Baruch Fischhoff
Professor, Howard Heinz University
Carnegie Mellon University
Member, Institute of Medicine,
National Academy of Sciences (US)
Chair,
FDA Risk Communication
Advisory Committee

Granger Morgan
Professor, Lord University
Carnegie Mellon University
Member, National Academy
of Sciences (US)

The Science and Practice of Risk Ranking

Summary

Sound risk ranking is essential to effective risk management. Without it, small risks may receive unwarranted attention, while large ones are neglected. The challenges in ranking risks include the sheer number that need to be considered, the variety of ways to define "risk," and the differences among stakeholders, regarding which consequences matter most. Addressing these challenges requires an understanding of risks, risk analysis, and decision-making processes. A practical approach is offered for producing sound, transparent, and credible risk rankings.

Ranking risks to health, safety, and the environment is important because, while there are risks everywhere, we have limited resources for managing them. In an ideal world, we would regularly review our priorities, deciding which risks deserve more attention and which less. In the real world, systematic reviews of risk priorities are as rare in the public arena as they are in our private lives. That is, we usually muddle through, waiting until circumstances bring a risk to our attention, then decide whether to treat it more or less seriously.

In our private lives, we bear the consequences, if we spend our time, money, attention, or emotional resources poorly. However, the public as a whole suffers, when policy makers worry about the wrong things. When setting their priorities, policy makers face the same challenges as do individuals, one challenge being the sheer number of risks that might be considered. A second is deciding how to define "risk." A third is reconciling the variety of values of the different stakeholders in comparing risks.

The first section below considers these challenges from a theoretical perspective. It is followed by a short history of US Environmental Protection Agency (EPA) efforts to grapple with them. The next section describes an approach that combines risk research with practical experience in risk ranking followed by consideration of the compatibility of this approach with risk-management processes initiated by the Government of the United Kingdom and advocated by the Canadian Standards Association.

Before beginning, it is important to note that ranking risks is but one critical step in effective risk management. Ordering risks by their importance allows policy makers to focus on those that matter most. It does not, however, say what to do about them. It does not even determine which risks require action or who should take it. There may be small risks that are easily managed and large risks that merit no further attention, because there is nothing to be done about them, beyond investing in research that might, one day, make action possible.

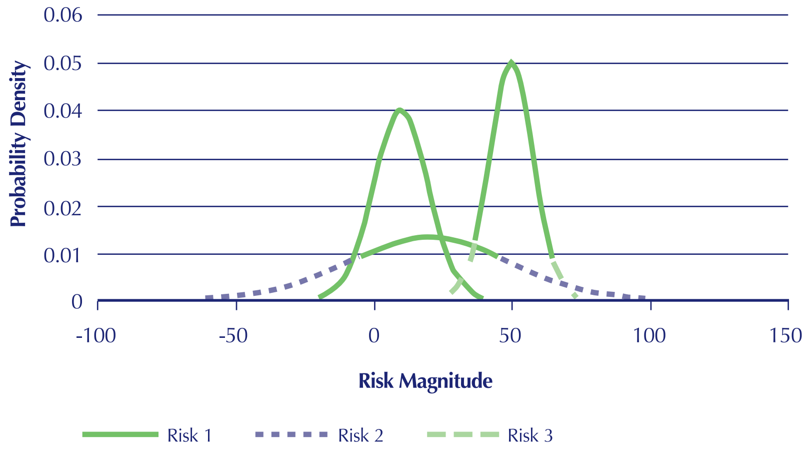
Challenges to Risk Ranking

Risk analysis is an interdisciplinary field that develops and applies computational and empirical methods to understanding risks. It has identified three challenges to ranking risks: too many risks, too many definitions of risk, and too many values.

Too Many Risks. The list of risks facing an agency, firm, or family can be long and varied. For example, on a given day, a parent might need to decide how much attention to pay to a child's cough, a car's rattle, an aging parent's recent fall, a wave of neighbourhood burglaries, a worrisome skin rash, and blood sugar irregularities. At a given meeting, a school board might need to decide how much attention to pay to missing school bus seatbelts, playground fights, potential pandemics, broken stairs, and student obesity.

Normally, people pursue *sequential risk ranking*. That is, they wait until a risk draws their attention, then work hard to understand it better. Based on that improved understanding, they move that risk up or down in their ranking, hoping to afford it a more appropriate level of concern. Thus, a parent might conclude that the rattle is just annoying, then try to put it out of mind. A school board might conclude that it is living on borrowed time, for pandemic preparedness, then try to push other risks away, to give a possible pandemic the needed attention.

Figure 1
Some Complexities of Risk Ranking, in a Simple Case



Source: Long and Fischhoff (2000).

Over time, sequential risk ranking might lead to better priorities – or it might lead to focusing on vivid minor risks, while neglecting quite serious ones. Parents can neglect their own major health problems, while attending to minor concerns about their kids, cars, and home. School boards can neglect potential disasters, while dealing with routine problems and single-issue interest groups.

The success of sequential ranking depends on how many risks need to be ranked, how quickly uncertainty about them can be reduced, how they attract attention, and how precise the ranking needs to be (Long and Fischhoff, 2000). Sequential ranking can work well, for example, when public health surveillance programs pick up telltale signs of emerging diseases, whose seriousness can be quickly ascertained. It can work less well when it is driven by the 24/7 news cycle.

When sequential ranking proves impossibly inefficient, *simultaneous risk ranking* is needed: looking at all risks at once. As appealing as that idea might be, in principle, the challenges to its execution are substantial. Figure 1 shows, in abstract terms, issues that arise when ranking three risks measured on a single scale. As the number of risks increases, the complexity of simultaneous ranking can grow exponentially, diluting the attention paid to part of the work. At the extreme, attempting to understand everything can lead to understanding nothing. The remainder of this article considers practical ways to overcome three key challenges to simultaneous risk ranking.

Too Many Definitions of "Risk." Figure 1 has one major simplification: all risks are measured in a common unit (called the risk magnitude). Risk analysts have long realized that there is no single measure of "risk." Even when

Table 1
Risk Comparisons

One…legitimate purpose [for risk comparisons] is giving recipients an intuitive feeling for just how large a risk is by comparing it with another, otherwise similar, risk that recipients understand. For example, roughly one American in a million dies from lightning in an average year. "As likely as being hit by lightning" would be a relevant and useful comparison for someone who has an accurate intuitive feeling for the probability of being hit by lightning, faces roughly that "average" risk, and considers the comparison risk to be like death by lightning in all important respects. It is not hard to imagine each of these conditions failing, rendering the comparisons irrelevant or harmful:

(a) Lightning deaths are so vivid and newsworthy that they might be overestimated relative to other, equally probable events. But "being struck by lightning" is an iconic very-low-probability risk, meaning that it might be underestimated. Where either occurs, the comparison will mislead.

(b) Individual Americans face different risks from lightning. For example, they are, on the average, much higher for golfers than for nursing-home residents. A blanket statement would mislead readers who did not think about this variability and what their risk is relative to that of the average American.

(c) Death by lightning has distinctive properties. It is sometimes immediate, sometimes preceded by painful suffering. It can leave victims and their survivors unprepared. It offers some possibility of risk reduction, which people may understand to some degree. It poses an acute threat at some very limited times but typically no threat at all. Each of those properties may lead people to judge them differently — and undermine the relevance of comparisons with risks having different properties.

(d) It is often assumed that the risks being used for comparison are widely considered acceptable at their present levels. The risks may be accepted in the trivial sense that people are, in fact, living with them. But that does not make them acceptable in the sense that people believe that they are as low as they should or could be…

The second conceivable use of risk comparisons is to facilitate making consistent decisions regarding different risks. Other things being equal, one would want similar risks from different sources to be treated the same. However, many things might need to be held equal, including the various properties of risks…that might make people want to treat them differently despite similarity in one dimension…

The same risk may be acceptable in one setting but not another if the associated benefits are different (for example, being struck by lightning while golfing or working on a road crew). Even when making voluntary decisions, people do not accept risks in isolation but in the context of the associated benefits. As a result, acceptable risk is a misnomer except as shorthand for a voluntarily assumed risk accompanied by acceptable benefits.

Source: US NRC (2006; Pp. 37-38).

risk rankers care only about expected deaths, they must decide whether to treat all deaths as equal or, if not, how to weight them. For example, risks can be ranked differently, when measured by "expected probability of premature death" or "expected years of lost life" (which assigns extra weight to deaths of young people). Ranks might differ, too, when measured in units that consider benefits (e.g., deaths per coal miner vs. deaths per ton of mined coal) or in units that consider exposure (e.g., deaths per mile travelled) (Fischhoff et al., 1981; Crouch and Wilson, 1981).

Additional choices arise when deciding how to include various kinds of morbidity, in the measure of risk. Lively academic debates revolve around different measures of *quality-adjusted life years* (associated with different forms of harm). These measures try to put diverse risks on a common footing, by asking people how much they, personally, value different states – using structured surveys to resolve the ethical issues of defining risk.

Defining risk is complicated further when mortality and morbidity do not capture all the concerns of citizens. For example, they may also care about how voluntary exposure to a risk is, how equitably it is distributed across the population, how much of a sense of dread it evokes, how controllable it seems, how far in the future its effects extend, how well it is understood by science, how well it is understood by those exposed to it, how immediate its effects are seen, and how new it is (Fischhoff et al., 1978; Slovic, 1987). Ignoring these risk attributes can mean missing issues that are critical to policy makers or their constituents.

Defining risk is complicated further when mortality and morbidity do not capture all of citizens' concerns

Too Many Possible Values. Once the risks have been characterized, ranking can begin, bringing additional complications. Reasonable people can disagree about the relative importance of mortality and various forms of morbidity, or even about the importance of different aspects of mortality. For example, some people are more averse to risks that have *catastrophic potential*, in the sense that they can take many lives at once (e.g., aviation), compared to *chronic risks*, with the same expected death toll exacted at a more even rate (e.g., driving). Other people find it offensive not to treat all deaths equally. Those people might consider catastrophic potential, because of its *signal value*, feeling that risks that can take many lives at once may be poorly understood and managed. Similarly, some people want to have all risks treated similarly, regardless of whether exposure to them is voluntary, whereas others believe that people get more benefit from risks that they assume voluntarily (Slovic, 2000).

A common temptation for simplifying risk ranking is comparing risks that exhibit seemingly similar magnitudes, then arguing that they should be treated similarly. A "classic" comparison equates the risk of living 50 years beside a nuclear power plant to that of eating a tablespoon of peanut butter (due to potential aflatoxin contamination). Table 1 summarizes the logical flaws in such comparisons.

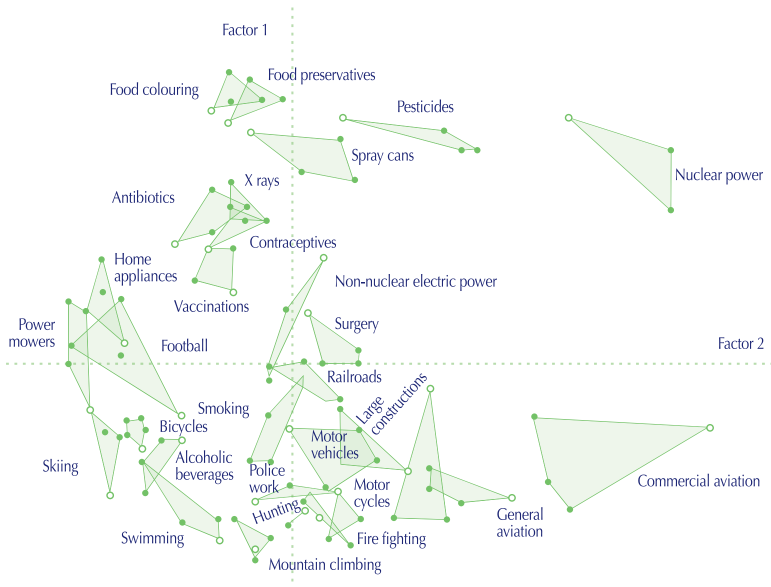
Figure 1 reveals an additional challenge to ranking risks, even when they have been reduced to a common unit. The rankings depend on what statistic is used to represent a risk whose value is not known with certainty (as is almost always the case). If means are used (as a "best guess"), the three risks would be ranked 3-2-1. If a high percentile is used (as a "worst case"), the order becomes 2-3-1. Other statistics are also possible, including different "best guesses" (in cases where the mean, median, and mode differ).

In these ways, defining "risk" raises fundamental value questions, which must be resolved before scientific evidence can be assembled, regarding the magnitude of the risks, and the ranking process begun. In principle, an organization could choose to resolve these value issues among its stakeholders, then let someone else assemble the science and compute the ranks. In practice, resolving value issues in an informed way typically requires vigorous discussion among individuals with suitably diverse perspectives. Without such a deliberative process, the issues are unlikely to be thoroughly understood (US NRC, 1996). Typically, they are too complex for individuals to grasp fully, without hearing other people's views Moreover, transparent, public deliberations, by trusted individuals may be needed for rankings to have external credibility. Two decades of research and practice have produced a foundation for methods to achieve these goals.

Ranking Risks at US EPA

The US Environmental Protection Agency (EPA) has long sought to set its regulatory and research agenda systematically. A landmark report, *Unfinished Business* (US EPA, 1987), summarized the judgments of 75 staff members ranking the risks addressed by the EPA's existing programs, as well as risks that it might, one day, regulate. A similar process, undertaken by the EPA's Science Advisory Board, produced *Reducing Risk: Setting Priorities and Strategies for Environmental Protection* (US EPA, 1990). Based on the framework that these reports created, the EPA established a program to encourage state and local risk-ranking exercises. After supporting several dozen such exercises, the EPA published *A Guidebook to Comparing Risks and Setting Environmental Priorities* (US EPA, 1993), with thoughtful advice on conducting respectful, scientifically informed deliberations. Seeing its foundational

Figure 2
Location of 30 Hazards within a Two-Factor Risk Space



Source: Slovic et al. (1985).

work as done, the EPA funded two regional centers to support additional ranking.

Central to the EPA's approach is letting participants drive the process, in terms of which risks are ranked and how "risk" is defined. Technical experts are entrusted with creating risk estimates relevant to participants' concerns. One price paid for this flexibility and responsiveness is reduced transparency. Individuals who were not in a group must trust the work of those who were, because the rationale for their ranking is not made explicit. A second price is limited comparability. Without a standard definition of "risk," one cannot tell whether different groups have reached consistent conclusions, or pool ranking results across domains, so overall priorities can emerge.

As a result of these methodological problems and changed political conditions, systematic risk-ranking has not been a priority for the EPA recently. The US Department of Homeland Security has committed to risk-informed decision making. However, its work has involved computation, without deliberation.

A Method for Risk Ranking

Drawing on research in risk analysis and behavioural decision research, a group centred at Carnegie Mellon University's Department of Engineering and Public Policy developed a risk-ranking procedure that adds standardization and transparency to the EPA's flexible, participatory approach. Like the EPA approach, Carnegie Mellon's recognizes the variety of risks and ways to value them. It, too, allows participants' concerns to drive the selection and presentation of risk estimates and uses risk analysis to aid judgment, rather than to replace it. It also views well-informed stakeholders as the final arbiters of risk priorities.

The Carnegie Mellon approach departs from the EPA practice of characterizing all risks in terms of a common set of attributes, rather than allowing each ranking exercise to choose its own attributes. Such standardization is possible for two reasons: there are some attributes that most people want to consider and, hence, belong in every exercise (e.g., human mortality), and many potentially relevant attributes are correlated (e.g., involuntarily assumed risks tend to be distributed inequitably). As a result, taking a representative (or two) from a cluster of correlated attributes should address that general set of concerns. Figure 2 shows such core clusters, represented as dimensions in a *risk attribute space*.

Many other studies, with varying activities and technologies, risk attributes, risk raters, and statistical procedures, have yielded similar patterns: (a) People rate risks similarly on these attributes, even when they disagree about the attributes' importance. (b) Attribute ratings are highly correlated, typically revealing two primary dimensions, given names like *Unknown* (vertical) and *Dread* (horizontal).

Based on these regularities, the Carnegie Mellon approach characterizes all risks in terms of the same attributes, as in Table 2. Each column uses two different (but correlated) attributes to represent one dimension of concerns, trusting them to convey its meaning. The first column has two measures of mortality; one considers the age of the dead, while the other ignores it. The second column has two measures of environmental impact, developed from the dozens of indicators used in different environmental impact analyses (Willis et al., 2004, 2005). The two right-hand columns have measures representing the two factors in Figure 2.

The display in Table 2, along with accompanying explanatory materials is designed to communicate the facts needed to rank risks based on the attributes that matter to people. Like any risk communication, they needed empirical evaluation, before being used for any serious purpose (Morgan et al., 2001b). That evaluation needed to approximate the conditions in which the materials were designed to be used: the sort of moderated, deliberative group process that any credible risk-ranking would entail.

To that end, an experimental test was created, with realistic profiles of 22 potential risks in a hypothetical middle school. Research participants ranked the risks, playing the roles of citizens advising the school board of a district with limited resources for managing risks. Each risk was described in a brochure that included a tabular summary like Table 2, along with a narrative description, subject to extensive pretesting.

The deliberative process sought to respect both individual and group perspectives, building on the EPA Guidebook and the US National Research Council's (1996) influential report, *Understanding Risk*. Before meeting as a group, individuals made personal risk rankings. At various points in the deliberations, the group publicly assessed its degree of consensus, while members privately recorded their personal views. Two different ways were used to elicit judgments so participants could triangulate on their values. The process assumed that these ranks needed to be constructed from individuals' basic values, as they reflected on the issues, informed by others' views (Fischhoff, 2005).

The method was evaluated with lengthy group sessions, involving both lay people and professional risk managers. Generally, participants tended to agree about the rankings, even when they disagreed about the importance of the attributes. Moreover, that agreement increased as the deliberations proceeded, without evidence of inappropriate group pressure. Details on the procedures and the evaluations can be found in Florig et al. (2001), Morgan et al. (2001a), and Willis et al. (2004, 2005), with exemplary materials at <http://sds.hss.cmu.edu/risk/>.

Table 2
A Standard Risk Characterization

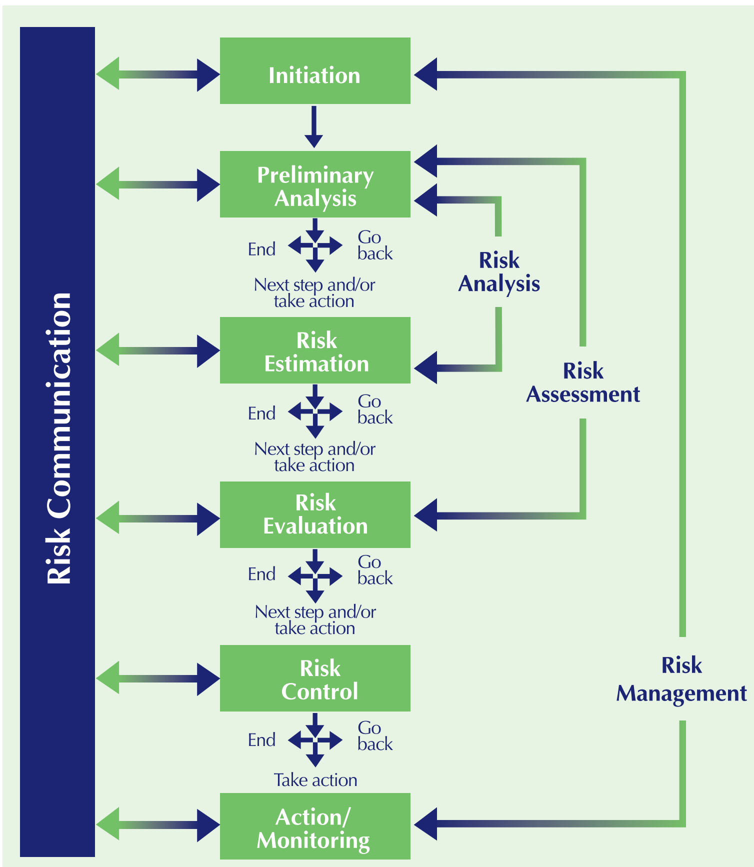
Number of People Affected	Environmental Impact	Knowledge	Dread
Annual expected number of fatalities	Ecosystem stress or change	Degree to which impacts are delayed	Catastrophic potential
0-450-600 (10% chance of zero)	50 km^2	1-10 years	1,000 times expected annual fatalities
Annual expected number of person-years lost	Magnitude of environmental impact	Quality of scientific understanding.	Outcome equity
0-9,000-18,000 (10% chance of zero)	modest (15% chance of large)	medium	medium (ratio = 6)

Source: Adapted from stimuli developed and used by Willis et al. (2005).

Risk Ranking in Practice

The Carnegie Mellon approach to risk ranking applies analytical and empirical risk research within the reality circumscribed by the EPA Guidelines. Its empirical evaluations suggest that it could be trusted to support real decisions, with a wide variety of risks and stakeholders. It is grounded in extensive research regarding what risk attributes matter to people, how to characterize them scientifically, and how to present them comprehensibly.

Figure 3
Steps in the Q850 Risk Management Decision-Making Process – Simple Model



Note: Risk communication with stakeholders is an important part of each step in the decision process.

A variant of the Carnegie Mellon approach has been endorsed by an initiative aimed at improving UK government risk management. Adapted through consultations with staff from several ministries, it is designed to be applied efficiently, without special training (HM Treasury, 2005). Called a method for "assessing concern," it characterizes risks on six attributes: familiarity, understanding, equity, dread, control, and trust. Risks are rated separately for how they are viewed by experts and by the public. These ratings complement scientific estimates of deaths and other harms, along with estimates of their monetary equivalents (to the extent possible).

In terms of the approach's suitability to Canadian conditions, Figure 3 presents a risk-management philosophy, promulgated by the Canadian Standards Association, which influenced the approach's development. The center column prescribes a risk-management process with standard steps – although with a noteworthy commitment to self-evaluation, not proceeding until a step has been satisfactorily accomplished. The left-hand bar prescribes continuing two-way interaction with the public. That interaction seeks to focus the process on public concerns and make its conclusions as credible as possible. The Carnegie Mellon risk-ranking approach could offer a scientifically sound approach to realizing this philosophy.

In this abstract representation, possible risk levels are measured on a single dimension called risk magnitude. On this scale, having zero risk means receiving no further attention. The height of each curve (a probability density function) shows the chances of having that risk level.

The narrowness of the curves for risks 1 and 3 means that they are relatively well understood. Their location on the scale shows that Risk 1 should clearly be ranked lower than Risk 3. The flatness of the curve for Risk 2 means it is much more poorly understood than either Risk 1 or Risk 2. Its rank is also less obvious. It will more likely have a much lower risk, but has some chance of having a higher risk. It might be given a higher rank by people who were especially concerned about large risks.

Individuals from four diverse groups rated 30 activities and technologies on nine attributes (e.g., voluntariness, dread). A statistical procedure (factor analysis) identified two underlying dimensions of risk. Risks high on the vertical factor, called "unknown risk," were rated as new. They are not well known to those exposed to them, not well known to science, involuntary, and with delayed effects. Risks high on the horizontal factor, called "dread risk," were rated as certain to fatal, if things go wrong, to threaten large numbers of people, and to evoke a feeling of dread. The four groups were students, League of Women Voters members, Active 20-30 Club members, and risk experts. The lines connect the highly similar results from the different groups.

References

Canadian Standards Association. 1997. *Risk Management: Guideline for Decision Makers* (reaffirmed 2002). Etobicoke: Canadian Standards Association.

Carnegie Mellon University, Center for Risk Perception and Communication. nd. Home page. <http://sds.hss.cmu.edu/risk/>.

Crouch, E.A., and R. Wilson. 1981. *Risk-Benefit Analysis.* Cambridge, MA: Harvard University Press.

Fischhoff, B. 2005. "Cognitive Processes in Stated Preference Methods." Pp. 937-968 in *Handbook of Environmental Economics*, ed. K.-G. Mäler and J. Vincent. Amsterdam: Elsevier.

———. (in press). "Risk Perception and Communication." Pp. 940-952 in *Oxford Textbook of Public Health.* Fifth edition, ed. R. Detels, R. Beaglehole, M.A. Lansang, and M. Gulliford. Oxford: Oxford University Press.

Fischhoff, B., S. Lichtenstein, P. Slovic, S.L. Derby, and R.L. Keeney. 1981. Acceptable Risk. New York: Cambridge University Press.

Fischhoff, B., P. Slovic, S. Lichtenstein, S. Read, and B. Combs. 1978. "How Safe Is Safe Enough? A Psychometric Study of Attitudes Towards Technological Risks and Benefits." *Policy Sciences* 9: 127-152.

Florig, H.K., M.G. Morgan, K.M. Morgan, K.E. Jenni, B. Fischhoff, P.S. Fischbeck, and M. DeKay. 2001. "A Deliberative Method for Ranking Risks (1): Overview and Test Bed Development." *Risk Analysis* 21: 913-922.

Long, J., and B. Fischhoff. 2000. "Setting Risk Priorities: A Formal Model." *Risk Analysis* 20: 339-351.

Morgan, K.M., M.L. DeKay, P.S. Fischbeck, M.G. Morgan, B. Fischhoff, and H.K. Florig. 2001a. "A Deliberative Method for Ranking Risks (2): Evaluation of Validity and Agreement among Risk Managers." *Risk Analysis* 21: 923-938.

Morgan, M.G., B. Fischhoff, A. Bostrom, and C. Atman. 2001b. *Risk Communication: The Mental Models Approach.* New York: Cambridge University Press.

Morgan, M.G., K. Florig, M. DeKay, and P. Fischbeck. 2000. "Categorizing Risks for Risk Ranking." *Risk Analysis* 20: 49-58.

Slovic, P. 1987. "Perception of Risk." *Science* 236: 280-285.

——— (ed.). 2000. *The Perception of Risk.* London: Earthscan.

Slovic, P., B. Fischhoff, and S. Lichtenstein. 1985. "Regulation of Risk: A Psychological Perspective." Pp. 241-278 in *Regulatory Policy and the Social Sciences*, ed. R. Noll. Berkeley: University of California Press.

United Kingdom, HM (Her Majesty's) Treasury. 2005. *Managing Risks to the Public.* London: HM Treasury.

US EPA (United States, Environmental Protection Agency). 1987. *Unfinished Business.* Washington, DC: Environmental Protection Agency.

———. 1990. *Reducing Risk: Setting Priorities and Strategies for Environmental Protection.* Washington, DC: Environmental Protection Agency.

———. 1993. *A Guidebook to Comparing Risks and Setting Environmental Priorities.* Washington, DC: Environmental Protection Agency.

US NRC (United States, National Research Council). 1996. *Understanding Risk.* Washington, DC: National Research Council.

———. 2006. *Scientific Review of the Proposed Risk Assessment Bulletin of the Office of Management and Budget.* Washington, DC: National Research Council.

Willis, H.H., M.L. DeKay, B. Fischhoff, and M.G. Morgan. 2005. "Aggregate and Disaggregate Analyses of Ecological Risk Perceptions." *Risk Analysis* 25: 405-428.

Willis, H.H., M.L. DeKay, M.G. Morgan, H.K. Florig, and P. Fischbeck. 2004. "Ecological Risk Ranking: Development and Evaluation of Method for Improving Public Participation in Environmental Decision Making." *Risk Analysis* 24: 363-378.