Risk Analysis, Vol. 21, No. 5, 2001

A Deliberative Method for Ranking Risks (II): Evaluation of Validity and Agreement among Risk Managers

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

A deliberative method for ranking risks was evaluated in a study involving 218 risk managers. Both holistic and multiattribute procedures were used to assess individual and group rankings of health and safety risks facing students at a fictitious middle school. Consistency between the rankings that emerged from these two procedures was reasonably high for individuals and for groups, suggesting that these procedures capture an underlying construct of riskiness. Participants reported high levels of satisfaction with their groups' decision-making processes and the resulting rankings, and these reports were corroborated by regression analyses. Risk rankings were similar across individuals and groups, even though individuals and groups did not always agree on the relative importance of risk attributes. Lower consistency between the risk rankings from the holistic and multiattribute procedures and lower agreement among individuals and groups regarding these rankings were observed for a set of high-variance risks. Nonetheless, the generally high levels of consistency, satisfaction, and agreement suggest that this deliberative method is capable of producing risk rankings that can serve as informative inputs to public risk-management decision making.

KEY WORDS: Risk ranking; comparative risk analysis; risk perception; risk attributes; multiattribute utility; policy capturing; school risks

1. INTRODUCTION

In the first article of this pair, Florig *et al.*[(1)] describe a deliberative risk-ranking method that involves five interdependent steps (see Fig. 1 of that article). Risk experts (with input from community representatives) define and categorize the risks to be ranked (Step A), identify the risk attributes that should be considered (Step B), and characterize the risks in a set of standardized *risk summary sheets* (Step C). Jurylike groups of laypeople or others then rank the risks using these sheets (Step D). Finally, the investigators who conduct the risk-ranking exercises describe the deliberations and the resulting rankings in policy-relevant terms (Step E). Florig *et al.*[(1)] also report on the development of an experimental test bed (the fictitious Centerville Middle School, or CMS), with particular attention paid to the preparation of risk summary sheets (Steps A–C). This article describes procedures for ranking the risks (Step D) and demonstrates these procedures using risk managers as participants.

1.1. Multiple Rankings and Multiple Procedures

In Step D, participants produce both individual and group rankings. Initial individual rankings are elicited before the group meetings, to help partici-

[1] Research Triangle Institute, Washington, DC.
[2] Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA.
*Address correspondence to Michael L. DeKay, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213-3890; dekay@andrew.cmu.edu.

0272-4332/01/1000-0923$16.00/1 © 2001 Society for Risk Analysis

pants articulate their own values. The groups provide an airing of different views to help participants evaluate and refine their opinions. Final individual rankings are collected after the group rankings to assess the effect of group discussion and the extent to which individuals dissent from their groups' rankings.

Risk judgments can be affected by irrelevant factors, such as the reference point used to describe a risk[2] and the type of communication materials used to present the information,[3–5] particularly when the concepts are unfamiliar and people need help articulating their own values.[6] In addition to providing concise, nontechnical, and consistent information about each risk, our method addresses these concerns by providing participants the opportunity to express their judgments in two different ways. Specifically, participants complete both a *holistic* ranking procedure (in which they rank the risks directly) and a *multiattribute* ranking procedure (in which they indicate the relative importance of different risk attributes, and the investigators use this information to construct implied rankings of the risks).[7,8]

1.2. Validity and Agreement

To make responsible use of risk rankings, policy makers need to be sure that the individuals and groups involved have a clear definition of risk. Because our holistic and multiattribute procedures are both designed to measure the same thing (relative levels of concern about specific hazards), the correlation between the rankings produced with the two procedures is a measure of internal consistency or *convergent validity.* If, as intended, group discussion provides participants with additional knowledge and the opportunity to reconsider their judgments, such consistency should increase over the course of the ranking exercise.

Other things being equal, the resulting rankings are likely to be more compelling to policy makers when participants are satisfied with the process and willing to stand behind their groups' rankings as representative of their concerns. Because people often evaluate the fairness of outcomes by assessing the fairness of the process that led to them,[9] participants' satisfaction with both the process and the final results are assessed, as measures of *face validity.*

Similarity among rankings produced by different individuals and groups is not, in itself, an objective of risk ranking, because participants may have legitimate reasons for disagreeing about the relative riskiness of hazards. Although good risk-communication materials should increase agreement among participants by providing missing knowledge and dispelling misconceptions,[10–12] residual disagreements may depend both on the risks in question and on the values and experiences of participants. Policy makers would be well served by knowing how much agreement exists and how the level of agreement changes over the course of the ranking exercise. Our procedures are designed to provide this information.

In this article, we evaluate our risk-ranking method in terms of (a) the consistency between holistic and multiattribute risk rankings, at both the individual and group levels; (b) participants' satisfaction with the ranking process and the resulting rankings; and (c) the agreement among individuals and among groups. Although rankings of risks and attributes are also reported, our goal is to evaluate the method, not to put forth specific rankings for policy purposes.

1.3. A Limited Study of Risk Managers

The study reported here differed from the full risk-ranking method described in the companion article[1] in a number of ways. First, the participants were risk managers rather than laypeople. The performance of the method with lay participants has been examined in other studies.[13–15] Second, the brevity and nature of the risk-ranking sessions precluded the use of the full set of 22 risks, the juxtaposition and integration of holistic and multiattribute risk rankings, and the collection of detailed data on groups' decision-making processes (e.g., recording or coding the actual discussions). These aspects of the method have also been addressed elsewhere.[13–15] Finally, the procedures employed in the various sessions differed slightly, because our ideas about the method and the proper metrics for assessing it evolved over the course of this research. The similarities among sessions were much greater than the differences, however, and the resulting data provide a rich assessment of several key aspects of our risk-ranking method.

2. METHOD

2.1. Participants

Two hundred eighteen persons who were enrolled in a short course entitled "Analyzing Risk: Science, Assessment, and Management" at the Harvard School of Public Health participated in this study. Data were collected at five different times: fall 1997 ($n = 48$), spring 1998 ($n = 47$), fall 1998 ($n = 53$), spring 1999 ($n = 26$), and fall 1999 ($n = 44$), referred to as Sessions 1 through 5, respectively. Participants were predominantly risk managers from government, industry, and consulting firms in the United States.

2.2. Materials

Materials for this study included a description of CMS and risk summary sheets for 21 of the 22 risks listed in the companion article[(1)] (when we began this study, the risk summary sheet for "building collapse" had not yet been written). More information on the materials is provided in that article and at http://www.epp.cmu.edu/research/EPP_risk.html.

Given the limited time available, each participant received one of six different subsets of the 21 risks. Risk Sets A through C (11 risks each) were used in Sessions 1 and 5, Sets D and E (9 risks each) were used in Sessions 2 and 3, and set F (9 risks) was used in Session 4. Risks for Sets A, B, C, and F were chosen to span the range of risk characteristics, with some overlap between sets. In pilot studies, greater variation was observed in the ranks assigned to some risks, such as "electric and magnetic fields from electric power." Set D was composed of these *high-variance* risks. We hypothesized that groups considering only high-variance risks would be less likely to reach agreement, and that members of these groups would find the tasks more difficult and would be less satisfied with their groups' decision-making processes and rankings. In contrast, Set E was composed of *low-variance* risks, such as "school bus accidents," that had been ranked rather consistently in pilot studies. Further details on the risk sets appear below.

In Sessions 2 through 5, the risk summary sheets contained attribute tables like the one shown in Fig. 2 of the companion article.[(1)] In Session 1, the risk summary sheets contained an earlier version of the attribute table that included "number of injuries, illnesses," "number of disability days," and "number of hospital days" rather than the four measures of morbidity adopted later. In addition, the attribute table used in Session 1 did not list "combined uncertainty in death, illness, injury" as a separate attribute, although it did include the high and low estimates for mortality and morbidity, on which this attribute was based.

2.3. Procedures

2.3.1. Participant Tasks

A few days before attending the risk-ranking session, each participant received the packet of materials, along with instructions to complete several tasks individually before attending the session. After reading the description of CMS and the risk summary sheets, participants ranked the risks, with 1 being the risk of greatest concern (the *initial individual holistic* rankings; Step D1). Participants then ranked the risk attributes according to their relative importance, with 1 being the most important attribute, and with explicit instructions not to rank those attributes they had not used when ranking the risks. These retrospective reports of attribute importance were used later to construct implied multiattribute rankings of the risks, using the procedure described in Section 2.3.2 (the *initial individual multiattribute* risk rankings; Step D2).

Immediately prior to the ranking session, participants attended a 50-min lecture on the psychology of risk perception and the multiattribute nature of risk. Participants who had individually considered the same risk sets were assigned to small groups ($n = 3–7$). Each group produced a single ranking of the risks to represent its concerns (the *group holistic* rankings; Step D3). Groups then provided retrospective rankings of attribute importance, from which multiattribute rankings of the risks were constructed later (the *group multiattribute* risk rankings; Step D4). Groups in Sessions 1 and 4 were not asked to provide attribute rankings, so group multiattribute risk rankings could not be constructed for them. Over the five sessions, Risk Sets A through C were considered by 6 groups each, Sets D and E were considered by 10 groups each, and Set F was considered by 5 groups.

Following the group exercise, individuals again ranked the risks (the *final individual holistic* rankings; Step D5) and attributes (for the *final individual multiattribute* risk rankings; Step D6). Participants in Session 4 did not provide final individual rankings of the risks or attributes. Instead, each individual identified "the three risks for which the group experienced the most difficulty in agreeing on the final ranks." For each of these three risks, participants were instructed to check all of the following statements that applied:

- Some members of the group disagreed about the relative importance of the attributes for this risk.
- Some of the group felt that this risk has important features not captured in the risk summary sheet.
- Some of the group had specific knowledge and experience with this risk, but others did not.
- Some of the group did not believe some of the information provided in the risk summary sheet.

- Some of the group were stubborn and unwilling to listen to reasonable arguments.
- Other reasons (please list):

Participants in all sessions evaluated three statements regarding the group decision-making process, using a scale from 1 (*strongly disagree*) to 7 (*strongly agree*):

- The group was willing to consider and discuss different points of view and encouraged each member to express his or her opinion.
- When group members disagreed about the way in which risks should be ranked, this disagreement was resolved primarily by voting or averaging.
- When group members disagreed about the way in which risks should be ranked, this disagreement was resolved primarily by persuasion and opinion change.

In addition, participants in all sessions answered two questions regarding their satisfaction with the group process and output, using a scale from 1 (*not at all satisfied*) to 7 (*very satisfied*):

- How satisfied are you with your group's decision-making process?
- How satisfied are you with your group's final ranking of the risks?

Participants in Sessions 3 and 5 also answered the following question, using the same scale:

- How satisfied are you with your group's final ranking of the attributes?

Each ranking session took about an hour, excluding the initial lecture and final discussion.

2.3.2. *Calculating Implied Multiattribute Risk Rankings*

For each participant and group, the attribute rankings from Steps D2, D4, and D6 were used to construct concern scores for each risk as follows:

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

where j is a risk, i is an attribute, n is the number of attributes, w_i is the weight for attribute i, v_i is the value function for attribute i, and x_{ij} is risk j's level on attribute i. The weight for each attribute was assumed to equal the reciprocal of the attribute rank, normalized so that the weights (without signs) summed to 1.[16] The value function for each attribute was assumed to equal the ranking of the levels of that attribute, normalized to range from 0 to 1 (e.g., for an attribute with levels of *low*, *medium*, and *high*, the values were 0, 0.5, and 1, respectively). After the concern scores were computed for each risk, they were ranked so that 1 represented the risk of greatest concern, as for the holistic rankings. When consistency with holistic rankings is used as a measure of convergent validity, the above assumptions imply multiattribute rankings for this set of risks that are as valid as those implied by more time-consuming elicitation procedures.[13,14]

In a simplification of procedures used in the full method,[1] participants were not asked to specify whether high or low levels of the attributes were associated with greater concern (i.e., the signs of the ws in the above expression). For the morbidity and mortality attributes, the correct signs are obvious. However, the correct signs may not be obvious for the four qualitative attributes at the bottom of the risk summary sheets (see Fig. 2 in the companion article[1]). "Quality of scientific understanding"; "combined uncertainty in death, illness, injury"; and "time between exposure and health effects" were intended to represent a factor that has been labeled *unknown risk*. The relation between this factor and perceived risk is often weak[17,18] and not always of the same sign.[17] "Ability of student/parent to control exposure" (along with the quantitative attribute "greatest number of deaths in a single episode") was intended to represent a factor labeled *dread risk*. This factor is positively associated with perceived risk,[17,18] although this relationship does not always hold for specific controllability attributes.[17] For school risks, Jenni[19] reported that scientific knowledge and personal control are negatively correlated with risk rankings, whereas immediacy is essentially uncorrelated with risk rankings. Additional insight was obtained from exercises in which lay groups actively debated the signs for the attributes in the risk summary sheets.[15] In this study, we assumed that greater concern was associated with immediate consequences, lower scientific understanding, greater uncertainty, and lower ability of the student or parent to control exposure, but we also assessed multiattribute rankings based on different assumptions.

3. RESULTS

Participants reported spending an average of 73 min on their initial individual rankings of the risks and attributes before attending the ranking sessions. Because group members' responses could not be considered to be independent after group discussion, the sta-

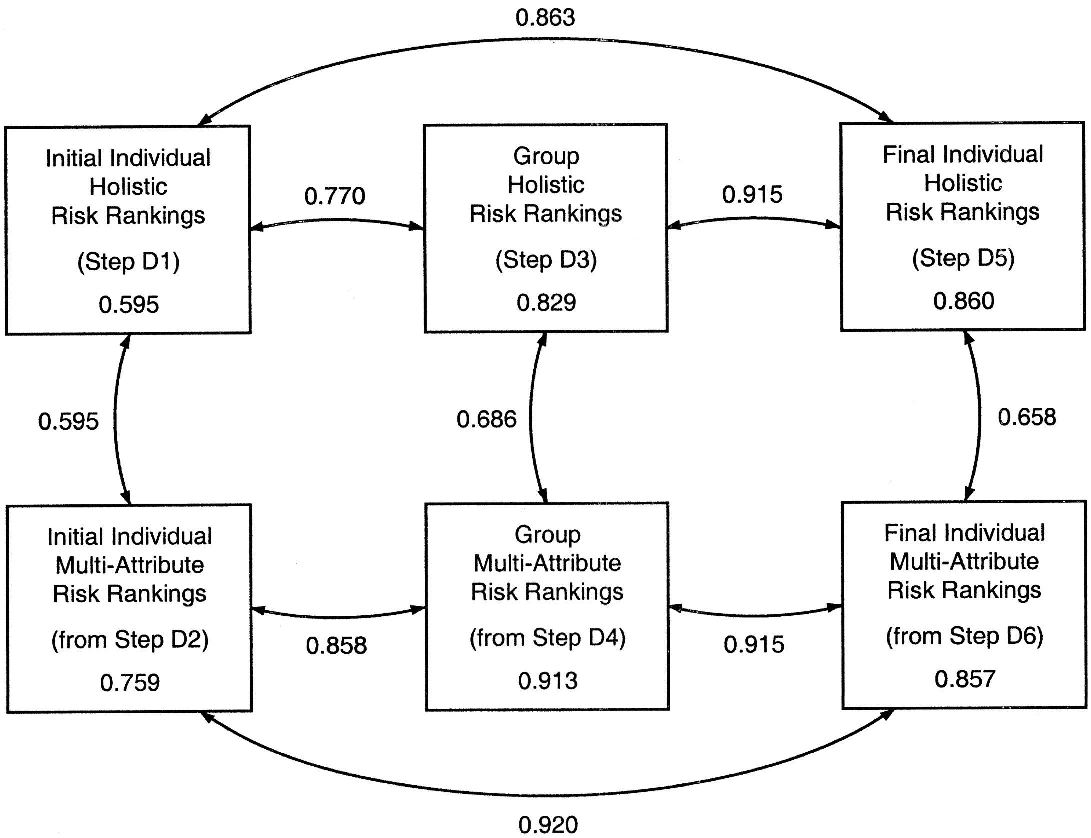

Fig. 1. Results from correlational analyses. Numbers next to arrows are mean Spearman correlations between risk rankings. When individual participants' rankings were involved, correlations were computed at the individual level and averaged within groups; the means reported here are the means of those group means. Numbers within boxes are mean pairwise Spearman correlations among individual participants' rankings (Steps D1, D2, D5, and D6) or among groups' rankings (Steps D3 and D4; see Sections 3.1 to 3.3 for details). Some tasks were not completed in some sessions (see Section 2.3.1 for details). Results of some analyses for individual risk sets are shown in Tables I and II.

tistics reported below were computed at the individual level, averaged within groups, and then analyzed at the group level. For ease of comparison, these procedures were used regardless of whether the individual data were collected before or after group discussion.

3.1. Consistency between Holistic and Multiattribute Risk Rankings

Mean Spearman correlations were used to assess the consistency between holistic and multiattribute risk rankings within individuals (both before and after group discussion) and within groups.

3.1.1. Within-Individual Consistency

Across all groups, the mean Spearman correlation between participants' initial individual holistic rankings (Step D1) and initial individual multiattribute risk rankings (Step D2) was 0.595. The mean Spearman correlation between participants' final individual holistic rankings (Step D5) and final individual multiattribute risk rankings (Step D6) was 0.658 (see Fig. 1).[3] Both mean correlations were significantly greater than zero, $t(42) = 17.05$, $p \leq 0.0001$, and $t(36) = 16.84$, $p \leq 0.0001$, respectively. Analyses of variance (ANOVAs) indicated that the initial correlations were marginally dependent on the risk set considered, $F(5, 37) = 2.27$, $p = 0.068$, whereas the final correlations were significantly dependent on the risk set considered, $F(4, 32) = 3.75$, $p = 0.013$ (see Table I). After group discussion, participants who considered the high-variance Risk Set D demonstrated significantly lower consistency between the two rankings than did participants who

[3] Because there were often several perfect correlations at the individual or group level, Fisher's Z transformation was not used in the analyses reported in this article.

Table I. Consistency (Mean Spearman Correlations between Holistic and Multiattribute Risk Rankings) for Each Risk Set

Rankings	Risk Set A	B	C	D	E	F[a]	All
Individual							
Initial (Steps D1 and D2)	0.760	0.648	0.588	0.518	0.654	0.376	0.595
Final (Steps D5 and D6)	0.781	0.684	0.767	0.447	0.722	—	0.658
Final minus initial[b]	0.041	0.051	0.075	−0.047	0.187	—	0.060
Group (Steps D3 and D4)	0.867	0.909	0.679	0.435	0.830	—	0.686

[a] Participants who considered Risk Set F did not provide final individual rankings or group attribute rankings.
[b] These means of differences may be slightly different from the differences between the above means because some participants did not provide the necessary rankings.

considered other risk sets, $F(1, 32) = 14.24$, $p = 0.0007$.[4]

The mean correlation between individuals' holistic and multiattribute risk rankings was greater by 0.060 after group discussion, $t(36) = 1.93$, $p = 0.062$. The increase for the high-variance risk set D was significantly less than for the other risk sets, $F(1, 32) = 4.19$, $p = 0.049$, whereas the increase for the low-variance Risk Set E was marginally greater than for Risk Sets A through C, $F(1, 32) = 3.22$, $p = 0.082$ (see Table I). When Risk Set D was excluded, the increase in consistency for the other risk sets was 0.100, $t(26) = 2.75$, $p = 0.011$.

The generally high correlations between holistic and multiattribute risk rankings suggest that the two procedures yielded valid measures of individuals' beliefs about the relative riskiness of these hazards. The fact that consistency did not increase for the high-variance Risk Set D suggests that the risks in this set presented particular difficulties, perhaps because they have features that were not captured by our set of attributes.

3.1.2. Within-Group Consistency

The mean Spearman correlation between groups' holistic rankings (Step D3) and groups' multiattribute risk rankings (Step D4) was 0.686 (see Fig. 1). The mean correlation for the high-variance Risk Set D was significantly lower than for the other risk sets, $F(1, 21) = 17.04$, $p = 0.0005$ (see Table I).

3.1.3. Making Different Assumptions about the Signs for the Four Qualitative Attributes

As noted above, the construction of multiattribute risk rankings required assumptions regarding the signs for the four qualitative attributes. As an initial test of the validity of these assumptions, the correlations between holistic and multiattribute risk rankings were recalculated after changing the sign for each qualitative attribute in turn. If our initial assumptions were correct, then reversing any of these signs should have led to lower consistency between holistic and multiattribute risk rankings. When the sign for "time between exposure and health effects" was reversed, so that risks with delayed effects were assumed to be associated with greater concern, the initial individual, final individual, and group correlations decreased from 0.595 to 0.576, from 0.658 to 0.614, and from 0.686 to 0.636, respectively. However, when the sign for any of the other three attributes was reversed, so that greater scientific understanding, lower uncertainty, or greater ability of the student or parent to control exposure was assumed to be associated with greater concern, these correlations all increased. When the signs for these three attributes were reversed simultaneously, the initial individual, final individual, and group correlations increased to 0.721, 0.796, and 0.795, respectively. Although participants did not weight these three qualitative attributes as we had anticipated, reversing their signs had almost no effect on the other findings reported in

[4] Orthogonal contrast codes were used to test our hypotheses regarding the difficulty of ranking the risks in Set D and the ease of ranking the risks in Set E.[20] The first code compared Set D with Sets A, B, C, E, and F (when data for Set F were present), and the second code compared Set E with Sets A, B, C, and F (when data for Set F were present). Complete sets of contrasts codes were used, but we did not test the significance of any *post hoc* comparisons, such as whether the correlations between the initial individual holistic and multiattribute risk rankings were lower for Set F than for the other risk sets.

this article. More detailed multiple regression analyses of attribute use appear in Section 3.5.

3.2. Participants' Evaluations of Group Processes and Rankings

3.2.1. Group Decision-Making Processes

Relative to the scale midpoint (4), participants agreed that their groups considered and discussed different viewpoints and encouraged each member to express his or her opinion, $M = 6.12$, $t(42) = 19.88$, $p \leq 0.0001$; that disagreements about ranking the risks were resolved primarily by voting or averaging, $M = 5.05$, $t(41) = 5.54$, $p \leq 0.0001$; and that disagreements about ranking the risks were resolved primarily by persuasion and opinion change, $M = 4.93$, $t(42) = 5.49$, $p \leq 0.0001$. Although the latter two results appear contradictory, participants may have focused on the fact that disagreements were resolved, rather than on the means of resolution. These results were not significantly different for the six risk sets.

The only significant correlation among these three measures was between responses to the first statement (consideration of different viewpoints) and the third statement (resolution of disagreements by persuasion and opinion change), $r(41) = 0.462$, $p = 0.002$. The fact that only one of the three correlations among these measures was significant indicates that participants' endorsement of the decision-process statements did not reflect a halo effect.

3.2.2. Self-Reported Satisfaction

Relative to the scale midpoint (4), participants indicated that they were satisfied with their groups' decision-making processes, $M = 5.76$, $t(42) = 12.63$, $p \leq 0.0001$; their groups' final rankings of the risks, $M = 5.80$, $t(42) = 17.90$, $p \leq 0.0001$; and their groups' final rankings of the attributes, $M = 5.41$, $t(20) = 6.63$, $p \leq 0.0001$. These results were not significantly different for the six risk sets. In the only planned comparison that approached significance, participants who considered the high-variance Risk Set D were somewhat more satisfied with their groups' rankings of the attributes than were participants who considered other risk sets, $M = 6.07$, $F(1, 15) = 3.26$, $p = 0.091$. This occurred despite the relatively low consistency between holistic and multiattribute risk rankings for Risk Set D (see Table I).

Responses to the three satisfaction questions were highly correlated, all rs ≥ 0.698, all ps ≤ 0.0004. In separate multiple regressions, each of the three satisfaction questions were regressed onto the three decision-process statements. Results indicated that groups able to resolve their disputes were more satisfied with their decision-making processes and the resulting rankings, adjusted R^2s = 0.356, 0.274, and 0.516 for the three satisfaction questions, respectively.

3.2.3. Regression-Based Measures of Satisfaction

Participants' satisfaction with their groups' rankings was also assessed using within-participant regressions. Participants' final holistic rankings (Step D5) were regressed onto their group's holistic ranking (Step D3) and their own initial individual holistic ranking (Step D1). The mean regression coefficient for the group ranking, $M(b_{D3}) = 0.654$, was significantly greater than that for the initial individual ranking, $M(b_{D1}) = 0.294$, $t(37) = 3.02$, $p = 0.005$, indicating that participants were satisfied with their groups' holistic rankings.[5] Across groups, 21.2% of members adopted their groups' holistic rankings as their final individual holistic rankings, whereas only 12.1% reverted to their initial individual holistic rankings (for an additional 1.7% of members, the three holistic rankings in Steps D1, D3, and D5 were identical).

A parallel analysis was conducted on the multiattribute risk rankings, which were constructed using our original assumptions regarding the signs for the qualitative attributes. The mean coefficient for the group ranking, $M(b_{D4}) = 0.577$, was not significantly greater than that for the initial individual ranking, $M(b_{D2}) = 0.426$, although it was significantly greater than zero, $t(24) = 9.75$, $p \leq 0.0001$, indicating that participants found the group discussion of attributes useful (see footnote 5). Across groups, members' final individual multiattribute risk rankings matched their groups' multiattribute risk rankings 22.7% of the time, and matched their initial individual multiattribute risk rankings 15.4% of the time (for an additional 6.3% of members, the three multiattribute risk rankings in Steps D2, D4, and D6 were identical).

In summary, self-reports and regression analyses of holistic and multiattribute risk rankings indicated that participants were very satisfied with their groups' decision-making processes and the resulting rankings.

[5] Occasionally, the group ranking and the initial individual ranking were identical, and the regression failed because of collinearity. When this occurred, both regression coefficients were set to 0.5. Results were similar when the coefficients for the group ranking and the initial individual ranking were set (more conservatively) to 0 and 1, respectively.

Table II. Agreement (Mean Pairwise Spearman Correlations between Holistic Risk Rankings and between Multiattribute Risk Rankings) for Each Risk Set

	Risk Set						
Rankings	A	B	C	D	E	F[a]	All
Holistic							
Individual							
Initial (Step D1)	0.716	0.698	0.583	0.583	0.600	0.356	0.595
Final (Step D5)	0.843	0.920	0.944	0.785	0.878	—	0.860
Final minus initial[b]	0.126	0.235	0.311	0.202	0.199	—	0.209
Group (Step D3)	0.899	0.874	0.798	0.730	0.855	0.789	0.829
Multiattribute							
Individual							
Initial (Step D2)	0.888	0.795	0.711	0.735	0.809	0.564	0.759
Final (Step D6)	0.933	0.769	0.874	0.802	0.930	—	0.857
Final minus initial[b]	0.056	0.005	0.016	0.067	0.019	—	0.037
Group (Step D4)	0.924	0.991	0.958	0.740	0.950	—	0.913

[a] Participants who considered Risk Set F did not provide final individual rankings or group attribute rankings.

[b] These means of differences may be slightly different from the differences between the above means because some participants did not provide the necessary rankings.

3.3. Agreement among Individuals and among Groups

As mentioned earlier, measures of similarity among rankings produced by different individuals and groups are likely to be very useful to policy makers, regardless of whether those measures indicate agreement or disagreement. Mean pairwise Spearman correlations were used to assess agreement among individuals within groups (both before and after group discussion) and agreement among groups that ranked the same set of risks.

3.3.1. Agreement among Individuals within Groups

For each group, Spearman correlations were computed between the initial individual holistic rankings (Step D1) for all possible pairs of individuals, and these correlations were averaged to obtain a measure of prediscussion agreement. These mean pairwise correlations were significantly greater than zero for 41 of 43 groups,[6] with an overall mean of 0.595 (see Fig. 1). They varied significantly as a function of the risk set considered, $F(5, 37) = 2.68, p = 0.036$ (see Table II).

The mean pairwise Spearman correlations between final individual holistic rankings (Step D5) were significantly greater than zero for 31 of 34 groups, with an overall mean of 0.860 (see Fig. 1). Compared with members of groups that considered other risk sets, members of groups that considered the high-variance Risk Set D demonstrated significantly lower agreement among their final holistic rankings, $F(1, 29) = 6.26, p = 0.018$ (see Table II). Agreement among individuals' holistic rankings was significantly higher after group discussion than before, $t(33) = 9.03, p \leq 0.0001$, and this increase did not vary significantly as a function of risk set (see Table II).

For individuals' multiattribute risk rankings, the mean pairwise Spearman correlations were significantly greater than zero for 40 of 43 groups prior to group discussion (Step D2), $M(r_s) = 0.759$, and for 30 of 32 groups after group discussion (Step D6), $M(r_s) = 0.857$ (see Fig. 1). Agreement was not significantly higher after group discussion than before. None of the results for agreement among individuals' multiattribute risk rankings varied significantly as a function of risk set (see Table II).

To this point, we have focused on holistic and multiattribute risk rankings. However, it is possible to conduct a parallel analysis of the attribute rankings used to construct the multiattribute risk rankings. Agreement on attribute rankings was significantly higher after group discussion (Step D6), $M(r_s) =$

[6] Mean pairwise Spearman correlations may be tested for significance by first transforming them to Kendall's coefficient of concordance, $W = [r_s(m-1)+1]/m$, where r_s is the mean pairwise correlation and m is the number of rankings.[21] The result of a second transformation, $Wm(n-1) = [r_s(m-1)+1](n-1)$, where n is the number of items being ranked, is distrubted as χ^2 with $n-1$ degrees of freedom.[21] In most of the cases in which agreement was not significantly greater than zero, some participants did not provide the required rankings, so m was low and the value of r_s required to reach the $p = 0.05$ significance level as high.

0.520, than before (Step D2), $M(r_s) = 0.308$, $t(32) = 3.92$, $p = 0.0004$. These relatively low correlations among attribute rankings led to much higher correlations among multiattribute risk rankings because our weighting scheme emphasized the most important attributes and because the attribute values were correlated across risks.

3.3.2. *Agreement among Groups*

Analogous procedures were used to assess agreement among groups that considered the same set of risks. For holistic rankings (Step D3), intergroup agreement was significantly greater than zero for all 6 risk sets, $M(r_s) = 0.829$ (see Fig. 1). For multiattribute risk rankings (Step D4), intergroup agreement was significantly greater than zero for Risk Sets A through E, $M(r_s) = 0.913$ (see Fig. 1). For both types of risk rankings, agreement was lowest for the high-variance Risk Set D (see Table II). Representative group holistic rankings for the six risk sets, created by averaging group rankings within risk sets and reranking the results, are shown in Table III. For reference, the risks are sorted by the expected number of student deaths per year.

The mean pairwise Spearman correlation between groups' attribute rankings was 0.354. Values for Risk Sets A through E appear in Table IV, along with representative group attribute rankings for these risk sets. Significant intergroup agreement was observed only for the high-variance and low-variance Risk Sets D and E, in part because group attribute rankings were not collected for Risk Sets A through C in Session 1 (see footnote 6). Again, these relatively low correlations among attribute rankings led to much higher correlations among multiattribute risk rankings. For Risk Set D, the four qualitative attributes were all ranked as more important than the four illness-and-injury attributes. This pattern was reversed or nearly reversed for the other four risk sets.

Table III. Representative Group Holistic Rankings for Each Risk Set

Risk	Ranking by expected mortality[a]	Risk Set					
		A	B	C	D	E	F
Commuting to school on foot, by bike, or by car	1	1	1	1		1	
Accidental injuries (excluding sports)	2	2				2	1
Less common infectious diseases	4	3	3	3	1		
Self-inflicted injury or harm	4		2				
Team sports	4	4	5	2		3	
Electric and magnetic fields from electric power	6.5		11[b]		7		8
Radon gas	6.5	8			3		
Intentional injury	8		4		2		2
Airplane crashes	9			8	6		
Common infectious diseases	10.5	5	6	4		5	4
School bus accidents	10.5	6				4	3
Allergens in indoor air	12			5		6	5
Bites and stings	13.5			6			
Lightning	13.5	11					
Electric shock	15		7			8	
Asbestos	16			11	8		9
Drowning	17.5	9	8.5	8		9	
Food poisoning	17.5	7				7	6
Fire and explosion	19	10	8.5	10	4		
Hazardous materials transport	20		10		5		7
Lead poisoning	21			8	9		
Mean Spearman correlation with expected mortality[c]		0.801	0.769	0.759	0.508	0.891	0.545

Note: Group holistic rankings were averaged within risk sets and reranked to generate these representative group holistic rankings.
[a] The ranking by expected mortality changed slightly across sessions due to revisions of the risk summary sheets. The ranking shown is for Session 5.
[b] One group that considered Risk Set B did not rank "electric and magnetic fields from electric power." The omitted rank would not affect the results presented unless it were 8 or less. Other groups that considered Risk Set B ranked this risk 8, 10, 10, 11, and 11.
[c] Correlations with expected mortality were calculated separately for each group, using the summary sheet data for the appropriate session, and averaged within risk sets.

Table IV. Representative Group Attribute Rankings for Each Risk Set

Attribute	Risk Set[a] A	B	C	D	E
Student deaths					
Number of deaths per year[b]	1	3	4.5	1	2
Chance in a million of death per year for the average student[b]	6.5	1.5	1	2	1
Chance in a million of death per year for the student at highest risk	9	1.5	2	6	5
Greatest number of deaths in a single episode	4	5	6	4	6
Student illness or injury					
More serious, long-term cases per year	3	6	3	9	3
More serious, short-term cases per year	6.5	8	7	10	7
Less serious, long-term cases per year	2	7	4.5	11	4
Less serious, short-term cases per year	5	11	8	12	9
Other factors					
Time between exposure and health effects	8	9	10	8	12
Quality of scientific understanding	11.5	4	12	3	10
Combined uncertainty in death, illness, and injury	11.5	12	9	5	11
Ability of student/parent to control exposure	10	10	11	7	8
Mean pairwise Spearman correlation between groups' attribute rankings	0.061	0.391	0.355	0.556	0.407

Note: Group attribute rankings were averaged within risk sets and reranked to generate these representative group attribute rankings.
[a] Risk Set F was used only in Session 4, in which group attribute rankings were not collected. Not all groups are represented for Risk Sets A through C, because group attribute rankings were not collected in Session 1.
[b] The first two attributes express the same information in different forms, but groups could still give more attention to one of the two attributes.

3.4. High-Variance and Difficult-to-Rank Risks

As noted in Section 2.2, the high-variance Risk Set D and the low-variance Risk Set E were constructed on the basis of pilot data. These characterizations were checked by computing the standard deviations of groups' ranks for each risk in each risk set (e.g., the standard deviation of groups' ranks for "airplane crashes" in Risk Set D was 1.79). The means of these standard deviations within each risk set were then computed. As expected, the mean standard deviation was greatest for the high-variance Risk Set D, $M(SD) = 1.25$, and lowest for the low-variance Risk Set E, $M(SD) = 0.81$, with the other four risk sets falling in between, $M(SD) = 0.97$, 1.02, 1.14, and 1.07 for Risk Sets A, B, C, and F, respectively. Each of these four sets included some risks from Sets D and E. Within these mixed risk sets, the mean standard deviation for the high-variance risks, $M(SD) = 1.37$, was greater than that for the low-variance risks, $M(SD) = 0.99$. As noted above, participants and groups who considered the high-variance Risk Set D showed lower consistency between their holistic and multiattribute risk rankings and lower agreement regarding these rankings (see Tables I and II).

In Session 4, participants listed the three risks from Set F for which their group experienced the greatest difficulty. Risks were coded as 1 (listed) or 0 (not listed), and these codes were averaged across participants within groups and then across groups to yield a mean difficulty score for each risk. For the nine risks in Set F, the correlation between risks' mean difficulty scores and the standard deviations of groups' ranks for those risks was marginally significant, $r(7) = 0.661, p = 0.052$. Thus, there was some indication that difficulties within groups were predictive of disagreements among groups.

Finally, participants in Session 4 could check one or more reasons for the difficulty experienced with each listed risk. Reasons were tallied for each participant and averaged across participants within groups and then across groups. The reasons most commonly endorsed were that some members disagreed about the relative importance of the attributes for a risk, $M = 1.75$; that some members felt that a risk had important features not captured in the risk summary sheet, $M = 1.20$; and that some members had specific knowledge or experience with a risk, $M = 1.02$.

3.5. Regression Analysis of Attribute Use

As noted in Section 3.1.3, individuals and groups appear to have used three qualitative attributes ("quality of scientific understanding"; "combined uncertainty in death, illness, injury"; and "ability of student or parent to control exposure") in an unexpected manner. These interesting results were explored in greater detail by regressing each group's holistic ranking onto a larger set of attributes. This *policy-capturing* analysis was complicated by the fact that there were as many attributes (10–12, depending on the session) as risks (9–11, depending on the risk set). This issue was addressed by (a) omitting some attributes that were highly correlated with others (e.g., "chance in a million of death per year for the student at highest risk"), (b) creating composite illness-and-injury and knowledge indices from multiple attributes, and (c) analyzing the results of within-group regressions across groups, as described below.

The injury-and-illness index was defined as a weighted sum of the four illness-and-injury attributes (4 × "more serious, long-term cases per year" + 2 × "less serious, long-term cases per year" + 2 × "more serious, short-term cases per year" + 1 × "less serious, short-term cases per year"), normalized to range from 0 to 1. For Session 1, in which the severity of illnesses and injuries was not characterized in this manner, "number of injuries, illnesses" was used. The knowledge index was defined as "quality of scientific understanding" − "combined uncertainty in death, illness, injury," normalized to range from 0 to 1. For Session 1, in which uncertainty was not a separate attribute, "quality of scientific understanding" was used.

For each group, the group's holistic ranking was regressed onto the following predictors:

- Number of deaths per year
- Illness-and-injury index
- Greatest number of deaths in a single episode
- Time between exposure and health effects
- Knowledge index
- Ability of student/parent to control exposure

The following models were analyzed: (a) an expected-mortality-and-morbidity model that included only the first two of these predictors, (b) four three-predictor models that included the first two predictors above plus one of the remaining four, and (c) the full six-predictor model. Results for the three-predictor models were quite similar to those for the full model and are not discussed further. For the full six-predictor model, coefficients were determined for each group and averaged within risk sets and across all groups. Table V reports the mean unstandardized regression coefficients, the significance levels for these means, and the adjusted R^2 values for the two-predictor expected-mortality-and-morbidity model and the full six-predictor model.

Because all predictors ranged from 0 to 1, the mean regression coefficients have very straightforward interpretations. Controlling for the other predictors, an increase from the lowest value to the highest value on a predictor is associated with a change in the predicted risk rank equal to the reported coefficient. For example, the first entry in the rightmost column indicates that the difference between the predicted ranks of risks with the lowest and highest numbers of expected fatalities is 5.17. The ranks were coded so that positive coefficients mean that higher predictor values are associated with greater concern.

Across risk sets, there was uniform agreement

Table V. Mean Unstandardized Regression Coefficients for Group Holistic Rankings

	Risk Set						
Predictor	A	B	C	D	E	F	All
Number of deaths per year	6.65****	7.93	3.98*	4.15**	5.97****	2.51*	5.17****
Illness-and-injury index[a]	2.87**	7.12*	2.77*	−3.94*	1.97*	9.47****	2.31**
Greatest number of deaths in a single episode	2.21*	1.19	−1.66†	2.76**	3.73*	10.05***	2.95****
Time between exposure and health effects	0.14	−10.46*	−0.57	2.38**	1.91	6.26**	0.46
Knowledge index[a]	1.36	−0.31	2.45	3.39†	0.73	0.43	1.54*
Ability of student/parent to control exposure	1.81***	−3.67	2.40*	5.88****	1.14**	2.54***	2.13****
Mean adjusted R^2 for the two-predictor expected-mortality-and-morbidity model	0.792	0.702	0.751	0.308	0.747	0.349	0.599
Mean adjusted R^2 for the full six-predictor model	0.818	0.858	0.758	0.396	0.932	0.743	0.732

Note: For this analysis, ranks were coded so that positive coefficients indicate that higher levels of the predictors are associated with greater concern. Mean coefficients and their significance levels are reported for the full six-predictor model.

[a] See Section 3.5 for the definitions of these indices.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p \leq 0.0001$; † $p < 0.1$.

that larger numbers of expected fatalities were associated with greater concern. There was nearly uniform agreement that larger numbers of expected illnesses and injuries were associated with greater concern, and that greater catastrophic potential (as measured by the greatest number of deaths in a single episode) was associated with greater concern. There was very little agreement regarding the implications of time between exposure and health effects; the mean regression coefficient across groups did not approach significance. There was a weak but relatively consistent indication that greater knowledge (greater scientific understanding and lower uncertainty) was associated with greater concern. Contrary to conventional wisdom, there was nearly uniform agreement that greater ability of the student or parent to control exposure was associated with greater concern. Based on the adjusted R^2 values, both the two-predictor expected-mortality-and-morbidity model and the full six-predictor model performed poorly for the high-variance Risk Set D, compared with the other risk sets, $F(1, 37) = 23.16, p \leq 0.0001$, and $F(1, 36) = 13.38$, $p = 0.0008$, for the two models, respectively.

Spearman correlations between individual attributes and groups' holistic rankings yielded similar results, except that longer times between exposure and health effects were associated with less concern, $M(r_s) = -0.306$, $t(42) = -8.82$, $p \leq 0.0001$. Policy-capturing analyses on individuals' initial holistic rankings also yielded similar results.[13]

4. DISCUSSION

4.1. Consistency, Satisfaction, and Agreement

In Section 1.2, we asserted that the consistency between the rankings produced with the holistic and multiattribute procedures is a measure of the validity of our risk-ranking method. Our results clearly demonstrate such consistency, suggesting that these procedures capture an underlying construct of riskiness. We hypothesized that group discussion would lead to increased consistency between individuals' holistic and multiattribute risk rankings. Although the observed increase was only marginally significant when all risk sets were considered, it was significant when the high-variance Risk Set D was excluded from the analysis.

We also asserted that risk rankings will be more useful to policy makers when participants are satisfied with the process and willing to stand behind their groups' rankings as representative of their concerns. Participants in this study reported high levels of satisfaction with their groups' decision-making processes and the resulting rankings, particularly when their groups were able to resolve disagreements. Moreover, individuals' final holistic rankings were more closely related to their groups' holistic rankings than to their own initial holistic rankings, indicating that participants "bought into" their groups' rankings. These results may allay concerns that participants were coerced into publicly accepting group decisions that they did not support.

Because individuals and groups may have legitimate reasons for disagreeing about the relative riskiness of different hazards, similarity among rankings produced by different parties is not a good measure of the quality of a ranking method. Of course, policy makers might prefer to have a clear signal of agreement rather than disagreement. In this study, there was substantial agreement on the ordering of risks, both among individuals and among groups. Previous results[13] suggest that individuals' initial agreement on holistic rankings is fostered by the risk summary sheets, which present relevant information in a consistent, comprehensible format that facilitates comparisons among risks. Group discussion also appears to foster agreement; for holistic rankings and attribute rankings, agreement among individuals was significantly higher after group discussion than before.

Although agreement on attribute rankings was lower than agreement on risk rankings, both among individuals and among groups, this may not create difficulties for policy makers. Because attributes are likely to be correlated across risks in almost any domain, the resulting multiattribute risk rankings may be somewhat immune to disagreements about the relative importance of attributes. Indeed, one possible reason for lower agreement on attribute rankings is individuals' limited ability to introspect on the determinants of their own judgments. Participants may have had particular difficulty discerning the impacts of the ranks they assigned to correlated attributes. The attribute-ranking procedure and the resulting multiattribute risk rankings are not unimportant, however. Other research suggests that individuals and groups revise their holistic rankings on the basis of their multiattribute risk rankings when given the opportunity to do so.[13–15]

4.2. High-Variance and Difficult-to-Rank Risks

Risks that evoke more divergent reactions may be of particular concern to policy makers. Such risks may be identified by comparing the rankings from several different groups and by assessing the difficulties that those groups experience. In this study, participants and groups who considered the high-variance Risk Set D showed lower consistency between their holistic and multiattribute risk rankings, lower agreement regarding these rankings, and weaker relationships between their holistic rankings and expected mortality and morbidity (see Tables I, II, III, and V). Participants considering Risk Set D may have focused their attention on different attributes for different risks, making it difficult for them to assess the relative importance of attributes across all risks in the set. Similarly, these risks may have induced greater use of anecdotal evidence than other risks, leading participants' attention away from the attributes. However, the relatively high satisfaction with groups' attribute rankings and the relatively high intergroup agreement on attribute rankings suggest that participants and groups who considered Risk Set D were reasonably attentive to the attributes. Alternatively, the risks in Set D may have unique features that are not adequately captured by our set of attributes. Attention to such features would be expected to lower the consistency between holistic and multiattribute risk rankings, weaken the relationships between holistic rankings and expected mortality and morbidity, and (if different participants and groups weight these features differently) lower the agreement regarding holistic rankings.

The results from Session 4 suggest that disagreements among groups can be predicted from difficulties within groups. Such difficulties were most likely when group members disagreed about the relative importance of attributes for particular risks, when particular risks had important features that were not captured by our set of attributes, and when group members had different levels of knowledge and experience with particular risks. The practical importance of predicting public disagreements means that replicating these results could be especially valuable.

4.3. Understanding, Uncertainty, and Controllability

Our results indicate that greater knowledge about the risks (greater scientific understanding and lower uncertainty) was associated with greater concern. Although Jenni[19] obtained the opposite result for rankings of school risks, our study is not the first to report a positive relationship between knowledge and perceived risk. For example, Slovic *et al.*[17] reported weak positive relationships in two studies, but weak to moderate negative relationships in a third study.[7] In more recent studies of perceived ecological[22–25] and health[26] risks, knowledge has been positively associated with perceived risk, worry, and priority for personal and government action. Although it is possible that researchers have learned more about hazards that people perceive as being particularly risky, the positive relationship between knowledge and perceived risk strikes us and others[25] as somewhat counterintuitive. In the study reported here, the regression results may have been driven by three high-uncertainty risks ("electric and magnetic fields from electric power," "asbestos," and "radon gas") for which the lower bounds for fatalities, injuries, and illnesses were given as zero. The summary sheets also reported the chance of zero risk as 80% or 90%, 40%, and 10% for these three risks, respectively. This information may have led participants to be less concerned about these risks, resulting in a positive relationship between the knowledge index and concern. Table III confirms that groups were less concerned about these three risks than might be anticipated on the basis of expected mortality alone.

We also found that greater ability of the student or parent to control exposure was associated with greater concern. Although controllability is often associated with lower perceived risk,[16,17] there are exceptions in the literature. Slovic *et al.*[17] reported different relations for different participant groups in one study, and recent studies of perceived ecological risk have reported nonsignificant[22–25] or positive[27,28] relationships between controllability and perceived risk. One explanation for the positive relationship observed in the study reported here is that participants disagreed with our controllability ratings for some risks. For example, we rated "self-inflicted injury or harm" as controllable, but participants may have thought otherwise. Alternatively, participants may have considered how the school or community might control these risks, despite our explicit instructions not to consider such risk-management options. If participants also thought that risks that could be controlled should be controlled, they may have given such risks higher priority.

[7] This reversal was attributed to the large number of chemical risks in the third study. Because the knowledge attributes were coded in the opposite direction (e.g., "not known to science" and "not known to those exposed"), Slovic *et al.*[17] reported that the effects were negative in the first two studies and positive in the third study.

4.4. Limitations

The study reported here has a number of limitations. First, the reported rankings clearly depend on the risks that were used, the information provided in the risk summary sheets, and the people who ranked the risks. These rankings are intended to illustrate how to evaluate a risk-ranking method, and should not be used for policy purposes. Comparable levels of consistency, satisfaction, and agreement might be obtained in similarly structured risk-ranking exercises. Second, the brevity and nature of the risk-ranking sessions precluded the use of the full set of risks, the integration of holistic and multiattribute risk rankings, and the collection of detailed data on groups' decision-making processes. We are currently conducting longer, more detailed risk-ranking exercises with lay groups to address these issues.[15] Finally, the study reported here considered only health and safety risks. We are currently extending the method to incorporate ecological risks and their attributes.[27,28]

5. CONCLUSIONS

The risk-ranking method developed at Carnegie Mellon University is designed to reveal participants' relative levels of concern about health and safety risks by providing clear, concise, and consistently formatted risk information in a setting that encourages individual and group deliberation about the important features of the risks. The data presented here indicate that the method can yield reasonably high levels of consistency and participant satisfaction, as well as substantial agreement on risk rankings, at least among risk managers. The method also allows individuals who have not participated in a risk-ranking exercise to ascertain the original views of the participants and examine how group discussion affected those views. Such transparency should make the resulting rankings more valuable to policy makers, who might need to explain or defend the rankings to parties not present at the time of their creation. We invite others to use the method and report their experiences with different risks and participants.

ACKNOWLEDGMENTS

This work was supported by a U.S. Environmental Protection Agency (EPA) STAR Graduate Fellowship and by grants from the National Science Foundation (SRB-9512023 and SES-9975200), the EPA (R8279200-1-0), the Electric Power Research Institute (W02955-12), the Alcoa Foundation, and the Chemical Manufacturers Association. We are grateful to the participants for their time and thoughtfulness, and to George Gray and John Graham for providing the opportunity to collect these data. We also thank Karen Jenni, Jun Long, Claire Palmgren, Henry Willis, and Patti Steranchak for their advice and assistance on this project.

REFERENCES

1. 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.
2. Stone, E. R., Yates, J. F., & Parker, A. M. (1994). Risk communication: Absolute versus relative expressions of low-probability events. *Organizational Behavior and Human Decision Processes, 60*, 387–408.
3. Loomis, J. B., & duVair, P. H. (1993). Evaluating the effect of alternative risk communication devices on willingness to pay: Results from a dichotomous choice contingent valuation experiment. *Land Economics, 69*, 287–298.
4. Golding, D. S., Krimsky, S., & Plough, A. (1992). Evaluating risk communication: Narrative vs. technical presentations of information about radon. *Risk Analysis, 12*, 27–35.
5. Kaplan, R. M., Hammel, B., & Schimmel, L. S. (1986). Patient information processing and the decision to accept treatment. *Journal of Social Behavior and Personality, 1*, 113–120.
6. Fischhoff, B. (1991). Value elicitation: Is there anything in there? *American Psychologist, 46*, 835–847.
7. Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives.* New York: Cambridge University Press.
8. von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research.* New York: Cambridge University Press.
9. Tyler, T. R. (1990). *Why people obey the law.* New Haven, CT: Yale University Press.
10. Atman, C. J., Bostrom, A., Fischhoff, B., & Morgan, M. G. (1994). Designing risk communications: Completing and correcting mental models of hazardous processes, part I. *Risk Analysis, 14*, 779–788.
11. Bostrom, A., Atman, C. J., Fischhoff, B., & Morgan, M. G. (1994). Designing risk communications: Completing and correcting mental models of hazardous processes, part II. *Risk Analysis, 14*, 789–798.
12. Morgan, M. G., Fischhoff, B., Bostrom, A., & Atman, C. J. (2002). *Risk communication: A mental models approach.* New York: Cambridge University Press.
13. Morgan, K. M. (1999). *Development and evaluation of a method for risk ranking.* Unpublished doctoral dissertation, Carnegie Mellon University.
14. Morgan, K. M., Fischbeck, P. S., & DeKay, M. L. (1999, November). *Assessing a multi-attribute model for ranking risks.* Paper presented at the annual meeting of the Institute for Operations Research and the Management Sciences, Philadelphia, PA.
15. Fischbeck, P. S., DeKay, M. L., Fischhoff, B., Morgan, M. G., Florig, H. K., Palmgren, C. R., & Willis, H. H. (2000, December). *Evaluating a risk-ranking methodology.* Presentation at the annual meeting of the Society for Risk Analysis, Arlington, VA.
16. Barron, F. H., & Barrett, B. E. (1996). Decision quality using ranked attribute weights. *Management Science, 42*, 1515–1523.

17. Slovic, P., Fischhoff, B., & Lichtenstein, S. (1985). Characterizing perceived risk. In R. W. Kates, C. Hohenemger, & J. X. Kasperson (Eds.), *Perilous progress: Managing the hazards of technology* (pp. 91–125). Boulder, CO: Westview Press.
18. Slovic, P. (1986). Perception of risk. *Science, 236*, 280–285.
19. Jenni, K. E. (1997). *Attributes for risk evaluation.* Unpublished doctoral dissertation, Carnegie Mellon University.
20. Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach.* Orlando, FL: Harcourt Brace Jovanovich.
21. Kendall, M., & Gibbons, J. D. (1990). *Rank correlation methods* (5th ed.). New York: Oxford University Press.
22. McDaniels, T., Axelrod, L. J., & Slovic, P. (1995). Characterizing perception of ecological risk. *Risk Analysis, 15*, 575–588.
23. McDaniels, T., Axelrod, L. J., & Slovic, P. (1996). Perceived ecological risks of global change. *Global Environmental Change, 6*, 159–171.
24. McDaniels, T., Axelrod, L. J., Cavanagh, N. S., & Slovic, P. (1997). Perception of ecological risk to water environments. *Risk Analysis, 17*, 341–352.
25. Lazo, J. K., Kinnell, J. C., & Fisher, A. (2000). Expert and layperson perceptions of ecosystem risk. *Risk Analysis, 20*, 179–193.
26. Baron, J., Hershey, J. C., & Kunreuther, H. (2000). Determinants of priority for risk reductions: The role of worry. *Risk Analysis, 20*, 413–427.
27. DeKay, M. L., & Willis, H. H. (2000, November). *Public perceptions of environmental risks.* Presentation at the annual meeting of the Society for Judgment and Decision Making, New Orleans, LA.
28. Willis, H. H., DeKay, M. L., Fischbeck, P. S., Fischhoff, B., Florig, H. K., Morgan, M. G., & Palmgren, C. R. (2000, December). *Extension of the Carnegie Mellon risk-ranking test bed to include environmental and ecological factors.* Presentation at the annual meeting of the Society for Risk Analysis, Arlington, VA.