Risk Analysis, Vol. 10, No. 3, 1990

What Do We Know About Making Risk Comparisons?

Emilie Roth,[1] M. Granger Morgan,[2,3] Baruch Fischhoff,[2] Lester Lave,[2] and Ann Bostrom[2]

Received April 5, 1990

The risks of unfamiliar technologies are often evaluated by comparing them with the risks of more familiar ones. Such risk comparisons have been criticized for neglecting critical dimensions of risky decisions. In a guide written for the Chemical Manufacturers Association, Covello *et al.*[1] have summarized these critiques and developed a taxonomy that characterizes possible risk comparisons in terms of their acceptability (or objectionableness). We asked four diverse groups of subjects to judge the acceptability of 14 statements produced by Covello *et al.* as examples of their categories. We found no correlation between the judgments of acceptability produced by our subjects and those predicted by Covello *et al.*.

KEY WORDS: Risk comparison; risk communication; risk perception.

1. INTRODUCTION

A tempting way to describe the risks of hazardous technologies is by comparison with other, better known risks,[2,3] such as: the cancer risk of living at the boundary of a nuclear power plant for 5 years equals the cancer risk of eating 40 tablespoons of peanut butter (due to aflatoxin).[3] Despite their appeal,[4] such comparisons have come in for considerable criticism.[5-7] There are two major thrusts to this criticism. One is that these comparisons reduce risks to a single dimension (e.g., loss of life expectancy), whereas many risks are multidimensional. As a result, risks are not fully represented. The second thrust is that risk comparisons are used not just to communicate how large risks are, but also to persuade listeners regarding how large risks should be (e.g., if you are willing to eat 40 tablespoons of peanut butter over the next 5 years, then you should be willing to live near a nuclear power plant). Such implicit rhetorical arguments ignore critical elements of people's risky decisions, such as how voluntary the choices are and what benefits they are expected to provide. Because people perceive risks in multiattribute terms, the fact that a risk has a low value on a single focal dimension (e.g., estimated fatalities in an average year) does not imply its acceptability.[8] As a result of these logical and ethical flaws, it should not be surprising that risk comparisons have provoked anger and mistrust (responses that can only be aggravated by skepticism about how far the risks estimates themselves can be trusted).

In order to help chemical industry spokespeople avoid these pitfalls, Covello *et al.*[1] developed a manual advising plant managers on how to present risk comparisons so that the public will perceive them as useful and legitimate. Their manual has been published and distributed widely by the Chemical Manufacturers Association.

The manual represents a significant contribution to the risk communication literature. It provides, for the first time, an analysis of the different ways that risk comparison statements have traditionally been employed, and offers a framework for evaluating them. Covello *et al.* enumerate 14 commonly used types of risk comparisons, which they then group into five categories, ranked according to their predicted acceptability

[1]Westinghouse Science and Technology Center, Pittsburgh, Pennsylvania 15235.
[2]Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213.
[3]To whom correspondence should be addressed.

0272-4332/90/0900-0375$06.00/1

to lay people (see Table I). The manual recommends that spokespeople select the highest ranking risk comparisons whenever possible, and use low ranking risk comparisons with caution, alert to the possibility that communications using them could backfire.

Because the research base is thin, Covello *et al.*'s ranking is based on their accumulated experience and intuitions. Because of its potential significance for guiding risk communication, their proposal warrants empirical evaluation. The present study focused on how well Covello *et al.*'s ranking predicted lay people's judgments of the acceptability of risk comparisons. Its results provide us with a point of departure for a throretical analysis of Covello *et al.*'s proposal.

2. THE STUDY

The Covello *et al.* manual provides concrete examples of their 14 categories of risk comparisons, set in the context of a specific scenario: A manager of a chemical plant in a small town is faced with the task of communicating to the community about the risk of a chemical produced by the plant (see Appendix). We asked several groups of laypeople to evaluate the acceptability of these statements.

Such an evaluation requires an operational definition of "acceptability." The definition intended by Covello *et al.* is suggested by the following quotation.

> The highest-ranking comparisons are assumed to be those that put the least strain on the trust relationship between a plant manager and the public. These comparisons tend to strike even skeptical listeners as relevant, appropriate, and helpful information. The lowest-ranking comparisons, on the other hand, are those that have no intuitively obvious claim to relevance, appropriateness, or helpfulness. Such comparisons are more likely to be seen as manipulative or misleading—that is, as efforts to preempt judgments about the acceptability of the risk. (p. 17)

Thus, there are several distinct elements that contribute to acceptability. As a result, we devised seven rating scales that seemed to tap different elements of Covello *et al.*'s definition of "acceptable." These scales appear in Table II. Scale 1 asks about how clear and easy to understand the statement is. Scales 2 and 3 consider the perceived relevance and helpfulness of the risk comparison. Scale 4 ask whether the risk comparison seems misleading, in the sense of underemphasizing or overemphasizing the risk. Scales 5 and 6 ask how the risk comparison will affect public trust in the plant manager. Scale 7 provides an overall measure of acceptability, by asking whether the statement should be included in the plant manager's talk. Our subjects' response should reveal how these alternative criteria are correlated with one another as well as with Covello *et al.*'s predictions.

2.1 Method

2.1.1. Participants

Four groups participated in the study: (A) second-year graduate business students ($N = 13$); (B) members (or their spouses) of a suburban garden club from a mid-

Table I. Covello *et al.* Risk Comparison Categorization and Ranking System

First-rank risk comparisons
1. Comparisons of the same risk at two different times
2. Comparisons with a standard
3. Comparisons with different estimates of the same risk
Second-rank risk comparisons (second choice—less desirable)
4. Comparisons of the risk of doing and not doing something
5. Comparisons of alternative solutions to the same problem
6. Comparisons with the same risk as experienced in other places
Third-rank risk comparisons (third choice—even less desirable)
7. Comparisons of average risk with peak risk at a particular time or location
8. Comparisons of the risk from one source of a particular adverse effect with the risk from all sources of that same adverse effect
Fourth-rank risk comparisons (fourth choice—marginally acceptable)
9. Comparisons of risk with cost, or of cost/risk ratio with cost/risk ratio
10. Comparisons of risk with benefit
11. Comparisons of occupational with environmental risks
12. Comparisons with other risks from the same source, such as the same facility or the same risk agent
13. Comparisons with other specific causes of the same disease, illness, or injury
Fifth-rank comparisons (last choice—rarely acceptable—use with extreme caution!)
14. Comparisons of unrelated risks.

Table II. Scales Used to Rate Covello *et al.* Statements.

1	This statement is clear, easy to understand.	☐	☐	☐	☐	☐	This statement is unclear, difficult to understand.
2	This statement will help townspeople to better understand the risk.	☐	☐	☐	☐	☐	This statement will *not* help townspeople to better understand the risk.
3	This statement gives information needed by townspeople in their personal decisions about the risk.	☐	☐	☐	☐	☐	This statement gives *no* information needed by townspeople in their personal decisions about the risk.
		This statement's tone correctly conveys the risk.					
4	This statement's tone underemphasizes the risk.	☐	☐	☐	☐	☐	This statement's tone overemphasizes the risk.
5	This statement is likely to reassure the townspeople.	☐	☐	☐	☐	☐	This statement is likely to scare the townspeople.
6	This statement is likely to increase the townspeople's trust in the plant manager.	☐	☐	☐	☐	☐	This statement is likely to decrease the townspeople's trust in the plant manager.
7	This statement should definitely be included in the plant manager's talk.	☐	☐	☐	☐	☐	This statement should definitely be left *out of* the plant manager's talk.

dle-to-upper income community ($N=33$); (C) members of a synagogue ($N=28$); and (D) members of a Protestant church ($N=21$) from middle and lower income communities in Pittsburgh. The 95 total participants included a wide range of ages, socioeconomic backgrounds, religions, and both sexes. Participants were either paid $10 or had a $10 donation made to their organization.

2.1.2. Materials

In order to introduce the evaluation task, we converted the scenario described in the manual into a cover story which read as follows:

> Suppose that the manager of a chemical plant that manufactures ethylene oxide in the small midwestern town of Evanston has been asked to give a talk to a local community meeting about risks posed by his plant. The local newspaper plans to reprint the speech in its entirety and make it widely available. People in the town are concerned about the possible risks posed by the plant, but there is no crisis situation or serious confrontational atmosphere.
>
> The plant manager has been a friend of yours for many years. He is concerned about making this speech and, as an old friend, has asked you for your candid advice about some things he is considering saying.
>
> Before starting, here is some background information: Ethylene oxide is used in almost all hospitals and other medical facilities as a disinfecting agent. However, it can cause cancer. A risk assessment has shown that the cancer risk that the Evanston plant poses for citizens living in the town is about two additional cancers per year for every million people exposed (there are in fact only 3500 people in Evanston). The plant manager is looking for appropriate and acceptable ways to communicate this risk to the public and to compare it with other risks.
>
> He wants to give a clear honest picture of the risks. He feels that this is both his ethical responsibility and that if he were to misrepresent the situation, eventually that would be discovered and hurt his credibility. He is concerned, however, that even an accurate statement can come out sounding wrong or have the wrong impact. He also wants to keep the talk fairly short and simple, while still doing the topic justice.
>
> The following are 14 different pieces of text that the plant manager is considering using in his talk. Some of them overlap a bit in content. Assume that he will edit them so that they fit together well without much overlap. For each statement, please give your advice on the following questions.

This cover story appeared on the front page of a booklet that contained the 14 statements. There was one statement per page. Each statement appeared on the left side of its page, while the seven rating scales appeared on the right.

As indicated in Table II, each rating scale had five points with endpoints labeled. These were coded 1–5

from left to right. With the exception of scale 4 (tone of statement), a lower number indicates a more favorable value. In the case of scale 4, both endpoints of the scale represent unfavorable values (1 = underemphasizes the risk; 5 = overemphasizes the risk).

The order of presenting the 14 statements was varied across participants. Fifteen of group B received the statements in Covello *et al.*'s original order, while the remaining 18 received the statements in the reverse order. Two random orders of the 14 statements were also generated. Approximately half of the participants in each of the other three groups received the statements in each of these orders.

Groups A, C, and D completed the questionnaires in a group setting at the site of their organization or class. Group B members received brochures by mail.

2.2. Results

2.2.1. Results Across Groups

Table III shows mean responses for each statement on each scale for all 95 participants. With the exception of scale 4, Covello *et al.*'s proposal predicts that each successive group of statements will have higher means than its predecessors.[4] This was not found. Spearman rank-order correlations were computed between the mean ratings of each of the 14 statements and the rank order of the class to which it belongs. Table IV presents these correlations, both across all 95 participants and for each of the 4 groups.[5] None of the seven scales was significantly correlated with Covello *et al*'s. order in the direction predicted. For all participants combined, the correlation with scale 7 (whether to include the statement in the plant manager's talk) is close to zero ($r = -0.13$). The only significant correlation ($r = 0.51$, $p < 0.05$) is that with scale 1 (clarity of statement). However, its sign is opposite to that predicted by Covello *et al.* Each of the four groups produced a similar pattern of results, described more fully below.

Friedman two-way analyses of variance computed on the rank sums across the 95 participants were significant for all seven scales ($p < 0.001$). This nonparametric test indicates that there are reliable differences in

[4]On scale 4 a "3" was the most favorable value. Because all mean responses for scale 4 were less than 3, higher ratings indicate more favorable responses.

[5]Analyses were also performed on the rank sums for each statement. The rank sum for each scale was computed by determining each participant's rank ordering of the 14 statements. The rank sums across the 95 participants were highly correlated, with the mean scores appearing in the table (all correlations above 0.85). The results using this measure were essentially the same as when mean scores were examined.

Table III. Mean Responses for the 14 Sentences on Each Scale (Average Across all 95 Participants)

		Scales[a]						
		1	2	3	4	5	6	7
Rank	Statements[b]	Clarity	Aids understanding	Information needed	Under/over-emphasizes risk	Reassuring	Increases trust	Should be included
First	1	1.71	2.16	2.10	2.77	2.00	1.75	1.82
	2	2.29	2.76	2.55	2.57	2.54	2.50	2.55
	3	3.02	2.73	2.32	2.87	2.95	2.54	2.92
Second	4	2.19	2.67	2.37	2.67	2.94	2.71	2.68
	5	2.10	2.54	2.32	2.66	3.04	2.59	2.98
	6	1.69	2.69	2.58	2.24	2.19	2.33	2.35
Third	7	2.17	2.48	2.24	2.81	2.85	2.71	2.71
	8	2.50	2.51	2.34	2.85	2.76	2.44	2.61
Fourth	9	1.63	2.70	2.37	2.81	2.44	2.44	2.36
	10	1.56	2.81	2.10	2.53	2.65	2.53	2.27
	11	2.08	2.13	1.88	2.66	2.26	2.42	2.37
	12	2.15	3.67	3.47	2.12	3.25	3.44	3.63
	13	1.51	1.98	2.03	2.62	2.44	2.41	2.42
Fifth	14	1.82	2.03	2.18	2.57	2.35	2.48	2.39

[a]For scales 1 through 3 and 5 through 7, 1 is the most favorable response. For scale 4, 1 = underemphasizes risk, 5 = overemphasizes risk.
[b]The statements are listed in decreasing favorability, according to Covello *et al.'s* predictions.

Table IV. Spearman Rank-Order Correlation with the Covello *et al.* Ranking[a]

Scale	All groups	Garden club	MBA students	Synagogue	Church
Clarity	−.51	−.60	−.27	−.19	−.45
Aids understanding	−.24	−.04	−.16	−.12	−.24
Information needed	−.31	−.43	−.04	.03	−.32
Over/underemphasizes risk	−.35	−.36	−.06	−.32	−.18
Reassuring	−.10	−.42	.07	.02	−.09
Increases trust	.01	.23	.29	−.12	.11
Should be included	−.13	−.30	−.09	.08	−.02
	$N = 95$	$N = 33$	$N = 13$	$N = 28$	$N = 21$

[a]All correlations at or above .46 are significant at the .05 level. Correlations at or above .65 are significant at the .01 level.

the ratings among the 14 statements (not just the differences that were predicted).

Table V presents Pearson correlations among the seven rating scales, computed on mean ratings over all 95 participants. As can be seen, these means tended to be positively and significantly correlated,[6] indicating that statements judged positively in one respect were also judged positively in others. These results indicate that the weak correlations between scale ratings and the Covello *et al.* ranking cannot be be attributed to their being such poor measures that they cannot correlate with anything. Although all scales correlated with subjects' judgments of whether a statement should be included (scale 7), the strongest predictors were how reassuring it seemed and whether it seemed likely to increase trust.

The statements tended to be rated positively on all scales, with a rating of "1" given in almost 40% of all cases. One possible explanation is that the verbal labels anchoring the scales were too moderate (so that 1 connotes *good* rather than *excellent* performance). The resulting "ceiling effect" would reduce differences between statements, even though there were still statically reliable differences in acceptability (see Section 2.2.4). A second possibility is that most statements were actually pretty good, even though some were intended to represent seriously flawed risk comparisons (see Section 3.1).

Table V. Correlation Matrix for the Seven Scales

	Scale						
Scale	1	2	3	4	5	6	7
1 Clarity	1.00						
2 Aids understanding	0.28	1.00					
3 Information needed	0.22	0.88	1.00				
4 Over/underemphasize risk	0.32	−0.52	−0.66	1.00			
5 Reassuring	0.56	0.66	0.55	−0.09	1.00		
6 Increases trust	0.29	0.74	0.72	−0.49	0.82	1.00	
7 Should be included	0.53	0.71	0.75	−0.36	0.90	0.91	1.00

[6]As mentioned, higher ratings indicate more favorable responses on scale 4, so that the negative correlations there are consistent with the positive correlations on the other variables.

2.2.2. *Breakdown by Group*

The results are similar when the four groups are considered separately. For three groups, there was no significant correlation between mean scale ratings and the Covello *et al.* ordering. For group B, there was a negative correlation (-0.60; $P < 0.05$) between Covello *et al.*'s ranking and subjects' clarity ratings.

Every correlation between mean scale ratings of the different groups was positive, indicating a consistent degree of agreement. Correlations ranged between 0.23 and 0.88 with a mean, using Fisher's Z-transformation, of 0.63.

2.2.3. *Effects of Order of Presentation*

Mean ratings were computed separately for each of the four orders of presentation. Three of the four groups were highly similar to one another and to the overall averages. These were the two groups receiving random orders and the group rating the 14 statements in the order predicted to show decreasing acceptability. These means were all unrelated to Covello *et al.*'s prediction order. The ratings of the 15 participants who received statements in Covello *et al.*'s original order were significantly correlated ($P < 0.05$) in three cases. Two were in the predicted direction, scales 4 and 6 (-0.57 and 0.52, respectively); while one, scale 5 (-0.52) was in the opposite direction. Overall, the weak and inconsistent pattern with this small group does not shake the general conclusion that order of presentation did not affect subjects' ratings.

2.2.4. An Ordered Categorical Response Model

An ordered categorical response model, specifically a three-level ordered probit model, was used to clarify the differences in ratings among the 14 statements.[9,10][7] The model included the 14 statements, 7 scales, 4 orders of presentation, and 4 groups as predictor variables and the ratings as the dependent variable. Ratings were recoded into three categories, where 0 was "best" (rating "3" on scale 4; "1" and "2" on other scales), 1 was intermediate ("2" and "4" on scale 4, "3" on others), and 2 was "worst" ("1" and "5" on scale 4; "4" and "5" on others).[8] The model was estimated in LIMDEP,[11] using maximum likelihood estimation. The base case (represented by the intercept) was item 14, scale 7, order 1, and group 4 (D). This analysis characterizes predictors by beta coefficients that indicate changes in the underlying dependent variable, all else being equal. According to Covello *et al.* hypothesis, the beta coefficients for statements 1–13 should all be negative because each is contrasted with statement 14, which was predicted to be the worst. The coefficients should be increasingly negative as the statements become more attractive and statement number decreases. The beta coefficients for the 14 statements and their 95% confidence bands are presented in Fig. 1. They show reliable differences in ratings among the 14 statements that are not captured by the Covello *et al.* ranking system, even when effects of scale, order of presentation, and group are statistically controlled. The beta coefficients typically had the wrong sign (positive). There was no consistent trend over the five ranks.

The analysis yielded significant coefficients for scale and group, but not for order of presentation. The lack of an order effect with this more sophisticated analysis strengthens our inclination to discount the weak differences reported in Section 2.2.3. The overall fit of the model is moderately good. The χ^2 statistic from the log-likelihood ratio test is highly significant (727.5, 24 df, $p < 0.001$) and the model correctly predicts 55% of the observations.

[7]An ordered probit model assumes that the observed ratings are discrete and have ordinal properties (i.e., no interval relation between rating points is assumed), but that the underlying (unobservable) dependent variable (i.e., statement acceptability) is continuous and normally distributed, conditional on the predictive variables.

[8]The original five-point ratings for all scales, except 4, were also fit with an analogous model as was an alternative 3-point set of collapsed ratings (0 = 1; 1 = 2,3,4; 2 = 5 for all scales except 4, which was collapsed as above). Similar results were obtained and are available upon request.

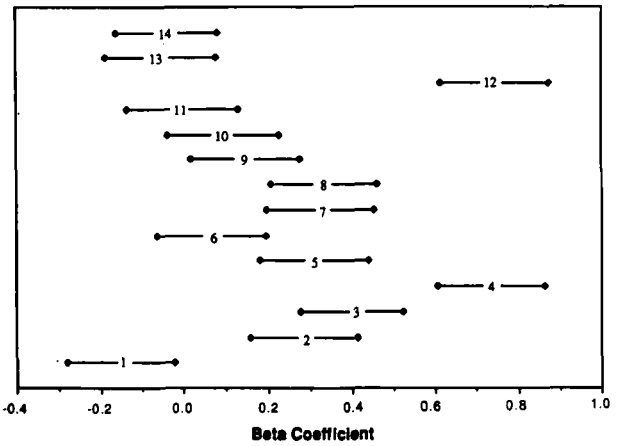

Fig. 1. Estimated beta coefficients for the statement dummies from the ordered probit regression model, with 95% confidence intervals marked (based on the coefficient's estimated standard deviation). Statement 14 is the base case (intercept).

3. DISCUSSION

Our subjects' ratings reliably distinguished among the statements, but not in the way predicted by Covello *et al.* This section discusses why Covello *et al*'s predictions might have fared so poorly and offers some alternative perspectives on risk comparison statements.

3.1. Risk: Comparisons Deviating From Predictions

One place to look for insight is at those statements whose ratings deviated the most from the Covello *et al.* predictions. As can be seen in Table III, three statements at the top of Covello *et al.*'s list were near the bottom of our subjects' ratings, while three of the four worst statements according to Covello *et al.* were rated among the best here.

3.1.1. Comparisons of Risks Across Domains Fared Better Than Expected

According to Covello *et al.* and others,[7] risk comparisons are particularly problematic when they involve risks with very different features. As a result, the examples that Covello *et al.* identify as worst involve risks from different domains. Their statement 13 (representing comparisons that invoke other specific causes of the same consequence) compares the risk of cancer from the chemical ethylene oxide to the risk of cancer from x-rays. Their least favored statement (14) compares ethylene oxide with other hazards whose consequences did

not include cancer (e.g.,lightning). Nonetheless, both statements were in the top half of the set for six of the seven scales. Indeed, they were the highest ranked statements on scale 2, how much a statement would "help townspeople to better understand the risk."

Covello *et al.*'s critique of cross-risk comparisons applies most strongly to cases where they are advanced with a rhetorical purpose—of the form "if you accept Risk A, then you ought to accept (equivalent) Risk B." Such comparisons have, however, no logical force unless the two risks are equivalent on all their risk features (not to mention their associated benefits and control options). A more modest use of risk comparisons is to convey a feeling for the magnitude of a risk, with no claim of acceptability. Such magnitude comparisons might focus on either the probability of negative consequences (e.g., as likely as being struck by lightning during an equivalent exposure period) or on their intensity (e.g., as painful as a root canal without anesthesia). Given their more limited ambitions, magnitude comparisons should be easier to make appropriately than acceptability comparisons.

Conceivably, Covello *et al.*'s own sensitivity to these issues kept them from creating truly bad risk comparisons, particularly ones containing indefensible acceptability arguments. As a result, our subjects were able to focus on the magnitude comparisons in the statements. These were, in turn, executed relatively well. If that is the case, then, in effect, Corvello *et al.* foiled their own prediction when they created the illustrative statements.

3.1.2. Comparison of Occupational with Enviromental Risks Fared Better Than Expected

A second unexpected success was statement 11, which was intended to exemplify comparisons between occupational risks and environmental risks. Rather than emerging near the bottom of the ratings, statement 11 appeared in the top half of all seven scales. It was ranked best on scale 3 ("gives information needed by the townspeople in their personal decisions about the risk") and was one of the top 3 statements on scale 2 ("will help townspeople to better understand the risk") Covello *et al.* do not explain why they expected such comparisons to be received particularly poorly. One possible reason is that the assumption of occupational risks often implies the acceptance of risk–benefit tradeoffs that seem quite inappropriate outside of working life.

However, although Covello *et al*'s statement 11 does refer to occupational and environmental risks, it does not invite risk–benefit comparisons. Rather, its main thurst is that the risk to employees is very small, implying that the risk to the community will be even smaller. Again, the example may have fared unexpectedly well because it lacked the particular feature of its category that people find objectionable.

3.1.3. Comparison with a Standard and Comparisons with Different Estimates of the Same Risk Fared Worse Than Expected

Covello *et al.* stressed the importance of being honest and forthright in providing risk information. Elements of such frankness include indicating uncertainties or disagreements regarding the size of the risk, discussing worst-case estimates as well as best-guess estimates, and noting how a risk compares to various proposed standards of acceptability.

Statements 2 and 3 were intended to exemplify this principle. Statement 2 compares the focal risk to five different emission standards, while statement 3 provides six alternative estimates of the size of the risk, based on different data, different assumptions, and different originating sources. Both statements should have been attractive. However, each was ranked in the bottom half of the set on six of the seven scales. They were among the worst three items on scale 1, measuring how "clear, easy to understand" a statement was. This last result suggests that these statements may have been ranked so poorly because of the quantitative and probabilistic information that they contained. The price paid for such candor may have been confusing recipients. Statement 3 may have been particularly difficult because it included small probabilities presented in decimal form (e.g., 0.007 cancers per 3500 persons). The Covello *et al.* manual itself explicitly warns against this format. Statement 8, which was designed to reflect a more effective way of communicating small probabilities, had some of the worst ratings on the clarity scale. Apparently, we still have much to learn about presenting such information.

A second possible source of confusion in these statements was the need to integrate the multiple perspectives that they presented. For example, what are recipients to make of a risk that meets one of several standards, especially when they know little about the organization that set each standard or the purpose for which it was set? Similarly, how are they to recocile competing scientific estimates of a particular risk without understanding the underlying science (and scientists) producing those estimates? Offering multiple perspectives may be a meaningless gesture unless recipients can

put them into context. Clearly, more research is needed here as well.

3.1.4. Comparisons of Risk of Doing and Not Doing Something Fared Worse Than Expected

Statement 4, which was intended to illustrate comparing the risks of doing and not doing something, received unexpectedly poor evaluations. It ranked in the bottom half of the set on six of the seven scales, faring particularly poorly on scales 1 (clarity), 5 (reassuring), and 6 (increases trust). It shared the bottom in the ordered probit analysis (Fig. 1). Here, too, presenting small probabilities in decimal form may have been problematic. In addition, statement 4 notes that the risk could be reduced (by a small amount) by purchasing new equipment, but without indicating whether the plant intends to do so. Silence on that issue may have raised suspicions and reduced ratings related to trust.

3.2. Explanations for Failure of Predictions

Reviewing our results in the light of these arguments suggests three reasons why Covello *et al.*'s predictions may have failed.

3.2.1. Flaws in Measurement

The first possibility is that Covello *et al.*'s theory is correct, but our rating scales failed to measure what they intended by "acceptability." As mentioned, we used a variety of rating scales in an attempt to capture the diverse elements of the complex notion of "acceptability" advanced by Covello *et al.* It is, of course, possible that none of our rating scales was related to the lay notion of "acceptability." However, the fact that so diverse a set of scales failed to correlate with Covello *et al*'s predicted ranking indicates the need to clarify the goals of risk comparisons as well as to study how to reach them.[8]

[8]Any other feature of our measurement procedure might also be called into question. For example, in their thoughtful response to this article, Slovic *et al.*[12] wonder about what would have happened had we used another cover story. Progress here requires accounting for both those patterns that did emerge in previous studies as well as for those that did not.

3.2.2. Flaws in the Examples

A second possible source of failure is that the 14 statements did not capture the essence of the categories that they were meant to represent. Section 3.1. raises some such possibilities (e.g., avoiding the risk acceptability arguments that can make some categories offensive, burdening relatively sound comparisons with unfamiliar decimal probabilities). The fact that recognized experts of this field might encounter such problems suggests the limits to our understanding of risk comparisons.

3.2.3. Flaws in the Underlying Theory

A third possibility is that the theory underlying the ranking system is flawed. It is always difficult to falsify a theory when there is uncertainty about how it should be implemented and evaluated. Nonetheless, it should be troubling to find failures with statements produced by the theory's creators and evaluation scales adapted from their stated objectives.

In Covello *et al.*'s theory, there are two obvious places to work on: its classification scheme and the predicted rankings of its categories. Covello *et al.*'s classification scheme sorts risk comparisons primarily according to what risks are being compared, and only secondarily according to the purpose of the comparison or the specific information that it contains. Elaborating these features may be a way to improve our understanding of risk comparisons.

3.3. Toward a Systematic Classification of Risk Comparisons

One significant contribution of the Covello *et al.* proposal is describing the variety of features of a risk that comparison statements can highlight. Indeed, each category in their system deals with a different aspect of risk. For example, statement 1 describes trends over time, while leaving the communication of absolute and relative magnitude to other statements. It seems unlikely that any criterion of acceptability could apply to messages having such a variety of purposes. Each is legitimate for some purposes and flawed for others, with its acceptability depending heavily on the quality of its implementation.

One way to conceptualize the potential purposes of risk comparisons is according to the roles that they may play in helping people to make decisions about risks. From a decision theory perspective, a decision involves

a choice among options, each of which can be characterized by a vector of attributes, representing its possible consequences. With risky decisions, at least some of those attributes involve uncertain negative consequences. When considering decision options (risky or otherwise), one needs to go through three stages: identifying the set of relevant attributes (i.e., the ones that might matter when one makes a choice), characterizing each option in terms of each attribute, and determining the relative importance of each attribute (in this set of options).

Risk comparisons have a legitimate role to play in supporting each of these stages. That is, they can help people to determine:

1. what attributes merit consideration;
2. how each option rates on each relevant attribute; and
3. how those attributes should be weighted.

3.3.1. Evoking Attributes of a Risk

Risk perception research has found that people are capable of rating risks on a large set of attributes (e.g., voluntariness, equity, dread), which are relevant to their judgments of risk acceptability.[13,14] The fact that these attributes are recognized when they are presented explicitly carries, however, no assurance that will be recalled spontaneously when a risk is mentioned. Indeed, the great number of possible attributes means that it would be hard to bear all in mind at once. A risk comparison might be able to help people by evoking decision-relevant attributes that they might otherwise neglect. Doing so in an unbiased fashion will pose a challenge to the design of communications. Considerations that are out of sight tend to be out of mind.[15,16] Conversely, those comparisons that are made may powerfully shape the attributes that people do consider (e.g., "This is the next dioxin" or "They tell us this is safe, but that's what they said about cigarettes and Agent Orange" or "They are just like tobacco company scientists").

3.3.2. Determining the Values on Risk Attributes

Once the attributes relevant to a decision have been identified, decision-makers must determine how each option rates on each attribute. Conveying information about the magnitude of consequences is one clear purpose of risk communications. As mentioned, risk comparisons might be a useful tool for doing so, by providing a familiar point of comparison for an unfamiliar hazard—as long as claims of risk acceptability can be avoided.

3.3.3. Crystallizing Preferences

A final role for risk communications is helping people examine and crystallize their own preferences. Simplistic models of decision-making assume a high degree of articulation in people's preferences, namely, they will know how to make all relevant tradeoffs, judging the relative importance of different outcomes. However, with options involving the sort of esoteric consequences involved with many risky decisions, people may welcome noncoercive suggestions of alternative perspectives.[17] Properly qualified risk comparisons might fulfill that role.

3.4. Reflections on Category Definition

The analysis above suggests that the details on content may be more important than the form of a risk comparison in determining its acceptability. This may explain some of the lack of predictive power of the Covello *et al.* classification scheme. In some cases, the categories in Covello *et al.*'s taxonomy are sufficiently broad to include statements with quite varied character. Conversely, statements that communicate very similar information by different means are sometimes classified separately. For example, statement 11 uses the experience of plant employees as an upperbound estimate of the risk to the townspeople. As such, this statement might arguably belong in category 6 with (other) comparisons that use the risk level experienced by one group as an input to estimating the risk to another.

Category 4 ("comparison of the risk of doing something versus not doing it") provides another example of a category that includes comparisons with varied content. For example, it includes both actions intended to increase risk and actions intended to reduce risks, which may invoke different attitudes. Moreover, all such comparisons invoke risk–benefit tradeoffs, insofar as as other consequences accompany these actions. As a result, category 4 overlaps category 9. The fact that these tradeoffs are left implicit in statement 4 may account for some reasons why it was judged more poorly than statement 9, where the tradeoffs are explicit.

4. CONCLUSIONS

Covello *et al.* have enumerated and classified a variety of risk comparisons. They were not, however, able

to predict the acceptability of statements generated to represent those categories, at least as measured by our subjects' responses. This failure seems to reflect a combination of (1) difficulty in translating the theory into concrete communications, (2) confounding the different possible purposes of risk comparisons within individual messages, and (3) the absence of adequate research on how to represent different kinds of information credibly. As a result, we need more and better theoretical and empirical research to build on Covello *et al.*'s challenging beginning.

APPENDIX

The following is the text of the 14 specific risk comparison statements, developed by Covello *et al.* (1988), which were evaluated in this research.

Statement 1

"Health risks from emissions of ethylene oxide at our plant are 40% less than a year ago, when we installed exhaust scrubbers. With more equipment coming in, we expect to reduce the risk another 40% by the end of the next year."

"Despite the extremely low health risks to the community from emissions of ethylene oxide at our plant, we are still looking for ways to lower these levels further. These are some of the plans we have under way to accomplish this: (provide specifics). As we implement these steps, we will keep you and the community informed of our progress. We will also continue to monitor our workers and keep track of health statistics within the community to ensure that the risks posed by our plant to our workers and to the community remain in the future as low as, if not lower than, they are today. Since some of you may have further questions about these and other matters concerning our plant operations, as plant manager, I am providing my work and home phone numbers so you can call me. I will do my best to supply you with answers to your questions as quickly as possible."

Statement 2

"Emissions of ethylene oxide from our plant are half the levels permitted by the U.S. Environmental Protection Agency and by our state's Department of Environmental Protection."

"Emissions of ethylene oxide from our plant are five times lower than the U.S. Environmental Protection Agency's safety standard."

"Plant emissions of ethylene oxide are five times below what was permitted under the old EPA standard, and two times below the level established by the new, stricter EPA standard.

Statement 3

"Laboratory studies on rats and mice suggest that current exposure to ethylene oxide may cause seven cancers in 1000 generations of residents in this city. This estimate is the maximum that would occur under worst-case conditions. Actual health effects from exposure to ethylene oxide are likely to be lower."

"Let me try to put this number into the context of other numbers. We've said that our worst case prediction is seven thousandths of one extra cancer within the next 70 years from our plant's emissions of ethylene oxide. Now, no one ever gets seven thousandths of a cancer. A better way to see the effect is that if 130 different communities the same size as Evanston had a plant just like this one, 129 of those towns would see no effect on their cancer rate. One of the 130 Evanstons might have a single extra cancer."

"Our best estimate of the risk is 0.001 cancers per 3500 persons using what we believe are realistic assumptions. This estimate is based on work done by our own scientists and by researchers at Evanston University. However, you should be aware that the state Department of Environmental Protection (DEP) has calculated a worst-case risk estimate of 0.007 cancers per 3500 persons. DEP made the assumption that all individuals living in Evanston would be expressed to emissions of ethylene oxide 24 hours a day for 70 years. This formula gave DEP a human-lifetime dose. DEP then took the best available laboratory information for ethylene oxide—data obtained from studies on the laboratory mice most likely to develop cancer in response to ethylene oxide—and calculated first the lowest dose that caused adverse health effects in mice and then the equivalent dose in humans. On the basis of these and other pieces of information, DEP concluded that the maximum cancer risk to people in the community is 0.007 cancers per 3500 persons over 70 years."

"Our worst-case estimate of the risk is seven thousandths of a cancer per 3500 persons over the next 70 years. How sure are we that the risk is really this low? The bad news is that we're not as sure as we'd like to be. Risk assessment is a pretty new science, based on models and assumptions rather than hard data. The good

news is that we're almost certain the risk is actually smaller than our estimate—we've instructed our scientists to make every assumption on the cautious side, to provide an extra margin of safety. And here's a piece of hard information. We've been manufacturing ethylene oxide in Evanston for 35 years now. We have continually monitored our employees for signs of adverse health effects associated with exposure to ethylene oxide. In all that time, as far as we know, not a single worker or retiree has had the sort of cancer normally associated with ethylene oxide. Please keep in mind that these workers are exposed to consistently higher levels of emissions than the surrounding population is. Therefore, on the basis of our workers' experience so far, the risk is zero. There are also people who think our risk estimate is too low. The Evanston chapter of the Sierra Club estimates seven hundedths of a cancer per 3500 persons over the next 70 years. That's 10 times higher than our estimate—but even if they're right, it's still an extremely small potential increase in the cancer rate. And we haven't found anyone with a higher estimate than theirs."

Statement 4

"If we buy and install the newest and most advanced emission-control equipment available, the worst-case situation is that the maximum total risk will be 0.005 additional cancers per 3500 persons, a very low number. If we don't buy new equipment and keep operating the plant with our current pollution-control system, the worst-case situation is that the maximum total risk will be 0.007 additional cancers per 2500 persons—also a very low number. Please keep in mind that both of these risk estimates are worst-case estimates."

Statement 5

"The maximum health risk from our plant's emissions of ethylene oxide is 0.007 additional cancers per 3500 persons. We could switch to producing the only known chemical substitute for ethylene oxide. However, the maximum health risk of emissions of that chemical is 50 times higher."

Statement 6

"We have installed in our plant the most advanced emission control system now operating in the country. Compared with those of older plants, such as the one in Middletown, our emissions are 10 times less."

Statement 7

"The risk posed by emissions of ethylene oxide is extremely low, no matter where you live or work in Evanston. However, the risk posed by emissions of ethylene oxide for people living two miles from the plant is 90% less, than for people living in the nearest home; and the risk for people living in the nearest home is 90% less than for people working within the plant gates. And our workers haven't had a single case of the type of cancer normally thought to be linked to ethylene oxide."

Statement 8

"Let me see whether these numbers will help. Roughly a quarter of all of us get cancer—a disease caused by smoking, diet, heredity, radon in the soil, pollution, and many other factors. Out of 3500 people, medical data show that one-quarter—or about 875—are going to get cancer sometime in a lifetime. So here's the predicted effect of ethylene oxide emissions from our plant on the overall cancer rate. In 129 of 130 hypothetical Evanstons, no effect—that is, no expected increase in cancer rates at all. In the 130th, cancer rates would rise from 875–876. Although this is only a tiny increased risk, it is still an increase. If we can find a way to make it even smaller, we should and we will. The most important thing is for all of us in Evanston to work together to find ways to bring down the total cancer rate, that unfortunate 875 out of 3500. But we at our plant have a special responsibility to be safe neighbors. Much higher risks due to other factors are no reason to ignore a small risk in our facility. Here's what we're doing to make sure we keep the risk from our plant as low as it can possibly get: (provide details)."

Statement 9

"During the next year, our plant will spend more than $2 million to reduce our already small emissions even further. This new investment will hurt us economically but will reduce the risk of cancer in the community by more than 25% when fully operational."

Statement 10

"If we stopped producing ethylene oxide today, many more people here and throughout the United States might die than could possibly be affected by emissions from our Evanston plant. Ethylene oxide is the best ster-

ilizing agent used by hospitals today. No equivalent substitute for ethylene oxide is available. Continued production of this production will contribute to saving many lives and will ensure that the surgical instruments that doctors and hospitals use are free from infectious agents."

Statement 11

"One way to look at the data is to compare the risks of emissions of ethylene oxide to plant neighbors with the risks to plant employees. We have been operating this plant for 35 years, with an average employment of 400 people. We therefore have about 10,000 person-years of worker exposure to ethylene oxide at this plant. Health monitoring at our plant indicates that the average workplace concentration of ethylene oxide is 0.5 ppm, a dose 200 times higher than that in the community. The primary health concern about ethylene oxide is its potential for causing certain types of brain cancer. We have not had a single case of brain cancer in our work force. Moreover, the overall incidence of cancer in our employees is lower than that of the U.S. population as a whole. Nor has Evanston's health department documented any brain cancers among our workers. On the basis of this information, I believe that the health risk posed by the plant to the community is insignificant."

Statement 12

"I believe that our ethylene oxide emissions do not pose a significant health risk to the community. I also believe that our emissions pose a much less serious problem than our hazardous waste problem, which is daily becoming more serious because the repositories in our state are filled and none are being built."

Statement 13

"One way to look at the cancer risk from emissions of ethylene oxide in our community is to compare the risk with the cancer risk from the x-rays you get during a health checkup. One chest x-ray per year presents a risk of developing cancer that is twice that of developing cancer from our plant's emissions of ethylene oxide."

Statement 14

"Another way to get some perspective on the risk of ethylene oxide emissions is by comparing it to some of the risks that we all face in our daily lives, such as the risk of being killed by lightning or the risk of beng killed in an auto accident. My purpose in making such a comparison is only to put the size of the risk in context. I recognize that such comparisons are like comparing apples and oranges. Still, I think the comparison can help us all understand and gain some perspective on the size of the risk we are talking about. For example, the risk of death by salmonella food poisoning from poultry bought at the local supermarket is at least five times greater than the risk of cancer from the highest exposure to ethylene oxide in this community."

"You may be wondering, 'But what does that mean to me as a resident of this community? What's the risk to me and my family?' First let me tell you that I am convinced that there is no threat to the health or safety of any member of our community at these extremely low exposure level. However, I recognize that the data still may be troubling. So it would probably be helpful to put these levels of risk from exposure to ethylene oxide into the context of other risks that we're all exposed to in our daily lives. For example, the risk to the average American of death from lightning is at least 140 times greater than the risk of cancer in Evanston from the highest exposure to ethylene oxide. Hurricanes and tornadoes also pose a risk about 140 times greater. Insect bites pose a risk about 70 times greater. The additional 0.007 cancer risk is about the same as the additional cancer risk you would incur spending four hours in Denver rather than at sea level because of Denver's high altitude and higher radiation level."

ACKNOWLEDGMENT

We thank C. Atman, C. Cortes, G. Hester, R. Lio, I. Nair, and P. Steranchak for their assistance in this work. We received helpful comments on a previous draft of the manuscript from G. Hester, J. Merz, D. Resendiz, P. Sandman, and P. Slovic. The work was supported by National Science Foundation grant SES-871564. The views expressed as those of the authors.

REFERENCES

1. V. T. Covello, P. M. Sandman, and P. Slovic, *Risk Communication, Risk Statistics, and Risk Comparisons: A Manual for Plant*

Managers (Washington, D.C., Chemical Manufacturers Association, 1988).
2. B. Cohen and I. S. Lee,. "A Catalog of Risks," *Health Physics* **36**, 707–722 (1979).
3. R. Wilson, "Analyzing the Risks of Everyday Life," *Technology Review* **81** 40–46 (1979).
4. Environmental Protection Agency, *A Citizen's Guide to Radon: What It Is and What to Do About It, 13 pp.* (Washington, D.C. 1986), 13 pp.
5. B. Fischhoff, P. Slovic, and S. Lichtenstein, "Weighing the Risks," *Environment* **21** 17–20, 32–38, (1979).
6. B. Fischhoff, S. Lichtentein, P. Slovic, S.L. Derby and R.L. Keeney, *Acceptable Risk* (New York, Cambridge University Press, 1981).
7. National Research Council, *Improving Risk Communication* (Washington, D.C., The Council, 1989).
8. B. Fischhoff, S. Watson, and C. Hope, "Defining Risk," *Policy Sciences* **17**, 123–139 (1984).
9. A. Agresti, "Tutorial on Modeling Ordered Categorical Response Data," *Psychological Bulletin* **105** 290–301 (1989).
10. R. D. McKelvey, and W. Zaviona, "A Statistical Model for the Analysis of Ordinal Level Dependent Variables," *Journal of Mathematical Sociology* **4** 103–120.
11. Greene, W. H. *LIMDEP* (Self-published statistical software package and manual, 1985).
12. P. Slovic, N. N. Kraus, and V. T. Covello, "Comment: What Should We Know About Making Risk Comparisons," *Risk Analysis* in press (1990).
13. B. Fischhoff, P. Slovic, S. Lichtenstein, S. Read, and B. Combs, "How Safe Is Safe Enough? A Psychometric Study of Attitudes Towards Technological Risks and Benefits," *Policy Sciences* **8**, 127–152 (1978).
14. P. Slovic, "Perception of Risk," *Science* **336**, 280–285 (1987).
15. B. Fischhoff, P. Slovic, and S. Lichtenstein, "Fault Trees: Sensitivity of Assessed Failure Probabilities to Problem Representation," *Journal of Experimental Psychology: Human Perception and Performance* **4**, 330–344 (1978).
16. A. Tversky, and D. Kahneman, "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology* **5**: 207–232 (1973).
17. B. Fischhoff, P. Slovic, and S. Lichtenstein, in T. Wallsten (ed.), *Cognitive Processes in Choice and Decision Behavior* (Hillsdale, New Jersey, Erlbaum, 1980).