Risk Analysis, Vol. 3, No. 4, 1983

Judged Lethality: How Much People Seem to Know Depends Upon How They Are Asked

Baruch Fischhoff and Don MacGregor[1]

Received January 26, 1983; revised September 17, 1983

Four formally equivalent response modes were used to elicit laypeople's beliefs regarding the lethality of various potential causes of death. Results showed that respondents had an articulated core of beliefs about lethality that yielded similar orderings of maladies by lethality regardless of the response mode used. Moreover, this subjective ordering was fairly similar to that revealed by public health statistics. However, the absolute estimates of lethality produced by the different response modes varied enormously. Depending upon the mode used, respondents were seen to greatly overestimate or greatly underestimate lethality. The implications of these discrepancies for public education and risk analysis are explored.

KEY WORDS: Risk perception; judgment; risk assessment; elicitation.

1. INTRODUCTION

A recurrent question in the management of hazardous technologies is "How well does the public understand them?" Different answers can point to rather different roles for the public in hazard management. A well-informed public can be trusted to use technologies wisely, fend for itself in the marketplace, and identify its best interest in political decisions. An ignorant public may need protection from regulatory agencies, help to grasp political questions, or special training and safeguards to prevent misuse of potentially dangerous machines and substances.

At first blush, assessing the public's knowledge would seem quite straightforward. Just ask questions like: What is the probability of a nuclear-core meltdown? How many people die annually from asbestos-related diseases? and How does wearing a seat belt affect your probability of living through the year? The responses can be compared with the best available technical estimates, and deviations can be interpreted as showing the extent of the respondents' ignorance. This straightforward (jusk-ask-them) strategy is clearly superior to relying on speculation or anecdotal evidence.

There are, however, a number of constraints on it. A first constraint on questioning is that the questions address pertinent topics.[1] Laypeople have no way of knowing the answers to questions that concern classified, proprietary, or otherwise unpublished information. There is no reason (other than curiosity) for them to know facts that cannot affect their behavior. A second constraint is that the question be clear.[2,3] Jargon must be avoided, as must terms such as "risk," that seem clear but are used differently by different people.[4,5]

Our concern here is with a third constraint, one that remains even with questions that are worth asking and wording that is clear. It is the need to request knowledge in a form that is compatible with people's customary way of thinking about the topic. To acquit themselves properly in an interview, people must be able to express what they know. If the mental representation of their knowledge is different from the formulation required by the interviewer,

[1]Decision Research, A Branch of Perceptronics, 1201 Oak Street, Eugene, Oregon 97401.

0272-4332/83/1200-0229$03.00/1

then some translation is necessary, first to retrieve what they know and, second, to express what they retrieve. The greater the incompatibility, the more cumbersome the translation process becomes and the more knowledge is lost in transmission.

As a concrete example of possible difficulties, consider a group of (somewhat morbid) individuals who conscientiously read the obituaries in their local newspaper and have perfect recall. They are asked by an interviewer to estimate the relative frequency of different causes of death (or the age distribution of deaths) in their community. Although the respondents have all the requisite knowledge, in order to satisfy the interviewer they must aggregate it into the particular summary categories requested and perform the needed mental arithmetic in the time allotted.

One solution to the compatibility problem is *convergent validation*, eliciting judgments in several ways and trusting only patterns that emerge however the question is posed.(6) Although methodologically valid, convergent validation is a conservative strategy. It ignores many data and evades the compatibility problem by taking a position neither on how knowledge is represented in people's minds, nor on how best to extract it. A more direct approach is developed here within the specific context of eliciting judgments of the lethality of potential causes of death. This method builds upon convergent validation to identify core knowledge, which emerges however questions are posed. However, it also provides enough insight into the mental representation of knowledge to make some informed guesses about what method is best when discrepancies are observed.

2. THE STUDY

Although "risk" can be (and often is) spoken of as a uniquely defined, unitary concept, it clearly is not.(7) There are many different aspects of risk(8, 9, 10) and various ways to measure each.(11, 12) One aspect of risk with an important influence on people's attitudes towards technological hazards is its degree of "lethality," the likelihood that if something goes wrong it will prove fatal.(5, 8, 13, 14) All other things being equal, more lethal problems are viewed as more "risky" and in need of stricter regulation.

The present experiments consider lay estimates of the lethality in the U.S. of the 20 potential causes of death appearing in Table I. As a standard of comparison, the right-hand column offers statistical estimates derived from public health statistics. Although used as a standard, these statistics are not infallible. Poor sampling, incomplete reporting, and inconsistent attribution of multiply-caused deaths are some of the problems that make this a comparison between lay estimates and technical estimates (rather than between "real" and "perceived" risk).

The lay estimates here were elicited by four formally equivalent response modes; exemplary versions of which are:

(a) *Estimate death rate*: In a normal year, for each 100,000 people who have influenza, how many people do you think die of influenza?

(b) *Estimate number died*: Last year, 80,000,000 people had influenza. How many of them do you think died of it?

(c) *Estimate survival rate*: In a normal year, for each person who dies of influenza, how many do you think have influenza but do not die of it during the year?

(d) *Estimate number survived*: In a normal year, 5,000 people die of influenza. How many people do you think have influenza, but do not die from it during the year?

The formal equivalence of these four questions carries no assurance of their psychological equivalence. Each requires respondents to approach, translate, and express what they know in a somewhat different way. To the extent that the four questions elicit consistent estimates, one can conclude that respondents have a core of knowledge about lethality that is equally accessible from all four perspectives, and whose translation into a numerical response poses no problem. Conversely, inconsistent responses reveal the differential compatibility between response modes and knowledge representation.

Some potentially significant differences among the response modes are: (a) the death rate and survival rate conditions called for estimates of rates, whereas the number died and number survived conditions called for estimates of numbers; (b) those two conditions provided some information (which did not "give the answer away," but might have confirmed or contradicted existing beliefs); (c) the death rate and number died conditions dealt with fatalities, whereas the survival rate and number survived conditions dealt with survivors; (d) the correct answers for the number survived condition were generally much larger numbers than for the death rate, number died, and survival rate conditions (the medians were 3,000,000; 80; 5,500; and 1,250, respectively).

Table I. Direct and Converted Lethality Rate Estimates Based on Geometric Mean Responses

	Death rate per 100,000 afflicted				
Malady	Estimated death rate[a]	Estimated number died	Estimated survival rate	Estimated number survived	Statistical death rate
Dental problems	10	1	2	1	1
Influenza	393	6	26	511	6
Mumps	44	114	19	4	12
Skin diseases	63	4	6	641	30
Asthma	155	12	14	599	33
Alcoholism	559	70	13	294	44
Venereal disease	91	63	8	111	50
Measles	52	187	18	28	75
High blood pressure	535	89	17	538	76
Drug abuse	1,020	1,371	19	95	80
Bronchitis	162	19	43	2,111	85
Pregnancy	67	24	13	787	250
Diabetes	487	101	52	5,666	800
Emphysema	1,153	1,998	70	5,417	1,423
Tuberculosis	852	1,783	188	8,520	1,535
Pneumonia	563	304	77	9,553	1,733
Automobile accidents	6,195	3,272	31	6,813	2,500
Strokes	11,011	4,648	181	24,758	11,765
Heart attacks	13,011	3,666	131	27,477	16,250
Cancer	10,889	10,475	160	21,749	37,500
Coefficient of concordance	.62	.67	.34	.67	
N	40	38	40	40	

[a]Only these rates were estimated directly. Participants in other groups estimated other quantities, which were converted to lethality rates as described in the text.

3. EXPERIMENT 1

3.1. Method

One hundred and fifty-eight individuals were recruited through an advertisement in a university newspaper and paid for participating in this and several other unrelated studies of judgment and decision making. They were evenly divided between men (median age = 24) and women (median age = 21). The task was described in written instructions that provided some pertinent risk statistics, including the overall lethality rate for the U.S. (expressed in the terms of the ensuing questions). The 20 questions were then presented in a single randomized order.

All responses were converted to a common response mode, death rate per 100,000, to facilitate comparisons. Individual subjects' converted responses were summarized by geometric, rather than arithmetic, means so as to reduce the influence of outliers.

3.2. Results

The bottom row in Table I presents coefficients of concordance for each group. This statistic measures the degree of agreement among subjects within a group, with regard to the *ranking* of maladies by judged lethality. It ranges from 1.0 representing total agreement to 0.0 meaning lack of any agreement. As can be seen, there was fairly high agreement within the death rate, number died, and number survived groups, but rather low agreement within the survival rate group. This suggests that individuals from this population have fairly similar ideas regarding the relative lethality of these maladies, but that this consensus cannot express itself in the survival rate response mode.

The body of Table I presents the geometric means of the derived death rates. The four columns differ markedly in the magnitude of the numbers they include. These differences provide a clear ordering of the response modes by the magnitude of the estimates they produce, with number survived estimates being greatest followed by death rate and survival rate estimates. In extreme cases (e.g., cancer, strokes), estimates produced by the different methods range over two orders of magnitude. Despite these discrepancies in absolute estimates, there was general agreement regarding the relative lethality of these 20 maladies. Rank correlations between the entries in Table I ranged from .72 to .83 (all statistically significant; $p < .001$).

The similarity of the survival rate ordering to those of the other groups, despite the large differences in absolute values, is further evidence that this mode was incompatible with subjects' natural mode of thought. Expressing their core of knowledge in this form required a translation process that took much effort and added noise to subjects' judgments. That noise was reduced agreement among individuals, as seen in the low coefficient of concordance. However, such random errors cancelled out when subjects' responses were aggregated.

In a correlational sense, all response modes produced judgments that were closely related to the statistical estimates. Rank correlations between geometric mean estimates and the statistical estimates ranged from .82 (survival rate) to .86 (number survived). As Table I shows, however, these high correlations obscure substantial differences in the accuracy of the actual estimates. In general, the statistical death rates fell in the middle of the four sets of estimated rates. Thus, whether these individuals tended to over- or under-estimate lethality depends upon how the question was asked.

One measure of accuracy is an *error factor*, equal to the ratio of the estimated rate for a malady to the statistical rate, when the former is larger, or the reciprocal of that ratio, when the latter is larger. When computed over all individual responses, the geometric mean error factor for survival rate subjects was 33.2. By contrast, subjects in the other groups were, on the average, off by only a factor of 10 or so (see Table III, bottom).

4. EXPERIMENT 2

Apparently, people have a core of knowledge regarding relative lethality that emerges however they are queried. Moreover, the ordering roughly matches that provided by public health statistics. Both the magnitude and the reliability of their responses are, however, quite sensitive to the precise response mode used, with the survival rate question producing particularly low and unstable responses. Before interpreting these results in too great detail, it is worthwhile establishing how robust they are and clarifying the psychological processes involved in them. Experiment 2 attempts to do that by repeating and elaborating the tasks of Experiment 1.

4.1. Method

One hundred forty-three individuals repeated the tasks of Experiment 1 with a subset of 10 of the maladies for which public health statistics seemed most trustworthy—thereby allowing subjects to focus on few items. There were 37 subjects in the death rate group, 36 in the number died group, 37 in the survival rate group, and 36 in the number survived group. After answering, subjects were given the correct values for each item. In order to encourage attention to those values, they scored their own answers as too high or too low. After an hour of unrelated tasks, they were unexpectedly asked to recall the true value. Arguably, the best recall and the greatest improvement in knowledge will be with the most natural representation, that mode most conducive to the integration and preservation of additional knowledge. Finally, they saw the lethality of infectious hepatitis expressed in each of the four modes. They rated those phrasings by how "natural" they seemed, and how closely each "corresponds to the way you usually think of the lethality of diseases and accidents." An additional 87 subjects performed only this rating task.

4.2. Results

As shown in Table II, the initial estimates here resembled those from Experiment 1 (presented in Table I). Across the four groups, 26 of the 40 geometric mean estimates were within a factor of 2 of the comparable estimates from Experiment 1; all 40 were within a factor of 5. Again, the coefficients of concordance showed considerable agreement among subjects within each group except survival rate. Again, the overall orderings of the maladies within the different response modes were similar to one another and to the statistical estimates. Again, the statistical estimates fell below some group estimates and above others.

Table II. Initial and Recalled Lethality Rates: Experiment 2 (Geometric Means)

	Estimated death rate		Estimated number died		Estimated survival rate		Estimated number survived		Statistical rate[a]
	Initial	Recall	Initial	Recall	Initial	Recall	Initial	Recall	
Influenza	136	4	11	10	140	36	284	370	6
Asthma	59	49	12	35	33	397	858	115	30
Measles	57	57	401	407	67	321	61	37	75
Pregnancy	57	115	25	124	20	299	549	444	250
Diabetes	287	344	436	374	54	579	8,435	2,236	800
Emphysema	1,503	902	1,008	751	277	787	8,658	4,475	1,400
Tuberculosis	650	462	4,346	4,563	310	882	11,057	1,115	1,500
Pneumonia	482	352	392	156	199	854	9,279	9,580	1,700
Stroke	3,745	3,153	4,045	3,823	380	3,655	19,072	22,919	12,000
Cancer	6,110	12,106	9,211	8,433	327	7,388	17,526	33,128	37,000
Coefficient of concordance	.63	.58	.64	.66	.35	.33	.71	.80	
Rank correlation with statistical rate	.64	.87	.73	.64	.56	.78	.78	.78	

[a] Rates are given to subjects and are rounded to two significant figures.

After receiving the true values, subjects scored their own estimates as being too high or too low. One measure of the attention they paid is that there were only 47 errors in 1,480 scoring opportunities (= 3.2%).

The top section of Table III shows that in the unexpected recall task subjects infrequently remembered the statistical values that they had been given. The memory rate for individual maladies showed a serial position effect. The highest rates were for the first and last items (36.1%, influenza; 48.3%, pregnancy). The two worst remembered were fourth and sixth (0.6%, emphysema; 1.7%, tuberculosis). Personal relevance had some contribution to memorability insofar as cancer had the third best memory rate (22.4%) despite being fifth on the list. The second row of that table shows that when subjects failed to remember the correct value, they seldom supplanted it with their own initial estimate. Thus the two estimates were distinct enough in subjects' minds not to be confused.

The lower section of Table III shows that for all response modes, subjects' recollections were more accurate than their initial estimates. Thus, although subjects did not remember the statistical estimates, they did learn something from them. This learning was most pronounced with the survival rate group, whose recall estimates were, in the aggregate, as accurate as those of the other groups. Provision of the correct values seems to have enabled subjects to translate their ordinal knowledge of lethality into much more accurate numerical estimates for this response mode.

Eighty-seven "fresh" subjects rated the naturalness of the different modes for expressing information about the lethality of infectious hepatitis. Clearly, these subjects thought it more natural to think about lethality in terms of death than in terms of survival. There was no difference in preferences for statistic (rate or number). The rankings of the subjects who had previously completed the estimation and recall tasks were quite similar. Overall, mean rankings decreased by an average of 0.24 for subjects who had used a phrasing. Thus, although naturalness judgments are quite robust, they can be affected by immediate experience.

5. GENERAL DISCUSSION

In the aggregate, these results indicate that people have a fairly robust and consensual subjective ordering regarding the lethality of this set of maladies. The same ordering emerges with response modes sufficiently different to yield very different absolute estimates. This consistency means that it is possible to look at the substance of the lethality rankings regarding which maladies' relative lethality is overestimated or underestimated, although we will not do so.

Table III. Contrast between Original Estimates and Recall of True Values (Ten Items of Experiment 2)

	Estimated death rate	Estimated number died	Estimated survival rate	Estimated number survived
	Percentage of cases			
Recall = true value	19.3	15.3	21.3	10.4
Recall = initial estimate	4.1	4.4	3.0	5.2
	Geometric mean error factor			
Experiment 1 initial	10.9	10.2	33.2	12.5
Experiment 2 initial	12.5	10.7	43.0	12.1
Experiment 2 recall	4.2	6.5	7.6	9.2

That core of beliefs is not, however, as readily translated into all of the formally equivalent numerical expressions, as evidenced by differences in accuracy, within group agreement and naturalness ratings. The survival rate mode is clearly the outlier among these methods. It produced the least agreement among subjects and the worst absolute estimates. These results indicate a marked incompatibility between that response mode and subjects' customary ways of thinking about lethality. When respondents attempted to bridge that gap by themselves the result was noisy and biased responses. Along with number survived this mode was also rated least natural. Nonetheless, subjects were still able to exploit evidence presented in this mode, as shown by their vastly improved recall estimates. Thus, it appears harder to get information out of people with this mode than it is to get information into them.

Several simple accounts for these discrepancies in absolute judgments prove inadequate: (a) the *availability* explanation would argue that people are unduly influenced by the factors that are made most salient to them.[15] That should produce higher estimates of lethality with the response modes focused on death than with those focused on survival. However, the two survival response modes produced the largest and smallest lethality rates. (b) A *statistic* explanation would argue that the summary measure, a rate or numerical estimate, somehow affected performance. However, no such tendency was observed. (c) The same evidence would also reject a *storage mode* explanation: If people organize their information on a case-by-case basis, then the translation to a rate should be problematic; the converse would be true if subjects organized their knowledge in terms of rates. Yet, neither rates nor numbers were systematically higher or lower, more or less accurate, or more or less natural. (d) The number response modes provided some *additional information* (either the death toll or the affliction toll). In itself, that was not enough to improve performance consistently. (e) A *large number* explanation would argue that subjects have difficulty with response modes that require very large numbers,[16] which they are unaccustomed to using in daily life. For example, the number survived group was required to produce the largest numbers. Inability to do so would mean underestimating the number of survivors and emerge as overestimation of the lethality rate, the result obtained. The other groups, however, were required to produce numbers in a similar range, but showed quite different systematic biases. (f) An *anchoring and adjustment* explanation holds that respondents make quantitative estimates by picking some initially relevant number as a starting point and then adjusting it to accommodate additional information. In practice, that adjustment tends to be inadequate, turning the starting point into an anchor.[17] Unfortunately, the application of this heuristic with present tasks is unclear without independent knowledge of how people choose anchors. For example, was the number died group anchored on the total number of deaths, the number of deaths per 100,000 people in the U.S., the number of survivors, the number of deaths from accidents or from violent causes (all of which appeared on their form), or some other number(s) of their own creation?

Thus, none of these single factor explanations can account for the differences in the size of the magnitude estimates. Each might, of course, be "saved" if one could make an exception for one group or another. The most legitimate exception would be the survival rate group. If it is excluded, most of these explanations would prove quite serviceable, suggesting that each tells something about how people process such information.

5.1. Implications

The stable ordinal judgments observed here replicate the basic pattern observed in Lichtenstein's *et al.*[6] multi-method study of fatality judgments and Slovic's *et al.*[8] multi-method study of risk judgments. People have a consistent and fairly accurate feeling for the relative threat posed by different hazards. Where ordinal knowledge is all that is required, any response mode is good enough. However, if absolute estimates are needed, the methods matter greatly. People might respond quite differently to a threat if they assess its lethality by thinking about the survival rate or the number of survivors. A public health official could conclude that people underestimate or overestimate lethality, depending upon the question asked.

Our overall appraisal of the evidence produced by this multi-method approach suggests that the death rate and number died response modes provide the two best expressions of people's beliefs about lethality. They produce reliable and similar estimates; moreover, they are both judged to be quite natural. If this summary is correct, then it can be said that there is little systematic bias in people's lethality estimates.

We believe that some such multi-method analysis is essential before interpreting the responses produced with any response mode. The convergence found here is not assured. People might have had no coherent core of knowledge, knowing instead different things about death rates, survival rates, numbers died, and numbers survived. Responses to four such response modes would then tell four different stories. Assessing what people know would require evoking each perspective. Educators might be required to use several perspectives in order to ensure that people get the picture.

A needed extension of these methods is to the elicitation of information from technical experts in the context of risk analyses.[18, 19] For example, a supervisor might be asked how frequently workers fail to follow a particular operating procedure; an atmospheric chemist might be asked to assess a cumulative probability distribution for the oxidation rate in some complex situation; a mechanical engineer might be asked to estimate the failure rate for a familiar valve in an unfamiliar use. Such questions may be formulated for the convenience of the consumer of that knowledge (the risk analyst) or its producer (the technical expert). However, being an expert in a topic need not mean being an expert in answering questions about it. In that case, all formally equivalent questions are not psychologically equivalent. Question design may be as important an aspect of risk analysis as system modeling.

ACKNOWLEDGMENTS

This research was partially supported by the Advanced Research Project Agency of the Department of Defense under Contract N00014-79-312-0730 (ARPA Order No. 0194) under subcontract from Decisions and Designs, Inc. and by the National Science Foundation under Grant SES 8213452 to Perceptronics, Inc. Any opinions, findings and conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. B. Fischhoff, P. Slovic, and S. Lichtenstein, "The 'Public' vs. the 'Experts': Perceived vs. Actual Disagreements About the Risks of Nuclear Power," in V. Covello, G. Flamm, J. Rodericks, and R. Tardiff (eds.), *Analysis of Actual vs. Perceived Risks* (Plenum Press, New York, 1983).
2. National Academy of Sciences, *Survey Measure of Subjective Phenomena* (NAS, Washington, D. C., 1981).
3. S. L. Payne, *The Art of Asking Questions* (Princeton University Press, Princeton, New Jersey, 1980).
4. B. Fischhoff, P. Slovic, and S. Lichtenstein, "Lay Foibles and Expert Fables in Judgments about Risk," *American Statistician* **36**, 240–255 (1982).
5. P. Slovic, B. Fischhoff, and S. Lichtenstein, "Rating the Risks," *Environment* **21**, 14–20, 36–39 (1979).
6. S. Lichtenstein, P. Slovic, B. Fischhoff, M. Layman, and B. Combs, "Judged Frequency of Lethal Events," *Journal of Experimental Psychology: Human Learning and Memory* **4**, 551–578 (1978).
7. B. Fischhoff, S. Lichtenstein, P. Slovic, S. L. Derby, and R. L. Keeney, *Acceptable Risk* (Cambridge University Press, New York, 1981).
8. P. Slovic, B. Fischhoff, and S. Lichtenstein, "Characterizing Perceived Risk," in R. W. Kates and C. Hohenemser (eds.), *Technological Hazard Management* (Oelgeschlager, Gunn & Hain, Cambridge, Massachusetts, 1983).
9. C. Vlek and P. J. Stallen, "Judging Risks and Benefits in the Small and in the Large," *Organizational Behavior and Human Performance* **28**, 235–271 (1981).
10. D. von Winterfeldt, R. S. John, and K. Borcherding, "Cognitive Components of Risk Ratings," *Risk Analysis* **1**, 277–288 (1981).
11. E. A. C. Crouch and R. Wilson, *Risk/Benefit Analysis* (Ballinger, Cambridge, Massachusetts, 1982).

12. S. R. Watson, "On Risks and Acceptability," *Journal of the Radiological Protection Society* **1**, 21–25 (1981).
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. C. H. Green, "Risk: Attitudes and Beliefs," in D. V. Canter (ed.), *Behaviour in Fires* (Wiley, Chichester, 1982).
15. A. Tversky and D. Kahneman, "Availability: A Heuristic for Judging Frequency and Probability," *Cognitive Psychology* **4**, 207–232 (1973).
16. E. C. Poulton, "Biases in Quantitative Judgments," *Applied Ergonomics* **13**, 31–42 (1982).
17. A. Tversky and D. Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science* **185**, 1124–1131 (1974).
18. M. G. Morgan, M. Henrion, and S. C. Morris, *Expert Judgment for Policy Analysis* (BNL 51358) (Brookhaven National Laboratory, Upton, New York, 1980).
19. U. S. Nuclear Regulatory Commission, *Fault Tree Guide* (NRC, Washington, D. C., 1981).