year as equally important, regardless of how the health adjusted life year is distributed among the population (e.g., either as a large gain to a few people or a tiny gain to many people). In contrast, BCA treats a dollar of benefit as equally desirable, regardless of how it is distributed (24). Like BCA, CEA measures the costs of policies to reduce health risk using the monetary value of the resources consumed, thus aggregating costs on the principle that a dollar of cost is equally harmful, regardless of how it is distributed. Because it requires the use of a ratio of costs to effectiveness rather than a difference, the rankings of projects using CEA can depend on whether specific effects are defined as costs or effects. For

example, the reduction in productivity (either in the workplace or in other activities) when an individual is ill could in principle be counted as part of the effect and included in the loss of health quality or could be counted as part of the cost of illness. Use of CEA to compare actions requires a common definition of which elements are counted as effects and as costs.

Definition and Measurement of Benefits and Costs

Both BCA and CEA attempt to measure the "social" or "resource" benefits and costs of a policy or activity. Social benefits and costs represent the sum over a population of the private benefits and costs to individuals and firms. Some private benefits and costs increase or decrease social welfare directly, but others are simply transfers of resources within a society and do not affect the total. For example, a requirement to install catalytic converters to reduce pollution from automobile exhausts is a private cost to people who buy new automobiles, since they must pay for the additional equipment. It is also a social cost, because workers' time and materials used to produce the catalytic converters cannot be used to produce something else of value. In contrast, a tax on gasoline purchases is not a social cost but a transfer of resources from the gasoline consumer to the government which can use these resources to purchase other goods. (Imposing a gasoline tax may create a social cost by preventing consumers who would be willing to pay the cost of the gasoline, but not the cost plus the tax, from using gasoline.)

Benefits and costs are typically estimated using different methods. Costs are estimated from market prices or engineering models. Benefits are estimated from revealed-preference or stated-preference methods. Usually, there is substantial uncertainty about the magnitudes of both benefits and costs, with benefits often being the more uncertain.

The social cost of an action to reduce environmental or occupational risk is the value of the resources used to reduce the risk. It may include the cost of "end-of-pipe" control technology that reduces emissions (such as catalytic converters), changes in process (such as substituting fuel injection for carburetion), or changes in inputs (such as substituting natural gas for petroleum fuel). When multiple technologies are employed, one may be able to estimate the incremental cost by comparing the prices of the technologies, possibly controlling for other factors using statistical regression methods. When only one of the technologies is employed, costs may be estimated using engineering models to project the quantities of resources required to produce the new technology and the market prices or other estimates of the social value these

resources. The costs of new technologies that have not yet been produced at commercial scale are necessarily rather speculative.

Estimating Benefits

Benefits are estimated by "revealed-preference" or "stated-preference" methods (23, 24, 32). Revealed-preference methods are based on the assumption that an individual chooses from a set of available alternatives the one he prefers most (or one of the most preferred, if he is indifferent among two or more alternatives). This method requires that the analyst be able to obtain data on choices people make between alternatives that differ in the attribute to be valued (e.g., environmental quality or health risk) and monetary cost.

The "compensating-wage-differential" method is a revealed-preference method that is commonly used to estimate the value of changes in mortality risk, using data on workplace fatality risk and wages. Using regression methods to control for differences in education, type of work, and other factors that influence wages, analysts find that wages are positively associated with occupational fatality risks (33). A typical estimate would be that a worker receives an additional $500 in annual wages as compensation for an incremental annual fatality risk of 1/10,000. The worker is assumed to like the job he holds at least as much as others for which he is qualified, such as one paying $490 more and having a 1/10,000 higher annual fatality risk, and another paying $510 less and having a 1/10,000 lower annual fatality risk. This result is interpreted as implying that the value to the worker of a change is about $500 per 1/10,000 risk, commonly expressed as $5 million per statistical life.

Revealed-preference methods can only be used in cases where the benefit to be valued is currently available and one can observe people who choose to obtain the benefit and others who choose not to. Although the value of a reduction in the probability of fatal workplace accidents may differ from the value of a comparable reduction in the chance of fatality from exposure to environmental pollutants, revealed-preference estimates of the value of reducing environmental fatality risks are much more difficult to construct and so the value estimated from workplace accidents is often used as an approximation.

The other major approach, stated-preference methods, relies on asking survey respondents what choices they would make in a hypothetical setting and interpreting their answers under the assumption that the choices they report are the ones they most prefer. Stated-preference methods are not limited to situations where people's actual choices can be observed.

As a result, they are often viewed as less persuasive than revealed-preference estimates since survey respondents may be unfamiliar and inexperienced with the hypothetical choices they are offered, may have limited incentive to carefully consider the choices they would make, and may respond to other aspects of the survey setting unrelated to their preferences for the hypothetical choices such as a desire to provide an answer that will please the interviewer.

The most widely used form of stated-preference method is contingent valuation. With this method, survey respondents are presented with a clearly defined choice between two alternatives differing in the attribute to be valued (such as air quality), the method by which the change in the attribute is to be produced (such as a change in environmental regulations), the monetary cost to the individual, and the form in which the cost would be paid (such as higher taxes or higher prices of other goods). Although the choice is hypothetical, it is recommended that it be described as realistically as possible to encourage respondents to take the choice seriously (34, 35).

Accuracy of Estimates

It is sometimes claimed that costs of regulations to improve environmental quality or workplace safety are overestimated prospectively and that when rules are implemented the costs of compliance prove to be substantially smaller than anticipated. Moreover, it is also claimed that benefits are often underestimated, because it is difficult to estimate the value of reductions in health risk and of other benefits, such as improvements in environmental quality.

Such claims are plausible, for several reasons. With respect to costs of complying with a regulation, it would be logical for firms and other regulated entities to invest more effort in finding lower-cost ways to comply with a rule after it is promulgated than beforehand, when they are uncertain whether the rule will be adopted. In principle, cost analysts should account for the likelihood that lower-cost compliance methods will be found, but it is difficult to credibly estimate the probability and magnitude of cost-saving discoveries and this possibility is usually not incorporated in cost estimates. Moreover, in accounting for cost-saving innovations, it would be appropriate to account for the value of the other innovations the firm might have produced instead, had its efforts not been directed to complying with the new rule. To a first approximation, it may be reasonable to suppose these other innovations would be equally valuable as the compliance-cost savings.

On the benefit side, benefits of health risk reductions are often estimated using cost-of-illness methods rather than willingness to pay. Cost-of-illness methods estimate the monetary value of the loss from sickness or death as the sum of the costs of treating the illness (e.g., physician and other care-giver time, medications) and the lost productivity (the value of the goods and services an individual would have produced had he not been ill or died). These methods underestimate the economic value of health risk because they fail to include the loss in well-being to the individual exclusive of any market consequences. However, to the extent that the use of conservative assumptions in risk assessment yield overestimates of the initial risk and of the reduction due to reducing human exposures, benefits might be overestimated.

In evaluating the accuracy of prospective estimates of benefits and costs, it is important to note that in many cases published estimates may not represent expected values. While it is clear that policy advocates and opponents may have an interest in publicizing estimates of benefits or costs that favor their desired outcome, it is perhaps less well appreciated that the estimates reported by government agencies may also be biased. For example, an agency wishing to promulgate a rule may be required to demonstrate that the benefits are likely to exceed the costs. In this case, the agency is well-served by an analysis showing that even conservative or lower-bound estimates of benefits exceed conservative or upper-bound estimates of costs. Alternatively, an agency may be required to show that its proposed regulations are economically feasible. In this case, an analysis showing that even upper-bound cost estimates are not unduly burdensome may be appropriate. Moreover, agencies recognize that their decisions may be challenged in court and so may prefer to offer defensible bounds on benefits and costs than to engage in the more speculative task of attempting to estimate expected values.

A number of studies have attempted to compare estimates of regulatory costs and benefits produced before a regulation was enacted with retrospective estimates. Note that the retrospective values are also estimates because, although one can observe some of the consequences once the rule is implemented, one cannot observe what the consequences would be if the rule had not been adopted and so the counterfactual situation must be estimated. In addition, the health benefits of a regulation may remain quite uncertain in cases where the individuals suffering the health effects due to the agent that is regulated may not be identifiable, e.g., for deaths associated with fine particulate air pollution and for cancers associated with chemicals for which only laboratory animal data are available.

Harrington et al (36) reviewed the available studies which compared prospective and retrospective estimates of costs and benefits for 21 regulations, restricting attention to the prospective estimates produced by government agencies. They found that the question of accuracy depends critically on definition, and that the results differ between regulations that specify the control technology or emission level to be achieved (so called "command-and-control" regulations) and regulations that establish a system of tradable emission permits, taxes, or other economic incentives. For the command-and-control regulations, the total cost was overestimated more often than not. However, the total effectiveness of the regulation, measured by the reduction in the quantity of the agent released, was also overestimated in most cases, with the net result that the cost per unit of effectiveness was overestimated and underestimated with equal frequency. In contrast, for the economic-incentive rules the quantity of emission reduction was typically underestimated and the cost of compliance was typically overestimated, with the net result that the cost per unit effectiveness was usually overestimated prospectively. This difference in the bias of prospective estimates can be seen as evidence that economic-incentive regulations provide greater incentives than command-and-control rules for firms to develop and adopt lower-cost compliance methods.

Discounting

When the benefits or costs of an action are incurred at different points in time, it is conventional to adjust for the difference in timing using discounting. Typically, the time streams of expected benefits and costs are converted to "present value" by multiplying each value in the year it will be incurred by $[1/(1 + r)]^t$, where t is the number of years in the future when the benefit or cost will be incurred and r is the discount rate, which is typically between about two and seven percent per year.

The concept behind discounting is that resources can be put to productive use and so it is more valuable to receive a resource sooner than later. For example, if financial capital can be invested in projects that yield a three percent annual return, then a benefit valued at $10 million when it is received in 20 years is as valuable as $5.5 million received today. This follows because $5.5 million invested at three percent per year would yield $10 million in 20 years.

Discounting can appear to give insufficient weight to future consequences. Indeed, the effect of discounting can be surprisingly large over long time periods and for large discount rates. For example, the present value of $10 million to be received in 100 years is only $11,000

using a seven percent rate. The logic of discounting is clear, but the questions about its validity concern the inevitable uncertainty about what rate of return can be realistically expected in the future. Moreover, discounting over long time periods inevitably confounds issues of efficiency and distribution, since the people who will receive a benefit or bear a cost 100 years in the future are different from those living now (37, 38).

Economic evaluation provides a structured method to account for all the significant effects of a proposed action, including beneficial and adverse consequences. It attempts to answer the question whether the people who face the consequences of an action, including both the beneficial and adverse effects, judge themselves to be better off with the action or without. Compared with alternative approaches, economic evaluation requires explicit documentation of the effects that are considered together with the importance of changes in each effect (i.e., the monetary values or units of effectiveness). Although economic evaluation can appear to be a technical exercise, it is fundamentally populist in that it counts the preferences of the people affected by a regulation, not the preferences of the decision maker or of politically powerful lobbies. Moreover, by documenting the factors considered and the importance or values assigned to each, the rationale for a decision can be made explicit and evidence that the numerical values used are incorrect can be presented and used to refine the analysis and determine whether that affects its conclusion. A key limitation is that economic evaluation addresses only the question of efficiency, defined by whether those who benefit from an action could compensate those who are harmed. Judgments about whether the distributional effects of an action are more important than the efficiency aspects, and whether there are alternative methods to correct for any adverse distributional effects, must typically be made outside the formal analysis. However, the economic analysis can also provide information about the loss in benefits or increase in cost that is incurred by choosing an action that is inefficient but potentially preferred because of its distributional implications.

Conclusion

Humans are in many respects healthier now than at any time in history; life expectancies have increased substantially over the last century in most populations. A fundamental challenge to risk analysis of environmental and occupational risks is that society is concerned about many risks that are too small to be accurately measured with current methods. Moreover, for the most

part exposures to environmental and occupational hazards do not produce a signature disease or cause of death but instead increase the probability of a disease or cause of death that can result from other factors (there are exceptions, such as mesothelioma which is caused by asbestos). Distinguishing the deaths or cases of disease that result from a specific environmental or occupational exposure from those that result from other causes is generally impossible. For example, even though exposure to airborne fine particulate matter is estimated to cause perhaps 60,000 deaths per year in the United States, primarily through heart attacks and other cardiovascular events, the individuals dying because of air pollution cannot be identified among the total of 700,000 deaths per year from heart disease.

Another factor that limits the ability to accurately estimate the risks associated with specific exposures is the large number of potentially hazardous agents in the environment (both naturally occurring and synthetic) to which every human is exposed. The possibility of interactions among agents—both synergistic and antagonistic—cannot be ruled out *a priori*. Indeed, cigarette smoke interacts synergistically with both radon and asbestos exposures, so the increases in cancer risk due to exposure to radon or asbestos are larger for smokers than for nonsmokers. In contrast, exposure to selenium reduces the risk associated with exposures to both cadmium and arsenic through competition for molecular binding sites. Cases where an agent that is not itself hazardous at the relevant exposure level increases or decreases the risk associated with a second agent are described as potentiation and inhibition, respectively (39). Moreover, exposure to some agents (such as vaccines and viruses that cause childhood illnesses, which may themselves prove debilitating or fatal) may induce the body to build defenses or immunities against future exposure, so that exposure to these agents can reduce the risk associated with future exposure. The existence of interactions between agents raises the specter of having to assess risks not only of individual agents, but of the much larger set of mixtures and other combinations of agents. For new risks, such as those associated with an emerging disease or a newly developed technology, there are by definition few or no data concerning the risks that may result as these agents diffuse through the environment and so risk estimates must be based almost entirely on theory and laboratory data.

One response to uncertainty in risk assessment has been a tendency to adopt methods believed to produce “conservative” estimates, i.e., estimates that are more likely to overestimate than to underestimate a risk. As described above, in extrapolating from animal experiments to

human risk it is conventional to use data from the species and sex showing the largest risks, to use a linear no-threshold model to extrapolate from high to low dose, and to choose the health endpoint yielding the smallest estimate of the benchmark dose (BMD), no observed adverse effect level (NOAEL), or lowest observed adverse effect level (LOAEL). Estimates of human exposure are often constructed in ways to produce something closer to an upper bound than a population-average exposure, using concepts such as the maximally exposed individual (MEI).

The emphasis on conservative estimates is valuable for helping to identify and mitigate even small risks, but less helpful for determining how to balance the benefits of reducing exposure against the potential countervailing risks and other costs. The worst case is often not well defined, but may be limited only by one's imagination. Moreover, emphasis on conservative or worst-case risk estimates can lead to the perverse result of focusing efforts on the most uncertain risks (for which the worst case is far worse than the expected value) rather than on probably larger but more certain risks, a phenomenon labeled the "perils of prudence" (40). A better approach is to characterize the uncertainty about the risk in the form of a probability distribution which shows the plausible range of values as well as the relative probability that each is correct (41).

Estimates of the health risks associated with environmental and occupational exposures, and of the values of benefits and costs used in economic valuation, are subject to substantial uncertainty. A degree of humility is required in presenting results, and a degree of tolerance for uncertainty is required in using risk estimates. Nevertheless, the probabilities of harm from exposure to environmental, occupational, and other factors vary over many orders of magnitude and the benefits of reducing exposure to these agents also vary widely. Choices about what actions to take to reduce a risk inevitably involve judgments about what level of sacrifice is justified, a judgment which is critically dependent on the magnitude of risk reduction. It seems unassailable that more sensible choices can be made when decision makers have access to judicious, quantitative descriptions of the risk, how it may be influenced by risk-control measures, the costs of the control measures and their effects on other risks, and the attendant uncertainty about these matters, than when they must choose without the best understanding of the probabilities and severity of the potential consequences that can be provided by current scientific understanding.

Acknowledgement

This chapter was written while I was appointed to a Pierre de Fermat Chaire d'Excellence at the Université de Toulouse. I thank colleagues at the Institut d'Economie Industrielle and Laboratoire d'Economie des Ressources Naturelle for hospitality and support, and John Evans, George Gray, Jonathan Levy, Glenn Rice, Jonathan Wiener, Andrew Wilson, and Richard Wilson for helpful discussions and comments.

References

1. Raiffa H. Decision Analysis: Introductory Lectures on Choices Under Uncertainty. Addison-Wesley, Reading, MA, 1968.
2. Clemen RT. Making Hard Decisions: An Introduction to Decision Analysis. PWS-Kent Publishing Co., Boston, 1991.
3. Hammond JS, Keeney RL, and Raiffa H. *Smart Choices: A Practical Guide to Making Better Decisions*. Harvard Business School Press, Boston, 1999.
4. Stokey E and Zeckhauser RJ. *A Primer for Policy Analysis*. W.W. Norton, New York, 1978.
5. Wilson R and Edmund AC. Crouch, *Risk-Benefit Analysis*. Harvard University Press, Cambridge, 2001.
6. Covello VR and Merkhofer MW. Risk Assessment Methods: Approaches for Assessing Health and Environmental Risks. Plenum Press, New York, 1993.
7. Rodricks JV. Calculated Risks: The Toxicity and Human Health Risks of Chemicals in our Environment. Cambridge University Press, Cambridge, 1992.
8. U.S. Environmental Protection Agency. *Risk Assessment Guidelines*, http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=55907, no date.
9. National Research Council, *Risk Assessment in the Federal Government: Managing the Process*. National Academies Press, Washington, DC, 1983.
10. Crump K, Hoel D, Langley H, and Peto R. Fundamental Carcinogenic Processes and their Implications to Low Dose Risk Assessment. *Cancer Research* 1976; 36: 2973-2979.
11. Crawford M and Wilson R. Low Dose Linearity: The Rule or the Exception? *Human and Ecological Risk Assessment* 1996; 2: 305-330.
12. U.S. Environmental Protection Agency. *The Use of the Benchmark Dose Approach in Health Risk Assessment*. EPA/630/R-94/007, Washington, DC, 1995.
13. U.S. Environmental Protection Agency. *Exposure Factors Handbook.* EPA/EPA/600/P-95/Fa, Fb, Fc and http://cfpub.epa.gov/ncea, Washington, DC, 1997.
14. Wilson R and Spengler JD (eds.). *Particles in Our Air: Concentrations and Health Effects*. Harvard University Press, Cambridge, 1996.

15. U.S. Department of Health and Human Services. *National Toxicology Program*. http://ntp-server.niehs.nih.gov, no date.

16. Ames BN, and Gold LS. Too Many Rodent Carcinogens: Mitogenesis Increases Mutagenesis. *Science* 1990; 249: 970-971.

17. U.S. Environmental Protection Agency, *Guidelines for Carcinogen Risk Assessment*. EPA/630/P-03/001F, Washington, DC, 2005.

18. Paustenbach DJ. The Practice of Exposure Assessment: A State-of-the-Art Review. *Journal of Toxicology and Environmental Health* 2000; 3: 179-291.

19. U.S. Environmental Protection Agency, *Guidelines for Exposure Assessment*. EPA/600/Z-92/001, Washington, DC, 1992.

20. Bennett DH, McKone TE, Evans JS, Nazaroff WW, Margni MD, Jolliet O, Smith KR. Defining Intake Fraction. *Environmental Science and Technology* 2002; 36: 206 A – 211 A.

21. Kahneman D, Slovic P, and Tversky A. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge, 1982.

22. Slovic P. *The Perception of Risk*. Earthscan Publications, London, 2000.

23. Hammitt JK. Valuing Mortality Risk: Theory and Practice. *Environmental Science and Technology* 2000; 34: 1396-1400.

24. Hammitt JK. QALYs versus WTP. *Risk Analysis* 2002; 22: 985-1001.

25. Layard R and Glaister S. *Cost-Benefit Analysis.* Cambridge University Press, Cambridge, 1994.

26. Hanley N and Spash CL. *Cost-Benefit Analysis and the Environment*. Edward Elgar, Cheltenham, 1993.

27. U.S. Environmental Protection Agency. *Guidelines for Preparing Economic Analyses*. EPA-240-R-00-003, Washington, DC, 2000.

28. Gold MR, Siegel JE, Russell LB, and Weinstein MC (eds.). *Cost-Effectiveness in Health and Medicine*. Oxford University Press, New York, 1996.

29. Neumann PJ. Using Cost-Effectiveness Analysis to Improve Health Care: Barriers and Opportunities. Oxford University Press, Oxford, 2005.

30. Graham JD and Wiener JB (eds.). *Risk Versus Risk: Tradeoffs in Protecting Health and the Environment*. Harvard University Press, Cambridge, 1995.

31. Okun AM. *Equality and Efficiency: The Big Tradeoff*. The Brookings Institution, Washington, DC, 1975.

32. Freeman AM III. *The Measurement of Environmental and Resource Values: Theory and Methods* (2d ed.). Resources for the Future, Washington, DC, 2003.

33. Viscusi WK and Aldy JE. The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World. *Journal of Risk and Uncertainty* 2003: 27: 5-76.

34. Mitchell RC and Carson RT. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Resources for the Future, Washington, DC, 1989.

35. Carson RT. Contingent Valuation: A User's Guide. *Environmental Science and Technology* 2000; 34: 1413-1418.

36. Harrington W, Morgenstern RD and Nelson P. On the Accuracy of Regulatory Cost Estimates. *Journal of Policy Analysis and Management* 2000; 19: 297-322.

37. Portney PR and Weyant JP (eds.). *Discounting and Intergenerational Equity*. Resources for the Future, Washington, DC, 1999.

38. Groom B, Hepburn C, Koundouri P, and Pearce D. Declining Discount Rates: The Long and the Short of It. *Environmental and Resource Economics* 2005; 32: 445-493.

39. U.S. Environmental Protection Agency, *Supplementary Guidance for Conducting Health Risk Assessment of Chemical Mixtures*. EPA /630/R-00/002, Washington, DC, 2000.

40. Nichols AL and Zeckhauser RJ. The Perils of Prudence: How Conservative Risk Assessments Distort Regulation. *Regulatory and Toxicological Pharmacology* 1988; 8: 61-75.

41. Morgan MG and Henrion M. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, Cambridge, 1990.