Risk Analysis, Vol. 24, No. 2, 2004

Ecological Risk Ranking: Development and Evaluation of a Method for Improving Public Participation in Environmental Decision Making

Henry H. Willis,[1]* Michael L. DeKay,[2] M. Granger Morgan,[2] H. Keith Florig,[2] and Paul S. Fischbeck[2]

This article reports an extension of the Carnegie Mellon risk-ranking method to incorporate ecological risks and their attributes. On the basis of earlier risk-perception studies, we identified a set of 20 relevant attributes for describing health, safety, and environmental hazards in standardized risk summary sheets. In a series of three ranking sessions, 23 laypeople ranked 10 such hazards in a fictional Midwestern U.S. county using both holistic and multiattribute ranking procedures. Results were consistent with those from previous studies involving only health and safety hazards, providing additional evidence for the validity of the method and the replicability of the resulting rankings. Holistic and multiattribute risk rankings were reasonably consistent both for individuals and for groups. Participants reported that they were satisfied with the procedures and results, and indicated their support for using the method to advise real-world risk-management decisions. Agreement among participants increased over the course of the exercise, perhaps because the materials and deliberations helped participants to correct their misconceptions and clarify their values. Overall, health and safety attributes were judged more important than environmental attributes. However, the overlap between the importance rankings of these two sets of attributes suggests that some information about environmental impacts is important to participants' judgments in comparative risk-assessment tasks.

KEY WORDS: Comparative risk assessment; ecological risk; environmental attributes; human health risk; risk attributes; risk perception; risk ranking

1. INTRODUCTION

Regulators and other risk managers receive frequent input from representatives of a wide variety of special interest groups. Sometimes, they also receive results from surveys in which members of the public have expressed their impressions of various risks. However, what they rarely receive, but very much need in a democracy, are judgments from samples of average citizens who have taken the time to develop thoughtful, informed views about a set of risks (Brown, 1996). Well-developed procedures to support risk ranking are needed to provide such input.

The U.S. Environmental Protection Agency (EPA) report entitled *Unfinished Business* highlighted the use of risk-ranking methods in environmental policy (U.S. EPA, 1987). That report compared the EPA's allocation of regulatory attention to a ranking of the importance of various risks by senior EPA managers and staff. Subsequently, EPA's Science Advisory Board has conducted two agencywide ranking projects (U.S. EPA, 1990, 2000), and the agency has supported several dozen local and regional

[1] RAND, Pittsburgh, PA, USA.
[2] Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA.
* Address correspondence to Henry H. Willis, RAND, 201 North Craig St., Suite 102, Pittsburgh, PA 15213; hwillis@rand.org.

0272-4332/04/0100-0363$22.00/1 © 2004 Society for Risk Analysis

comparative risk-ranking projects (Jones, 1997; Jones & Klein, 1999; Minard, 1996; U.S. EPA, 1993).

Although most of these risk-ranking projects have incorporated public participation at some level, the mechanisms for doing so have varied widely across projects; the range of participatory methods used has included surveys, focus groups, and citizen panels. For example, the Ohio Comparative Risk Project incorporated a 24-member public advisory group, facilitated meetings with over 6,000 citizens, surveyed more than 8,000 citizens, and polled 900 citizens by telephone (Morrone, 1995). Despite such efforts, researchers in these projects have not formally evaluated the validity of these methods for involving the public in risk-ranking exercises.

Over the past few years, researchers at Carnegie Mellon University have developed a method for eliciting informed judgments regarding health and safety hazards. The method is designed to elicit participants' relative concern about hazards by providing clear, concise, and consistently formatted information in a setting that encourages deliberation about the important characteristics of the hazards. In previous overviews (DeKay *et al.*, 2001; Florig *et al.*, 2001), we have described the five interdependent steps in this method (see Fig. 1). Briefly, risk experts: define and categorize the risks to be ranked, with input from community representatives (Step A); identify the relevant attributes of those risks (Step B); and describe the risks in a set of standardized *risk summary sheets* (Step C). Jury-like groups of laypeople or others then use these materials to rank the risks (Step D), and the investigators who facilitate the risk-ranking exercises report on the process and the resulting rankings (Step E).

In our initial studies, we applied this risk-ranking method to a *test bed* involving the health and safety risks present at the fictional Centerville Middle School (Florig *et al.*, 2001). Results indicated that participants were satisfied with both the risk-ranking process and results, and that the elicitation procedures produced rankings that were both internally consistent and similar among participants and among groups (DeKay *et al.*, 2001; Morgan *et al.*, 2001).

Although the method performs well in the domain of human health and safety risks, the school test bed is not conducive to the study of environmental risks. Ranking environmental risks requires information about hazards' impacts on ecological endpoints and judgments about the importance of those effects. Simultaneously ranking health, safety, and environmental risks requires participants to make difficult tradeoffs among a larger and more diverse set of relevant attributes. Unfortunately, we know much less about what information is relevant to the general public's assessments of environmental hazards, compared to health and safety hazards.

This article reports an extension of our risk-ranking method to incorporate environmental hazards and their attributes. Section 2 describes a new experimental test bed for ranking health, safety, and environmental risks; Section 3 provides an overview of the ranking procedures and analyses; and Section 4 describes three risk-ranking sessions involving participants from the lay public. The results, which are described in Section 5, allow for evaluation of the ranking method and participants' use of environmental attributes. When considered alongside results from previous ranking studies, these results provide additional evidence for the validity of the ranking method

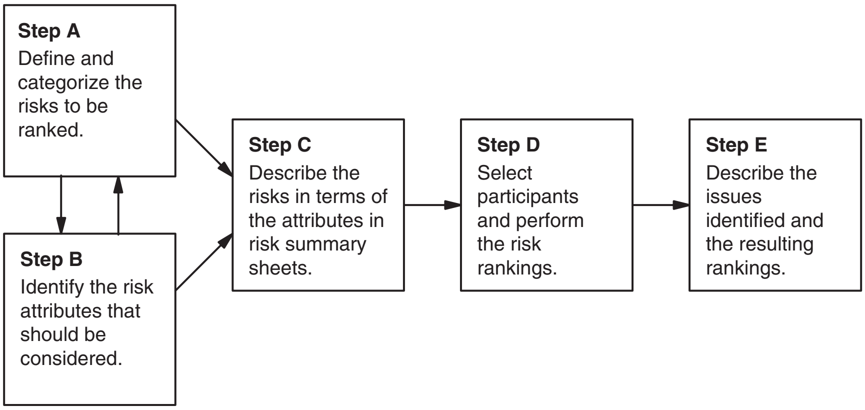

Fig. 1. Steps in the risk-ranking method. Steps A and B overlap in time and require some iteration. (*Source:* Florig *et al.* (2001))

and the replicability of the resulting rankings in these experimental test beds. Finally, Sections 6 and 7 provide additional discussion and conclusions.

2. EXTENDING THE RISK-RANKING METHOD TO ENVIRONMENTAL HAZARDS

To evaluate our risk-ranking method in a domain that includes environmental risks, we expanded the Centerville Middle School test bed to DePaul County, a fictional county in the Midwestern United States. The risk-ranking experiments were set in the context of county commissioners asking citizens to advise them on public concerns that the county can address through risk-management efforts. We developed a description of DePaul County that includes a county seat, residential centers, commercial and light industrial activities, a coal-fired power plant, a landfill, a typical transportation infrastructure, a navigable waterway, and a state park. The test bed also includes descriptions of several of the hazards existing in the county, as discussed below.

2.1. Categorizing the Hazards (Step A)

Although there are many ways to categorize hazards, risk ranking requires consistent choices regarding how hazards are grouped. For example, human health and safety hazards can be categorized in terms of predisposing conditions, proximal causes, outcomes, or in numerous others ways (DeKay *et al.*, 2001; Fischhoff & Willis, 2001; Morgan *et al.* 2000). Environmental hazards can be categorized in terms of the specific agent or stressor (e.g., lead), the activity giving rise to hazard (e.g., mining), the location (e.g., specific streams and rivers), the physical medium (land, water, or air), the endpoint of concern (e.g., biological diversity), and so on. Categorization is complicated because some risks can be considered both causes and effects. For example, acid rain could be considered a cause of ecological disturbances in affected forests or an effect of air pollution from coal-fired power plants.

Historically, comparative risk projects have used combinations of these classifications (Jones, 1997; Konisky, 2001; Minard, 1996; Morgenstern & Sessions, 1988). Unfortunately, there are no objective criteria for determining the best method of categorizing hazards for ranking purposes. Perhaps the best guidance is to define hazards in the manner that is most useful to the ultimate users of the rankings (Morgan *et al.*, 2000).

For the current study, we focused on activities and environmental stresses rather than on specific facilities or outcomes. Though the county test bed could be used to describe dozens of environmental and human health hazards, we had participants consider only the following 10 hazards: agricultural runoff, air pollution from electric power, food poisoning, genetically modified corn, invasive species, landfilling municipal solid waste, motor vehicle crashes, recreational motor boating, road salt and road-salt runoff, and transport of hazardous materials.

2.2. Identifying the Attributes (Step B)

Providing consistent attribute information about each hazard facilitates comparisons between hazards, but selecting an appropriate set of attributes is difficult (Jenni, 1997). The major challenge is choosing a set of attributes that comprehensively describes hazards' impacts while keeping the list short enough to be cognitively tractable. Each stakeholder group may prefer a different set of attributes. Although no single attribute set can fully satisfy every group's preferences, the chosen attribute set should not exclude any value set or perspective. In other words, it is more important for any group to be able to express their concerns through the selected attributes than for any single group to be completely satisfied with the attribute set.

2.2.1. Health and Safety Attributes

The literature on risk perception and previous studies on ranking health and safety risks provide a solid foundation for selecting attributes to describe hazards' health and safety impacts (Florig *et al.*, 2001; Jenni, 1997; Morgan, 1999; Morgan *et al.*, 2001; Slovic, 1992). Florig *et al.* identified four attributes for mortality impacts, four attributes for morbidity impacts, two attributes for uncertainty, and an attribute each for delay of effects and controllability. Extending the risk-ranking method to the county test bed required adding attributes to describe hazards' environmental impacts. To minimize the size of the extended attribute set, we reduced the number of human health attributes from 12 to 8 (see Fig. 2). We combined two mortality attributes that had described the risk of death in different statistical units. We dropped two attributes by differentiating injuries and illnesses by severity (i.e., more serious versus less serious

Road Salt and Road Salt Runoff

Summary:

As in other parts of the country which experience ice and snow in the winter, the DePaul County Highway Department and the Centerville Department of Public Works use salt as a deicing agent on roads in winter months. Salt runoff and spray can adversely affect plants and some other life forms. In some circumstances, sensitive trees near salted roads can be killed. Salt runoff can contaminate ground water. Salt also contributes to the accelerated deterioration of structural materials such as concrete and steel in structures such as bridges and auto bodies. This deterioration increases the risk of automobile-related injuries and fatalities.

Human Health and Safety Impacts	Low Estimate	**Best Estimate**	High Estimate
Risk of death			
For the average person –			
Chance in a million of death per year	4	**8**	15
Expected number of deaths per year	0.05	**0.1**	0.2
For the person at highest risk, chance in a million of death per year	11	**20**	35
Catastrophic potential, greatest number of deaths in a single event		**5–20**	
Risk of injury and illness			
Serious injuries and illnesses, number of cases per year	2	**4**	8
Minor injuries and illnesses, number of cases per year	4	**8**	15
Other factors			
Time between exposure and health effects		**immediate**	
Scientific understanding and predictability of health and safety impacts		**high**	
Ability of individual to control one's own exposure to health and safety risks		**medium**	

Environmental Impacts	Low Estimate	**Best Estimate**	High Estimate
Ecological effects			
Habitat affected –			
Acres	3,000	**5,300**	10,000
Square Miles	5	**8**	16
Animals killed or displaced, number		**few**	
Effects on variety of native species		**small**	
Ecological significance of affected species and habitat		**medium**	
Effects on natural processes and cycles		**low**	
Catastrophic potential, magnitude of worst-case effects		**low**	
Aesthetic effects			
Changes in landscape appearance		**small negative (–3)**	
Effects on noise, smell, taste, and visibility		**little or no change (0)**	
Other factors			
Time between exposure and environmental effects		**0–5 years**	
Duration of environmental effects, assuming the current activity or stress does not continue, but no other corrective actions are taken		**0–30 years**	
Scientific understanding and predictability of environmental impacts		**somewhat high**	
Negative effects on the environment's capacity to provide goods and services to people		**small**	

Fig. 2. Layout of the front page of a risk summary sheet showing the risk name, a summary paragraph, and a table of risk attributes. Additional pages include a few-paragraph narrative describing the risk in both national and local contexts, and a description of actions that local officials have taken to address the risk. A two- to four-page summary sheet was prepared for each of the 10 risks defined for DePaul County.

conditions) but not by duration (i.e., short-term versus long-term conditions). Finally, we combined the quality of scientific understanding and the predictability of health effects into a single attribute.

2.2.2. *Environmental Attributes*

The existing literature provides only limited guidance for choosing an appropriate set of attributes for describing the environmental impacts of hazards. A handful of psychometric studies on the perception of ecological risks (Lazo *et al.*, 2000; McDaniels *et al.*, 1997; McDaniels *et al.*, 1995, 1996) indicate that people perceive differences among ecological hazards on the following five factors: *impacts on species*, *impacts on humans*, *benefits to humans* (or *acceptability*), *avoidability* (or *controllability*), and *knowledge of impacts* (or *understandability*). In the McDaniels *et al.* (1997) and Lazo *et al.* (2000) studies, the first two of these factors (impacts on species and impacts on humans) collapsed into a single factor.

In a series of recent studies, we incorporated additional attributes from the literatures on ecology, conservation, and management, and from previous federal and state risk-ranking projects. In the first of these studies (Willis *et al.*, 2003, study 1), lay participants differentiated among risks on the five factors noted above, plus a sixth factor related to *impacts on aesthetics*. Oblique rotation of the factor solution (i.e., a rotation that did not constrain the factors to be orthogonal) indicated that the factors for ecological impacts, human impacts, and aesthetic impacts were correlated, as might be expected. The second study (Willis *et al*., 2003, study 2) focused on a subset of attributes related more closely to ecological risk, and replicated the appropriate portion of the factor solution (ecological impacts, aesthetic impacts, and scientific understanding). In addition, these studies indicated that factor solutions from aggregate data are useful in predicting riskiness judgments at both the aggregate and individual levels. In a third study (Willis & DeKay, 2003), analysis of data from four different stakeholder groups (laypeople, environmentalists, and environmental professionals in government and industry) indicated that the groups perceived risks similarly (i.e., they had very similar factor structures). However, there were some differences among groups in the way that these factors were related to judgments of riskiness. Compared to the other groups, laypeople placed less weight on ecological impacts and more weight on aesthetic impacts and scientific understanding.

On the basis of this research and other pilot studies, we selected 12 attributes to describe the environmental aspects of hazards present in DePaul County (see Fig 2). Definitions of all 20 health, safety, and environmental attributes appear in the Appendix.

2.3. Describing the Hazards (Step C)

The design of materials for the current study was guided by previous research on risk communication and risk ranking (DeKay *et al*., 2001; Florig *et al*., 2001; Morgan, 1999; Morgan *et al.*, 2002). We developed two- to four-page 8.5 × 11-inch risk summary sheets that describe each of the 10 hazards selected for DePaul County. As an example, the summary sheet for road salt and road-salt runoff appears in Fig. 2. Each summary sheet begins with a short narrative paragraph describing the scope of the hazard. The first page also includes a table of quantitative and qualitative data about the hazard for each of the 20 attributes discussed above. The remaining pages address three topics: (1) the sources, general mechanisms, and magnitudes of the hazard's impacts; (2) the sources and extent of the hazard in DePaul County; and (3) the efforts already made in the county to control the hazard's impacts. Descriptions of the hazards are based on the relevant literature and discussions with professionals familiar with the hazards. Quantitative data are based on national averages, but are adjusted to plausible levels of risk for the fictional county.

To help participants interpret the 20 attributes on the first page of the risk summary sheets, we created a separate 8.5 × 11-inch pamphlet, titled "Notes on the Numbers," that provides definitions for these attributes. To facilitate comparisons of hazards on these attributes, we also created an 11 × 17-inch sheet that contains rankings of the 10 hazards in terms of each of the 20 attributes.

3. ELICITING AND EVALUATING RANKINGS OF HAZARDS

3.1. Ranking the Hazards (Step D)

We employed the standard method described by Florig *et al*. (2001) to elicit participants' rankings of the risks. Participants produced their initial rankings individually, then worked with others to produce group rankings, and finished by producing their final rankings individually. In each of these three stages, hazards were ranked using two procedures. First, participants ranked the hazards directly, using the risk summary sheets. These rankings are called *holistic* rankings. Second, participants ranked the attributes of the hazards, and a model was used to produce hazard rankings from these attribute rankings. The resulting hazard rankings are called *multiattribute* rankings. In the first two stages, the participants and groups were given the opportunity to reconcile differences between their holistic and multiattribute risk rankings to create *revised* rankings. In all, eight separate rankings of the hazards were collected in Step D of Fig. 1 (Steps D1–D8, respectively). The procedures used to elicit these rankings are discussed in greater detail in Section 4.3.

3.2. Assessing Consistency, Satisfaction, and Agreement

In order to provide useful input to public risk-management decision making, risk-ranking exercises must yield valid results (Morgan *et al.*, 2001). In our studies, participants' rankings of hazards (i.e., their levels of concern) are assessed using both holistic and multiattribute procedures. The correlation between these rankings is a measure of internal consistency or *convergent validity*. Ideally, such consistency will increase over the course of the ranking exercise as discussions provide participants with additional knowledge and opportunities to reconsider their judgments.

If participants are satisfied with the ranking process and believe that the results faithfully represent their concerns, then the method has *face validity*. We assess participants' satisfaction explicitly, with a final questionnaire, and implicitly, with multiple regression models of their rankings of the hazards.

Producing complete agreement among participants or among groups is not a goal of our risk-ranking method (Morgan *et al.*, 2001). Different values, interests, and experiences provide a basis for valid disagreement, and ignoring this fact does a disservice both to participants and to decisionmakers. In some instances, identifying important areas of agreement and disagreement may be as important for risk managers as determining a specific ranking of hazards. Even so, effective risk communication and discussion should help reduce confusion and misunderstanding, and lead to greater agreement among the rankings of different parties over the course of the exercise.

4. METHODS

4.1. Participants

Twenty-three participants took part in the risk-ranking sessions: eight each in the first two sessions (Groups 1 and 2) and seven in the third session (Group 3). Sixteen of the participants were contacted through a community organization in Allegheny County, Pennsylvania; the remaining seven (not all in Group 3) responded to personal solicitations at Carnegie Mellon University. Participants' ages ranged from 23 to 80 years, with a mean of 53 years. Fourteen participants were female. All had graduated from high school; eight had attended a two-year college or trade school; five had obtained a bachelor's degree but not a graduate degree; and one had obtained a graduate degree. In return for their participation, we paid each person $100 or made an equivalent donation to the community organization.

4.2. Materials

Materials for this study included a description and map of DePaul County, the risk summary sheets for the 10 hazards listed in Section 2.1, the "Notes on the Numbers" pamphlet, and the large sheet containing rankings of the 10 hazards by each of the 20 attributes. These materials are available from the first author upon request or electronically at http://www.epp.cmu.edu/research/EPP_risk.html.

4.3. Procedures

One week prior to the risk-ranking session, we mailed each participant a package containing the risk summary sheets and supporting documents describing DePaul County. Working independently, participants reviewed these materials and ranked the hazards in terms of their level of concern, with 1 being the hazard of greatest concern. For this *initial individual holistic* ranking (Step D1), participants were free to use the information contained in the risk summary sheets and any other information or experiences that they thought relevant. Participants spent between 0.75 and 7.5 hours on this task, averaging about three hours.

We facilitated the remaining tasks (Steps D2–D8) during organized workshops that lasted between six and eight hours, held on Saturdays at Carnegie Mellon University. At the beginning of the session, we reviewed the materials that participants had been mailed, with particular attention to the 20 attributes on the front of the risk summary sheets.

For many attributes, such as mortality or the amount of habitat affected, people tend to agree about whether high or low values are worse. For other attributes, the direction corresponding to a worse condition may not be obvious (Morgan *et al.*, 2001). For five of the attributes listed in Fig. 2 (those related to the time between exposure and health or environmental effects, the scientific understanding and predictability of those effects, and one's ability to control his or her exposure to health and safety risks), participants worked independently to indicate whether high or low values are associated with greater concern. Participants then ranked the 20 attributes according to their relative importance, with 1 being the most important attribute. Participants were told not to rank attributes that did not affect their level of concern for the hazards.

We used these attribute rankings to calculate a total concern score for each hazard using a linear additive multiattribute model

$$\text{Concern}_j = \sum_{i=1}^{n} w_i \times v_i(x_{ij}), \tag{1}$$

where j is a hazard, i is an attribute, n is the number of attributes, w_i is the importance weight for attribute i, v_i is the value function for attribute i, and x_{ij} is hazard j's level on attribute i. In order to apply such a model, one must elicit or assume attribute value functions and attribute weights. To determine the value function for each attribute, we first ranked the levels of the attribute. For example, the three levels for the attribute "number of animals killed or displaced" (i.e., *none or almost none*, *few*, and *many*) were ranked 1 (best), 2, and 3 (worst), respectively. These ranks were then normalized to range from 0 to 1, leading to final values of 0 for *none or almost none*, 0.5 for *few*, and 1 for *many*.

We considered several alternative methods for converting attribute rankings to attribute importance weights. In a pilot study of 32 risk managers, we evaluated the rank-order-centroid, reciprocal-of-the-rank, and rank-sum attribute weighting functions (Barron & Barrett, 1996; Jia *et al.*, 1998) by assessing the consistency of the resulting multiattribute risk rankings with participants' holistic rankings of the 10 DePaul County hazards. Mean Spearman correlations with holistic rankings were 0.54, 0.51, and 0.53 for the three functions, respectively. In our previous studies of health and safety risks, we have used the reciprocal-of-the-rank attribute weighting function (Morgan, 1999; Morgan *et al.*, 1999; Morgan *et al.*, 2001). For the current study, we used the rank-sum function because it was simpler to implement and explain during the ranking workshops and because it has the property of distributing attribute weights more evenly across a larger number of attributes (Jia *et al.*, 1998). The following rank-sum formula converts attribute ranks to estimated attribute weights

$$w_i = \frac{n + 1 - R(i)}{\sum_{i=1}^{n} R(i)}, \tag{2}$$

where i is an attribute, $R(i)$ is the rank of that attribute, and n is the number of ranked attributes. If a participant did not use a particular attribute, that attribute was not considered in the calculation, and received a weight of zero. We used participants' reported attribute ranks directly, without reranking them to eliminate skipped ranks (e.g., rankings such as 1, 2, 3, 10 were not altered). In some cases, the highest reported ranks were greater than the number of nonzero ranks, n, and Equation (2) produced negative importance weights for those attributes. To avoid negative weights while still using participants' reported attribute ranks, we employed the following modified formula:

$$w_i = \frac{R_{\max} + 1 - R(i)}{\sum_{i=1}^{n} [R_{\max} + 1 - R(i)]}, \tag{3}$$

where $R_{\max}$ is the highest reported rank.

Attribute importance weights from Equation (3) were normalized so that they summed to 100 and then used in Equation (1). Finally, the resulting concern scores were ranked so that 1 represented the hazard of greatest concern, to yield *initial individual multiattribute* risk rankings (Step D2).[3]

These rankings were calculated and returned to participants for their consideration, along with spreadsheet output reflecting some of the intermediate results. The goal of this review was to help participants better understand the characteristics of the hazards and how their weights for these characteristics were related to their levels of concern for the 10 hazards. After considering discrepancies between their holistic and multiattribute risk rankings, participants produced *initial individual revised* rankings (Step D3).

Next, participants worked together to rank the risks. Each group of seven or eight people produced a single *group holistic* ranking (Step D4). Groups then ranked the attributes in terms of their importance, and we used this information to construct *group multiattribute* risk rankings, again using Equation (3) to determine the attribute weights (Step D5). Each group compared its holistic and multiattribute risk rankings and produced a *group revised* ranking (Step D6). Although we facilitated these discussions, groups were free to determine their own decision processes. Typically, participants referred to their initial individual revised rankings and provided reasons why they thought particular hazards should be ranked high or low. Persuasion and opinion change appeared to be

[3] For the 49 multiattribute risk rankings reported in this article (23 in Step D2, 3 in Step D5, and 23 in Step D7), the risk rankings implied by Equations (2) and (3) were identical in 43 instances (when negative importance weights from Equation (2) were set to zero). In the few cases where the rankings differed, the lowest correlation between rankings from the two weighting procedures was 0.96. Subsequent analyses involving multiattribute risk rankings were not sensitive to these small differences.

the norm, but there were many instances in which individual participants acquiesced in order to achieve a group ranking.

Because it is unrealistic to expect complete consensus, participants were given the opportunity to dissent from their group's rankings. Working independently, participants compared their group's revised ranking (Step D6) to their initial individual revised rankings (Step D3) and produced *final individual holistic* rankings (Step D7). Participants also compared their group's attribute ranking (Step D5) to their initial individual attribute rankings (Step D2) and produced final individual attribute rankings, which were used to calculate participants' *final individual multiattribute* risk rankings (Step D8).

At the end of the session, participants evaluated several aspects of the ranking process and results using seven-point scales. First, participants indicated how much their current knowledge of the risks and their final individual holistic rankings were influenced by completing their initial individual rankings and by their group's deliberation processes. Next, participants reported their levels of satisfaction with their group's decision-making process and ranking. Third, participants indicated how strongly they would approve or disapprove of using the results of their ranking session as an input to risk-management decisions in the fictional county, and how strongly they would approve or disapprove of actual government agencies using the method to obtain public input for policy decisions.

5. RESULTS

Responses from individual participants in the same risk-ranking sessions could not be considered independent after the group discussions. Despite the fact that there were only three groups, the statistics reported below were computed at the individual participant level, averaged within groups, and analyzed at the group level. For ease of comparison, the same procedures were also used for individual participant data collected before the group discussions.

5.1. Consistency Between Holistic and Multiattribute Risk Rankings

Consistency between holistic and multiattribute risk rankings was assessed using Spearman correlations, both for individual participants and for groups. Results are reported in Table I.

Table I. Consistency (Mean Spearman Correlations Between Holistic and Multiattribute Risk Rankings)

Rankings	Group 1	Group 2	Group 3	Average
Initial individual (Steps D1 and D2)	0.56	0.43	0.43	0.47
Group (Steps D4 and D5)	0.71	0.85	0.57	0.71
Final individual (Steps D7 and D8)	0.64	0.79	0.63	0.69

The mean Spearman correlation between participants' initial individual holistic and multiattribute risk rankings (Steps D1 and D2) was significantly greater than zero, $M = 0.47$, $t(2) = 11.27$, $p = 0.0078$, as was the mean Spearman correlation between participants' final individual holistic and multiattribute risk rankings (Steps D7 and D8), $M = 0.69$, $t(2) = 12.89$, $p = 0.0060$. Despite the notable increase in consistency, the difference between the initial and final correlations was not significant, $M = 0.21$, $t(2) = 2.65$, $p = 0.1175$, because of the small number of groups. The mean Spearman correlation between groups' holistic and multiattribute risk rankings (Steps D4 and D5) was 0.71, $t(2) = 8.71$, $p = 0.0129$.

These correlations are similar to those from our previous study in which risk managers considered health and safety risks (Morgan *et al.*, 2001), except that the increase in consistency from initial to final individual risk rankings is larger (though not significant) in the current study.

5.2. Participants' Evaluations of Procedures and Rankings

5.2.1. Self-Reported Evaluations

Compared to the midpoints of the response scales (3 on scales with 0 = *not at all* and 6 = *very much*), participants indicated that they learned from their groups' discussions (Steps D4–D6), $M = 4.82$, $t(2) = 20.27$, $p = 0.0024$, but not from completing their initial individual holistic rankings (Step D1), $M = 3.35$, $t(2) = 2.34$, $p = 0.1439$. The mean rating for how much participants learned from their initial individual multiattribute risk rankings (from Step D2) was marginally greater than the scale midpoint, $M = 3.78$, $t(2) = 4.25$, $p = 0.0511$. Similarly, participants indicated that their final individual holistic rankings (Step D7) were influenced by their groups' revised rankings (Step D6), $M = 4.21$, $t(2) = 5.80$, $p = 0.0285$, but not by their initial individual holistic rankings (Step D1),

$M = 3.44$, $t(2) = 1.01$, $p = 0.4174$, or their initial individual multiattribute risk rankings (from Step D2), $M = 3.55$, $t(2) = 1.85$, $p = 0.2059$.

On a similar 0–6 scale, participants indicated that their groups considered and discussed different viewpoints and encouraged each member to express his or her opinion, $M = 4.96$, $t(2) = 47.00$, $p = 0.0005$. However, there was no clear tendency for groups to resolve disagreements among members in a particular manner. On a scale with -3 = *voting or averaging* and 3 = *persuasion and opinion change*, the mean response was not significantly different from zero, $M = 0.71$, $t(41) = 0.80$, $p = 0.5099$.

Compared to the scale midpoints (0 on scales with -3 = *very dissatisfied* and 3 = *very satisfied*, for example), participants reported that they were satisfied with their groups' decision processes, holistic rankings, attribute importance rankings, multiattribute risk rankings, and revised rankings, all Ms ≥ 1.61, all ts ≥ 4.45, all ps ≤ 0.0470, and indicated that their groups' revised rankings represented their concerns well, $M = 1.57$, $t(2) = 7.20$, $p = 0.0187$. Finally, participants supported using the risk-ranking process to advise policy both in the fictional county, $M = 1.90$, $t(2) = 11.09$, $p = 0.0080$, and in real-world decision-making situations, $M = 2.10$, $t(2) = 22.00$, $p = 0.0021$. Responses to these eight satisfaction questions were highly correlated (mean Pearson correlation between measures = 0.58), indicating that an overall feeling of satisfaction (i.e., a halo effect) may have influenced responses. Nonetheless, these encouraging results are consistent with those reported in our previous work involving only health and safety risks (Morgan *et al.*, 2001).

5.2.2. Regression-Based Evaluations

Participants' implicit satisfaction with the procedures in our method was assessed using within-participant regressions. For example, participants' initial individual revised rankings (Step D3) were regressed onto their initial individual holistic rankings (Step D1) and their initial individual multiattribute risk rankings (Step D2). The mean regression coefficients for holistic and multiattribute rankings were both significantly greater than zero, $M(b_{\text{D1}}) = 0.62$, $t(2) = 8.27$, $p = 0.0143$, and $M(b_{\text{D2}}) = 0.38$, $t(2) = 4.45$, $p = 0.0470$, respectively. Three participants stuck with their holistic rankings, one participant adopted her multiattribute risk ranking, and 19 participants reconciled the differences between their holistic and multiattribute risk rankings in creating their revised rankings. These results indicate that participants learned from constructing their multiattribute risk rankings, and that this new knowledge influenced their revised rankings. This finding does not contradict participants' explicit reports that their initial individual multiattribute rankings had little influence on their *final* individual holistic rankings.

In contrast to the results for participants' initial individual rankings, groups did not appear to incorporate their multiattribute rankings into their revised rankings. When groups' revised rankings (Step D6) were regressed onto their holistic rankings (Step D4) and their multiattribute risk rankings (Step D5), the mean regression coefficients for holistic and multiattribute risk rankings were $M(b_{\text{D4}}) = 0.96$, $t(2) = 39.84$, $p = 0.0006$, and $M(b_{\text{D5}}) = 0.04$, $t(2) = 1.79$, $p = 0.2149$, respectively. Indeed, one of the three groups did not revise its holistic ranking at all in light of its multiattribute risk ranking. It appears that the usefulness of the multiattribute ranking procedure is greater in the earlier individual stage than in the group deliberation stage. There are two potential explanations for this finding. First, refinements to initial individual rankings on the basis of the multiattribute procedure may be carried forward and incorporated into the groups' holistic rankings, which therefore need little additional refinement. Second, the extensive discussions and effort devoted to generating the group holistic rankings may contribute to stronger support for these rankings than for the initial individual holistic rankings. These two explanations are not incompatible.

A similar analysis may be used to assess participants' implicit satisfaction with their groups' revised rankings. Participants' final individual holistic rankings (Step D7) were regressed onto their own initial individual revised rankings (Step D3) and their groups' revised rankings (Step D6). The mean regression coefficient for the initial individual ranking was not significantly different from zero, $M(b_{\text{D3}}) = 0.17$, $t(2) = 2.61$, $p = 0.1205$, but the mean regression coefficient for the group ranking was, $M(b_{\text{D6}}) = 0.84$, $t(2) = 18.34$, $p = 0.0030$. In one session, four participants adopted their group's revised ranking as their final individual holistic ranking. No participants reverted to their own initial individual revised rankings. Apparently, participants were more satisfied with their groups' rankings than with their own rankings from earlier in the day.

Taken together, these self-report and regression-based measures indicate that participants were satisfied with the ranking processes and results. These

Table II. Agreement (Mean Pairwise Spearman Correlations Between Holistic Risk Rankings, Between Multiattribute Risk Rankings, and Between Revised Risk Rankings)

Rankings	Group 1	Group 2	Group 3	Average
Initial individual				
Holistic (Step D1)	0.30	0.57	0.29	0.39
Multiattribute (Step D2)	0.56	0.66	0.80	0.67
Revised (Step D3)	0.40	0.59	0.44	0.48
Group				
Holistic (Step D4)				0.67
Multiattribute (Step D5)				0.87
Revised (Step D6)				0.71
Final individual				
Holistic (Step D7)	0.79	0.97	0.85	0.87
Multiattribute (Step D8)	0.66	0.85	0.93	0.81

findings are very similar to those from our earlier study involving only health and safety risks (Morgan *et al.*, 2001).

5.3. Agreement Among Individuals and Among Groups

Mean pairwise Spearman correlations of rankings provide measures of agreement among individuals and among groups. Results are reported in Table II.

5.3.1. Agreement Among Individuals Within Groups

For each group, Spearman correlations were computed between individuals' rankings for all possible pairs of participants, and those correlations were averaged to create measures of agreement among participants within groups. Each of these 15 mean pairwise correlations (five sets of individual rankings from each of the three groups) was significantly different from zero, all r_{avg}s ≥ 0.29, all χ^2_rs ≥ 24.69, all ps ≤ 0.0033 (see Table II).[4] The mean pairwise Spearman correlations between final individual holistic rankings (Step D7), $M = 0.87$, were significantly greater than those between initial individual holistic rankings (Step D1), $M = 0.39$, $t(2) = 9.81$, $p = 0.0102$, and those between initial individual revised rankings (Step D3), $M = 0.48$, $t(2) = 33.81$, $p = 0.0009$. One explanation for this increase in agreement is that the group deliberation process helped participants to correct their misconceptions about the hazards and come to a common understanding regarding the relative importance of the hazards' characteristics. An alternative explanation is that the procedures forced participants into agreement. However, strong evidence for participants' satisfaction with the procedures and the resulting rankings suggests that the former explanation is more plausible. Again, these results are similar to those reported by Morgan *et al.* (2001).

[4] Mean pairwise Spearman correlations may be tested for significance by first transforming them to Kendall's coefficient of concordance, $W = [r_{avg}(m-1)+1]/m$, where r_{av} is the mean pairwise correlation and m is the number of rankings. The result of a second transformation, $\chi^2_r = Wm(n-1) = [r_{avg}(m-1)+1](n-1)$, where n is the number of items being ranked, is distributed as χ^2 with $n-1$ degrees of freedom. The test based on χ^2_r is often referred to as the Friedman test (Kendall & Gibbons, 1990).

5.3.2. Agreement Among Groups

Spearman correlations were also computed between groups' rankings for the three possible pairs of groups, and those correlations were averaged to create measures of agreement among groups. This was done separately for the holistic, multiattribute, and revised rankings. Each of these three mean pairwise correlations was significantly different from zero, all r_{avg}s ≥ 0.67, all χ^2_rs ≥ 21.14, all ps ≤ 0.0121 (see Table II and footnote 2). Groups' revised rankings (Step D6) of the 10 hazards in DePaul County are shown in Table III. The mean pairwise Spearman correlation between these rankings was 0.71, compared to 0.83 for the mean correlation between group holistic rankings in our study of health and safety risks (Morgan *et al.*, 2001). For reference, the hazards in Table III are sorted by the risk of death for the average person (i.e., expected mortality). The mean Spearman correlation between groups' revised rankings and the ranking by expected by mortality was 0.44, significantly greater than zero, $t(2) = 5.71$, $p = 0.0293$. The analogous mean correlation in our earlier study was 0.71.

Air pollution from electric power and motor vehicle accidents were the only hazards ranked as the hazard of greatest concern. Genetically modified corn and recreational motor boating were always the hazards of least concern. Agricultural runoff and invasive species, hazards described as primarily affecting environmental endpoints, were ranked higher than food poisoning, a hazard affecting primarily health and safety endpoints, again suggesting that participants paid attention to environmental attributes.

Groups' rankings of the 20 attributes used to construct groups' multiattribute risk rankings (Step D5) are shown in Table IV. The mean pairwise Spearman correlation between these attribute rankings was 0.62, significantly greater than zero, $\chi^2_r = 42.61$, $p = 0.0015$

Table III. Group Revised Rankings of Hazards

Hazard	Ranking by Expected Mortality	Group Revised Rankings			
		Group 1	Group 2	Group 3	Average[a]
Air pollution from electric power[b]	1	1	1	2.5	1
Motor vehicle crashes[c]	2	2	3	1	2
Road salt and road-salt runoff[b]	3	6	5	7	6.5
Food poisoning[c]	4	4	8	8	8
Recreational motor boating[b]	5	9	9	9	9
Transport of hazardous materials[b]	6	7	6	4	5
Agricultural runoff[d]	7	5	2	5	3
Genetically modified corn[d]	9	10	10	10	10
Invasive species[d]	9	8	4	6	6.5
Land disposal of municipal solid waste[b]	9	3	7	2.5	4
Spearman correlation with expected mortality		0.58	0.44	0.31	0.44[e]

[a]Group revised rankings were averaged across sessions and reranked.
[b]Effects on both health and environmental endpoints.
[c]Effects primarily on health endpoints.
[d]Effects primarily on environmental endpoints.
[e]The average of the Spearman correlations for the three groups, not the Spearman correlation for the average ranking.

Table IV. Group Rankings of the Importance of Hazard Attributes

Attribute	Group 1	Group 2	Group 3	Average[a]
Human health and safety impacts				
Risk of death				
For the average person	1	11	2	3
For the person at highest risk	2	5	3	2
Catastrophic potential	3	1	1	1
Injury and illness				
Serious injuries and illnesses	4	12	4	6
Minor injuries and illnesses	9	14	10	10.5
Other factors				
Time between exposure and health effects	10	18.5	19.5	16
Scientific understanding and predictability of health effects	7	18.5	9	12
Ability of individual to control one's own exposure to health and safety risks	5	3	7	4
Environmental impacts				
Ecological effects				
Habitat affected	17	4	12	10.5
Animals killed or displaced	15	13	13	14
Effects on variety of native species	16	10	16	15
Ecological significance of affected species and habitat	14	9	14	13
Effects on natural processes and cycles	12	8	6	9
Catastrophic potential	8	7	8	7
Aesthetic effects				
Changes in landscape appearance	20	16	18	19
Effects on noise, smell, taste, and visibility	19	15	15	18
Other factors				
Time between exposure and environmental effects	18	18.5	19.5	20
Duration of environmental effects	11	2	11	8
Scientific understanding and predictability of environmental effects	13	18.5	17	17
Negative effects on the environment's capacity to provide goods and services to people	6	6	5	5

[a]Group attribute rankings were averaged across sessions and reranked.

(see footnote 2). The analogous mean correlation in our earlier study was 0.35. It is encouraging that independent groups can achieve this level of agreement regarding the relative importance of such a large set of attributes.

Overall, health and safety attributes were judged more important than environmental attributes. Mean rankings for health and safety attributes and environmental attributes (computed separately for each group and then averaged across groups) were 7.48 and 12.51, respectively. However, this difference was largely due to the three mortality attributes, which were ranked more important than the other five health and safety attributes, $M = 3.22$ and $M = 10.03$, respectively. Indeed, the three most important attributes in the average ranking were all related to the risk of death (see Table IV).

There were some differences between groups, with Group 1 placing the most emphasis on health and safety attributes and Group 2 placing the most emphasis on environmental attributes. Even so, each group ranked two or more environmental attributes as more important than some health and safety attributes, and four environmental attributes made the top 10 in the average ranking (see Table IV). These attributes were "negative effects on the environment's capacity to provide goods and services to people" (average rank = 5); "catastrophic potential, magnitude of worst-case effects" (7); "duration of environmental effects" (8); and "effects on natural processes and cycles" (9). Participants placed less emphasis on species and habitats, and very little emphasis on aesthetic effects. These results suggest that participants' environmental concerns were based on large-scale disruptions of ecological functioning rather than on more specific effects. Other studies that focused solely on environmental risks have indicated that aesthetic effects may be somewhat more important than they appear in the current study (Willis *et al.*, 2003; Willis & DeKay, 2003).

6. DISCUSSION

The results in Section 5 provide evidence for the validity and robustness of the Carnegie Mellon risk-ranking method and the replicability of the resulting rankings. Consistency measures indicate that participants and groups had relatively clear definitions of risk. Self-report and regression-based measures of satisfaction show that participants were satisfied with the ranking procedures and results. In particular, regression results indicate that participants found the multiattribute procedure useful when developing their initial individual rankings, but less so when developing group rankings. Participants were also supportive of using such a method to advise real-world risk-management decisions. Agreement among participants increased over the course of the ranking exercise, suggesting that the materials, procedures, and deliberations helped participants to correct their misconceptions or clarify their values. Overall, these results are very similar to those observed in our previous work on health and safety risks.

Developing the risk summary sheets involved considerable effort in identifying a set of attributes to describe the hazards' ecological impacts (Willis & DeKay, 2003; Willis *et al.*, 2003). In this study, analyses of groups' risk and attribute rankings provide insight into how participants used these environmental attributes. Groups placed more importance on health and safety impacts (particularly on the risk of death) than on environmental impacts, but there was some variation among groups. The most important environmental attributes were those reflecting substantial ecological disturbance. Of course, the reported rankings depend on the sets of risks and attributes selected, the way this information was communicated in the risk summary sheets, and the people who participated in this study. Although the rankings are useful for demonstrating the feasibility of using our method to rank health, safety, and environmental risks simultaneously, these results should not be used to inform specific risk-management decisions.

The results of this study suggest that comparative risk assessments should include information on at least some important environmental attributes of hazards. Although participants in this study appeared to pay little attention to attributes related to aesthetics, other studies of ecological risk perception reveal that aesthetic considerations may be significantly related to riskiness judgments (Willis & DeKay, 2003; Willis *et al.*, 2003). At this time, it would be premature to exclude such attributes from comparative risk assessments simply because the groups in this study did not find them important. It may be best to include a broad set of attributes, similar to the set used in our risk summary sheets, so that all participants will be able to make risk judgments that reflect their full range of concerns.

The ranking method presented here was designed to determine the public's concerns about hazards for the purpose of informing risk-management decisions. We believe that the method is most appropriate (and most likely to succeed) when there is a genuine motivation to achieve this goal. The method is not expected to be immune to the difficulties faced by other stakeholder participation processes when risk-management issues are highly politicized or linked to broader agendas that have little or nothing to do with the specific risks at hand. Complications could also arise in real-world applications if the organization preparing the risk summaries and conducting the ranking procedures is not viewed as neutral, honest, and trustworthy.

Another potential limitation is that the underlying science may not be adequate for all of the hazards considered. This concern is relevant to all comparative risk projects, not just those using the method presented here. A distinct advantage of the current method is that it makes the level of scientific uncertainty evident to the participants in two ways: by presenting a range of estimates for quantitative attributes and by summarizing this information in a pair of attributes for "scientific understanding and predictability" (one for health and safety effects and one for environmental effects). Being explicit about the quality of the science is important because it is usually necessary to make risk-management decisions on the basis of imperfect knowledge.

Finally, it is reasonable to ask how the method presented here might fare in an application with many more participants, hazards, or attributes. The number of participants does not present a serious problem. Although there is some practical limit on the number of participants in each group (in our experience, six to eight participants per group seems about right), there is no such limit on the number of groups that can be convened. For example, our studies of health and safety risks in the Centerville Middle School test bed involved more than 200 participants in over 40 groups (Morgan *et al.*, 2001). Using multiple small groups to perform rankings helps to ensure that the results are representative of the public's concerns, and not those of a few persuasive or charismatic individuals. The level of agreement (or disagreement) among group rankings can be assessed using procedures similar to those in Section 5.3.2.

Regarding the number of hazards, we note that we have conducted numerous ranking sessions using a larger set of 22 health and safety hazards in the school test bed (DeKay *et al.*, 2001). In practice, the ranking procedures begin with a coarse evaluation of each risk (e.g., three piles for risks of low, medium, or high concern), followed by more focused comparisons involving two or three risks at a time. Because participants are not required to keep all of the risks in mind simultaneously, increasing the number of risks does not require major changes in the ranking procedures, provided that adequate time is allowed for the full consideration of all risks. Even so, there may be some upper limit on participants' tolerance for making detailed risk comparisons. When time or other resources are in short supply, it may be wise to use broader risk categories in order to reduce the number of risks that participants must consider (Morgan *et al.*, 2000).

The constraint on the number of attributes may be more serious. Even though the ranking method is designed to focus participants' attention on those attributes that they think are most important, we believe that participants would have difficulty working with more than about 20 attributes. The empirical work cited in Section 2.2. suggests that the attributes used in this study provide an excellent basis for describing the health, safety, and environmental effects of hazards. However, if quality-of-life considerations (e.g., the social effects of diseases or crimes) are to be incorporated into a comparative risk assessment, the current set of attributes may need to be trimmed to make room for new attributes that capture these aspects of hazards (just as we trimmed our original list of health and safety attributes to make room for environmental attributes in this study).

Notwithstanding the limitations noted above, potential applications of this risk-ranking method exist in a variety of domains and at various geographic scales. The method could be used for a countywide risk assessment, as in our test bed, or it could be used to evaluate risks to workers and the public from a single industrial facility. On a larger scale, rankings could be elicited as input to a corporationwide risk assessment or an agencywide prioritization of resources for hazard management. Ultimately, the applicability and limits of this ranking method can be assessed only through actual field implementations. We encourage others to use this risk-ranking method and the principles on which it is based when evaluating risks in such situations.

7. CONCLUSIONS

High levels of consistency between rankings based on different procedures, high levels of satisfaction with the processes and the resulting rankings, and high levels of agreement among individual participants and among groups suggest that this deliberative method can produce rankings of health, safety, and environmental risks that are suitable for use as input to public risk-management decisions. Although groups placed more weight on health and safety attributes than on environmental attributes, they also considered attributes reflecting significant environmental impacts to be important. To ensure that the public's values can be well represented in future comparative risk projects, hazards should be described with a comprehensive set of attributes in clear, concise, and consistently formatted materials such as those used in this study.

APPENDIX

The following attribute definitions are abbreviated versions of those that appear in the "Notes on the Numbers" pamphlet. All of these attributes account for impacts that occur within the county and, if relevant, impacts that occur outside the county as a result of sources within the county.

Human Health and Safety Impacts

1. *Risk of death for the average person*. This attribute is the average annual risk of death for a randomly chosen person in DePaul County (or a larger geographic area, if relevant) as a result of the given hazard. The expected number of deaths per year is simply the risk of death multiplied by the population of the county (or the population of the larger area, if relevant). Values range from 0 to 240 in a million (0 to 21 deaths per year).

2. *Risk of death for the person at highest risk*. This attribute is the average annual mortality risk for the person at highest risk. Risk levels could be heightened both by exposure and susceptibility, but susceptibility information is not readily available. Therefore, we consider the most exposed person. Values range from 0 to 3,000 in a million.

3. *Catastrophic potential, greatest number of deaths in a single event*. Some of the hazards in DePaul County kill only one person at a time. Others can kill a group of people all at once. This attribute represents the greatest number of lives that could plausibly be lost in a single event involving the given hazard. Values range from 0 to 100–300 deaths.

4. *Number of serious injuries and illnesses per year*. Serious injuries and illnesses are those that require hospitalization, regardless of the duration of the condition. Examples include the loss of a limb, blindness, nonfatal cancer, disfiguring burns, permanent damage to organs, meningitis, pneumonia, severe allergy attacks, compound bone fractures, and severe food poisoning. Values range from 0 to 35 cases per year.

5. *Number of minor injuries and illnesses per year*. Minor injuries and illnesses are those that might require medical treatment, but not hospitalization. Examples include the flu, ankle sprains, and simple bone fractures. Values range from 0 to 18,000 cases per year.

6. *Time between exposure and health effects*. For hazards in DePaul County, the time between exposure and the onset of health effects ranges from *immediate* to 30 years.

7. *Scientific understanding and predictability of health and safety effects*. This attribute indicates how well scientists know the relationships between exposure and health effects for the given hazard, and how predictable those effects are. Ratings on these two components are combined to yield attribute levels of *low*, *somewhat low*, *medium*, *somewhat high*, and *high*.

8. *Ability of individual to control one's own exposure to health and safety risks*. Some hazards that people encounter can be avoided partly or entirely by actions that individuals can take on their own. For instance, people not wishing to incur risks from recreational boating can choose not to ride in a boat. Controllability is characterized as *low*, *medium*, or *high*.

Environmental Impacts

9. *Habitat affected*. This attribute indicates the extent of the ecological and aesthetic impacts of an activity or environmental stress in terms of the amount of plant and animal habitat affected. Because the effects of hazards are not confined by political boundaries, the amount of habitat affected by some hazards far exceeds the land area in DePaul County. For example, air pollution from a power plant can travel across and affect many counties and states. Values range from 0 to 310,000 acres (487 square miles).

10. *Number of animals killed or displaced*. Activities and environmental stressors can affect birds, fish, and mammals that live in DePaul County. This attribute reflects how many animals are killed or forced to find new habitat because of the given hazard. Hazards that do not kill or displace animals are scored as affecting *none or almost none*; those that do are scored as affecting *few* or *many*.

11. *Effects on variety of native species*. Native species variety is defined in terms of the number and relative abundance of native plant and animal species found in DePaul County. This attribute reflects the

magnitude of changes in the variety of species that have historically been found in the county. Values of *none or almost none*, *small*, *medium*, or *large* represent hypothetical consensus judgments by ecology experts, based upon species and population counts in the county.

12. *Ecological significance of affected species and habitat*. Ecological significance can be based on the rarity or uniqueness of the species or habitat, the position of the species in the food web, or the characteristics of the habitat that is affected. Values of *none or almost none*, *low*, *medium*, or *high* represent hypothetical consensus judgments by ecology experts regarding the significance of the affected species and habitat.

13. *Impacts on natural processes and cycles*. Many interconnected processes and cycles characterize an ecosystem. Cycles involve the processing of materials and energy in the environment. Population processes (e.g., the food web and natural selection) involve the regulation and evolution of plant and animal species over time. Values of *none or almost none*, *low* (should have been *small*), *medium*, or *large* represent hypothetical consensus judgments by ecology experts regarding the magnitude of effects on these processes and cycles.

14. *Catastrophic potential, magnitude of worst-case environmental effects*. All other attributes reflecting environmental impacts refer to the most likely outcome. This attribute indicates the magnitude of impacts for the worst-case scenario for the given hazard. Catastrophic potential is characterized as *none or almost none*, *small*, *medium*, or *large*, relative to the set of hazards considered.

15. *Changes in landscape appearance*. Scores are based on a hypothetical poll of county residents in which each person was asked to evaluate the effect of the given hazard on landscape appearance. Hypothetical responses on a 5-point scale with $-2 =$ *large negative* and $2 =$ *large positive* were averaged, and the possible minimum and maximum values were rescaled to –10 and 10, respectively. The words *large negative*, *negative*, *small negative*, and *little or no change* are associated with the resulting values, as appropriate (no hazards scored higher than $+1$).

16. *Effects on noise, smell, taste, and visibility*. Scores are based on a hypothetical poll of county residents in which each person was asked to evaluate the effects of the given hazard on these dimensions. Hypothetical responses were averaged across questions, and rescaled using the same procedure as that used for changes in landscape appearance.

17. *Time between exposure and environmental effects*. For hazards in DePaul County, the time between exposure and the onset of environmental effects ranges from *immediate* to 50 years.

18. *Duration of environmental effects*. This attribute indicates the hypothetical consensus judgment of ecology experts regarding the length of time necessary for the environment to fully recover from ecological and aesthetic damage resulting from the given hazard. This judgment assumes that the activity or stress is stopped, but that no clean-up or restoration activities take place. For example, in the case of invasive species, the judgment assumes that no further invasive species are introduced, but that previously introduced species are not removed. Values for this attribute range from 0 to *unlimited*.

19. *Scientific understanding and predictability of environmental effects*. This attribute indicates how well scientists know the relationships between exposure and environmental effects for the given hazard, and how predictable those effects are. Ratings on these two components are combined to yield attribute levels of *low*, *somewhat low*, *medium*, *somewhat high*, and *high*.

20. *Negative effects on the environment's capacity to provide goods and services to people*. This attribute reflects the extent to which the given hazard decreases the availability of goods and services provided by the environment (e.g., agricultural products, recreational opportunities, or scientific knowledge). Gains in goods and services associated with the hazard are not considered. For example, this assessment would include the decline in fishing opportunities resulting from the ecological effects of recreational boating, but would not include the positive contribution to recreation provided by boating. Negative effects are characterized as *none or almost none*, *small*, *medium*, or *large*.

ACKNOWLEDGMENTS

This research was funded by the U.S. EPA (grant R8279200-1-0) and the National Science Foundation (grant SES-9975200). The authors would also like to thank Baruch Fischhoff, Claire Palmgren, Colleen Vrbin, and Denise Murrin-Macey for assisting with data collection and contributing valuable insight and critique during this project.

REFERENCES

Barron, F. H., & Barrett, B. E. (1996). Decision quality using ranked attribute weights. *Management Science*, *42*, 1515–1523.

Brown, R. D. (1996). Epilogue. Looking backward: The idea of an informed citizenry at the end of the twentieth century. In *The Strength of a People: The Idea of an Informed Citizenry in America, 1650–1870* (pp. 196–207). Chapel Hill: University of North Carolina Press.

DeKay, M. L., Florig, H. K., Fischbeck, P. S., Morgan, M. G., Morgan, K. M., Fischhoff, B., & Jenni, K. E. (2001). The use of public risk ranking in regulatory development. In P. S. Fischbeck & R. S. Farrow (Eds.), *Improving Regulation: Cases in Environment, Health, and Safety* (pp. 208–230). Washington, DC: Resources for the Future.

Fischhoff, B., & Willis, H. (2001). Adolescent vulnerability: Measurement and priority setting. In B. Fischhoff, E. O. Nightingale, & J. G. Iannotta (Eds.), *Adolescent Risk and Vulnerability: Concepts and Measurement* (pp. 109–144). Washington, DC: National Academy Press.

Florig, H. K., Morgan, M. G., Morgan, K. M., Jenni, K. E., Fischhoff, B., Fischbeck, P. S., & DeKay, M. L. (2001). A deliberative method for ranking risks (I): Overview and test bed development. *Risk Analysis*, *21*, 913–921.

Jenni, K. E. (1997). *Attributes for Risk Evaluation*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Jia, J., Fischer, G. W., & Dyer, J. S. (1998). Attribute weighting methods and decision quality in the presence of response error: A simulation study. *Journal of Behavioral Decision Making*, *11*, 85–105.

Jones, K. (1997). *A retrospective on ten years of comparative risk*. Montpelier, VT: Green Mountain Institute for Environmental Democracy. Available at http://www.gmied.org/PUBS/papers/crdocs/crdocs.html (January 23, 2003).

Jones, K., & Klein, H. (1999). Lessons from 12 years of comparative risk projects. *Annual Review of Public Health*, *20*, 159–172.

Kendall, M., & Gibbons, J. D. (1990). *Rank Correlation Methods*, 5th ed. New York: Oxford University Press.

Konisky, D. M. (2001). Over a decade of comparative risk analysis: A review of the human health rankings. *Risk: Health, Safety & Environment*, *12*, 41–58.

Lazo, J. K., Kinnell, J. C., & Fisher, A. (2000). Expert and layperson perceptions of ecosystem risk. *Risk Analysis*, *20*, 179–193.

McDaniels, T. L., Axelrod, L. J., Cavanagh, N. S., & Slovic, P. (1997). Perception of ecological risk to water environments. *Risk Analysis*, *17*, 341–352.

McDaniels, T., Axelrod, L. J., & Slovic, P. (1995). Characterizing perception of ecological risk. *Risk Analysis*, *15*, 575–588.

McDaniels, T., Axelrod, L. J., & Slovic, P. (1996). Perceived ecological risks of global change. *Global Environmental Change*, *6*, 159–171.

Minard, R. A., Jr. (1996). CRA and the states: History, politics, and results. In J. C. Davies (Ed.), *Comparing Environmental Risks: Tools for Setting Government Priorities* (pp. 23–61). Washington, DC: Resources for the Future.

Morgan, K. M. (1999). *Development and Evaluation of a Method for Risk Ranking*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh, PA.

Morgan, K. M., DeKay, M. L., Fischbeck, P. S., Morgan, M. G., Fischhoff, B., & Florig, H. K. (2001). A deliberative method for ranking risks (II): Evaluation of validity and agreement among risk managers. *Risk Analysis*, *21*, 923–937.

Morgan, K. M., Fischbeck, P. S., & DeKay, M. L. (1999). *Assessing a Multiattribute Model for Ranking Risks*. Paper presented at the annual meeting of the Institute for Operations Research and the Management Sciences, Philadelphia, PA.

Morgan, M. G., Fischhoff, B., Bostrom, A., & Atman, C. J. (2002). *Risk Communication: The Mental Models Approach*. New York: Cambridge University Press.

Morgan, M. G., Florig, H. K., DeKay, M. L., & Fischbeck, P. (2000). Categorizing risks for risk ranking. *Risk Analysis*, *20*, 49–58.

Morgenstern, R., & Sessions, S. (1988). Weighing environmental risks: EPA's unfinished business. *Environment*, *30*(6), 14–17, 34–39.

Morrone, M. (1995). *Ohio State of the Environment Report*. Columbus: Ohio Environmental Protection Agency, the Ohio Comparative Risk Project.

Slovic, P. (1992). Perception of risk: Reflections on the psychometric paradigm. In S. Krimsky & D. Golding (Eds.), *Social Theories of Risk* (pp. 117–152). New York: Praeger.

U.S. EPA. (1987). *Unfinished Business: A Comparative Assessment of Environmental Problems*. Washington, DC: EPA Office of Policy Analysis.

U.S. EPA. (1990). *Reducing Risk: Setting Priorities and Strategies for Environmental Protection* (SAB-EC-90-021). Washington, DC: EPA Science Advisory Board.

U.S. EPA. (1993). *Guidebook to Comparing Risks and Setting Environmental Priorities* (EPA 230-B-93-003). Washington, DC: EPA Office of Policy, Planning, and Evaluation.

U.S. EPA. (2000). *Toward Integrated Environmental Decision Making*. Washington, DC: EPA Science Advisory Board, Integrated Risk Project Steering Committee.

Willis, H. H., & DeKay, M. L. (2003). *The Role of Group Membership and Worldview in Ecological Risk Perceptions*. Manuscript in preparation, Carnegie Mellon University, Pittsburgh, PA.

Willis, H. H., DeKay, M. L., Fischhoff, B., & Morgan, M. G. (2003). *Aggregate- and Individual-Level Analysis of Ecological Risk Perception*. Manuscript in preparation, Carnegie Mellon University, Pittsburgh, PA.