Risk Analysis, Vol. 9, No. 3, 1989

Are Large Differences in "Lifesaving" Costs Justified? A Psychometric Study of the Relative Value Placed on Preventing Deaths

John M. Mendeloff[1] and Robert M. Kaplan[2]

Received October 13, 1988; revised March 23, 1989

Government actions to reduce risks to health have varied greatly in their cost per death prevented, frequently by 10-fold or even 100-fold. This research asks whether disparities of this magnitude are justified by citizens' preferences abut the relative value of reducing deaths from different hazards. Four samples were asked to rank the relative priority of preventing deaths through 8 realistic programs, each addressed to a different hazard, and then to rate how large the differences in spending should be. Subjects were not asked to give absolute values on preventing deaths and were asked only for their relative valuation of the benefits of preventing a death, not to weigh the benefits and costs or to determine an optimal spending level. We found that in all samples the median respondent valued his top-rated program 5 to 6 times more than his bottom-rated program. However, because individuals disagreed upon the relative priority for different programs, the aggregated rankings barely showed more than a 2-fold difference in the amounts that should be spent. Thus, for the important programs considered by these samples, a large variation in spending does not appear to be justified on the basis of differentials in the values placed on preventing different types of deaths. A more deliberative methodology like the one used here appears fruitful for providing insights to policymakers about preferences in this sensitive area.

KEY WORDS: Benefit estimation; value of life; psychometric studies.

1. INTRODUCTION

Analyses of safety and public health programs have revealed great disparities in the cost per fatality prevented. For example, many OSHA and EPA health standards can be justified only by valuing each death prevented at millions and sometimes tens of millions of dollars. In contrast, many state highway departments appear to use valuations of only several hundred thousand dollars per death.[1–4] These differences have led many economists to argue that funds should be reallocated so that the marginal cost of death prevention is equal across programs, thus maximizing the number of deaths prevented for a given outlay.[1,5]

The logic of this argument depends, in part, upon the view that the prevention of all deaths is equally worthwhile. While many would acknowledge that adjustments should be made for the years of life added, and perhaps their quality, the more critical issue concerns the context in which the risk arises. Some assert that context makes little difference. Thus, Bailey[1] writes:

> Although there is no rational case for spending a huge sum to avoid a death from one cause while refusing to spend a relatively small sum to avoid a death from another cause, it can be shown that rational, well-informed citizens do not equalize

[1] School of Public Health, State University of New York at Albany, New York State Department of Health, Tower 2523, Empire State Plaza, Albany, New York 12237.
[2] School of Medicine, M-022, University of California, San Diego, La Jolla, California 92093.

0272-4332/89/0900-0349$06.00/1

> these incremental sums precisely in private choices. Hence a policy based on such choices will allow some, albeit minor, differences in these sums to remain.

Graham and Vaupel[2] present another view:

> ...policymakers, for numerous legitimate reasons, may explicitly decide to devote more resources to saving lives in some areas than in others. For example, some causes of death are particularly painful and anxiety-producing. To give another example, some causes of death may seem especially "unfair" since they result from largely involuntary exposure to, say, carcinogens in the air rather than from more voluntary factors such as cigarette smoking.

Yet, they acknowledge that the issue remains: "How much more is it reasonable to spend in some areas than in others? Do the huge disparities in lifesaving expenditures reflect defensible judgments?"

Many different research strategies have provided some insights into these questions. The most prominent have been studies of willingness to pay (WTP) for risk reduction and studies of "acceptable risk" via expressed preferences. Both have shortcomings for normative purposes. Studies of WTP that are based upon revealed behavior[6–9] (e.g., accepting risky jobs or buying safer products) are limited to the valuation of those directly at risk; caring by third parties is difficult to include. Also, the conditions required for a study of WTP are met in only a limited set of risk contexts (e.g., labor markets, where most studies of WTP have been conducted). Thus the issue of whether WTP differs greatly in different contexts is difficult to study. Some WTP[10–15] studies have used surveys of expressed preferences rather than studies of revealed behavior as the basis for estimates. These studies frequently find anomalies that seem related to subjects' discomfort with the use of dollar figures and unfamiliarity in dealing with small probabilities.

The literature[16,17] examining what people judge to be "acceptable" levels of risk has generated many insights, but these studies have required respondents to perform very difficult tasks requiring high levels of information. The judgement about what level of risk is "acceptable" can depend upon many factors, including not only the value placed upon a given reduction in risk, but also the cost of achieving that reduction. When the costs are quite subjective (e.g., the foregone liberty to drive a motorcycle without a helmet), an assessment of preferences remains important. Yet often the costs are resources, labor and materials, whose valuation is more straightforward. In those cases, since policymakers can estimate costs much better than average citizens can, it would make more sense to directly examine citizen preferences regarding the benefits of risk reduction.

Although all of the methods cited above make contributions, there is a need for supplementary approaches. Three choices guided the approach presented here. The first was to elicit information only about how people value a given benefit, not about the optimal budget for a program or the acceptable or optimal level or risk. The rationale was to minimize the cognitive effort and to focus on the issue where preferences matter most. The second choice was to elicit judgments about the *relative* values placed on preventing different types of deaths, but not about the absolute value of preventing those deaths. Previous studies that have elicited absolute values from respondents have not produced consistent results. The third choice was to ask question about programs that were real and important. The rationale was to enhance the usefulness of the findings for policymakers. The major objective of this research is to develop a method that will prove useful to policymakers.

2. METHOD

Four groups of subjects participated in the project. The first group included 190 undergraduate students from San Diego State University (SDSU). In exchange for their participation, the students received credit toward a requirement in an introductory psychology course. The second group consisted of 35 students enrolled in a course on health and safety policy taught by one of the authors (JM) at the University of California, San Diego (UCSD). Because of their participation in the class, the UCSD students were more familiar with issues in health resource allocation. The third group consisted of 18 retirees who were participating in the University of California's Institute of Continued Learning (hereafter called ELDERLY). Almost all of this group were over 60 years old. The fourth group included 38 UCSD employees (the "STAFF" sample) who worked in various clerical and administrative jobs.

The basic instrument instructed the respondent:

> We would like you to play the role of a policymaker who is faced with difficult choices among competing programs. We are interested in your views about the priorities that should be given to different programs for preventing fatalities. Assume that we can prevent 10 deaths in any of the programs listed on the next page [see Table I], but that we only have enough money to carry out some of the programs. Which of the programs, if any, do you think deserve priority over the others? Should we value the prevention of some deaths differently than others and thus be willing to spend more to prevent them? Or do you think that the prevention of all deaths should be valued equally and thus that the amounts we are willing to spend to prevent additional deaths should be the same for all programs?

Respondents were first asked to rank each of the

Table I. Program Description Used for All Four Samples

One program prevents a fatality by enforcing standards to reduce workplace exposures to cancer causing chemicals. The deaths would have occurred 30 years in the future. The average age of the victims is 65.
One program prevents a fatality by creating bicycle lanes to separate and protect cyclists from cars. The deaths prevented would have occurred this year. The average age of victims is 20.
One program prevents a fatality by enforcing regulations requiring that the slats on cribs be close enough together to keep a baby from getting its head caught and strangling. The deaths prevented would have occurred this year. The average age of victims is 1.
One program prevents a fatality by removing roadside obstacles (e.g., trees, boulders) along dangerous curves in order to protect motorists who drive off the road. The deaths prevented would have occurred this year. The average age of victims is 27.
One program prevents a fatality by increasing the number of paramedics and ambulances so that heart attack victims receive help more quickly. The deaths prevented would have occurred this year. The average age of victims is 65.
One program prevents a fatality by enforcing standards to reduce air pollutants that especially affect people with lung diseases like emphysema and bronchitis. The deaths prevented would have occured in 10 years. The average age of victims is 75.
One program prevents a fatality by building median barriers on roads to separate traffic and prevent head-on collisions. The deaths prevented would have occurred this year. The average age of victims is 27.
One program prevents a fatality by enforcing standards to prevent falls by workers at construction sites. The deaths prevented would have occurred this year. The average age of victims is 40.

programs from lowest to highest priority. Then they were asked to give a rating of 10 to the lowest ranked program and rate others in relation to that program (e.g., if they thought it merited spending twice as much per fatality prevented, it should get a rating of 20). The instructions emphasized that subjects should not question the presumed effectiveness of each program, and that they should assume that the costs of each program were the same. "The only issue we want you to think about is whether you value the prevention of 10 deaths in one program more than the prevention of 10 deaths in another."

Respondents were also asked to rate each program on three scales: (1) whether they thought that people, by their own actions, can reduce the risk from the hazard; (2) the extent to which they personally faced a risk from the hazard; and (3) the extent to which family or friends faced a risk from the hazard.

2.1. Variations

Several variations on the basic format were employed. Students in the SDSU sample were randomly assigned to five subgroups, the first of which followed the instructions above. The second group was first orally presented a brief "Discussion of Issues" (see Appendix), which reviewed the reasons why one might or might not choose to value the reduction of deaths from different programs by different amounts. The third group was asked to rate the highest ranked program as 100 and rate the others down from that level, rather than rating the lowest a 10 and moving up. The fourth group was also asked to use the 100-down approach, but was first presented the "Discussion of Issues." The fifth group heard the discussion of issues, used the 10-up approach, but was also given information on the average age of the victims and the latency period for the type of death potentially prevented by the program. The latency period was the length of time between the intervention and the benefit. Table I shows the eight programs with this information added.

The UCSD, ELDERLY, and STAFF samples were subjected to the same treatment as this fifth group at SDSU.

3. RESULTS

3.1. Variability in Each Individual's Ratings

For each subject, a ratio of highest to lowest ranked program was created. The results in Table II are for individuals. The median difference between the top and bottom ranked program were very similar in the four samples, ranging from five- to six-fold. Thus for half of the individuals, the maximum difference between top-rated and bottom-rated programs was less than that amount; and for half it was larger. Some members of both student samples valued their top ranked program 50 or more times higher than the lowest ranked; in the ELDERLY sample, the maximum difference was 10-fold.

Table II. Variations in Individual Maximums Within Samples[a]

ELDERLY		UCSD		STAFF		SDSU	
N	Max	*N*	Max	*N*	Max	*N*	Max
5	20	2	12	1	10	1	12
1	25	1	14	1	12	1	17
1	35	1	15	1	13	5	20
1	40	1	18	1	15	1	22
6	50	1	25	1	18	1	26
2	60	7	40	1	20	4	30
2	100	3	50	4	30	2	35
18	median = 50	1	53	4	40	3	40
		1	60	7	50	1	45
		2	70	4	60	5	50
		5	100	1	80	1	55
		1	120	1	90	1	60
		1	140	5	100	1	65
		1	200	1	140	1	75
		1	220	2	200	4	80
		1	300	1	400	1	90
		2	500	2	500	10	100
		1	525	38	median = 50	1	102
		1	600			1	150
		1	998			1	200
		35	median = 60			1	205
						1	350
						2	500
							(1 missing)
						50	median = 55 or 60

[a]Subjects gave their lowest ranked program a rating of 10. Thus, for example, in the ELDERLY sample, 5 subjects gave their top ranked program a rating of 20, implying that they valued preventing deaths in that program twice as highly as those in the lowest ranked. Two subjects in that sample gave their top ranked program a rating of 100, implying that the difference in valuation was 10-fold.

3.2. Variability Among Aggregated Ratings

Another way to review the data is to aggregate preferences across individuals (Table III). When we do this, the differences in ratings of programs shrink. Instead of five- or sixfold differences, the maximum differences in two of the samples are less than twofold. In a third, they were less than threefold. The largest difference in mean ratings within a sample was found in the UCSD sample, where the prevention of crib deaths got an average rating almost 4.5 times larger than the prevention of deaths from air pollution (108.6 vs. 24.3).

Because individuals used different scales in their ratings—some going above 500, others staying below 20—it is useful to use ratios to represent ratings of each program in relation to one another. There were 28 ratio comparisons [$(8 \times 7)/2$] in each of the four samples. The median ratios of these comparisons in each sample are shown in Table IV.

Thus, for the SDSU sample, the median ratio of the value of preventing cancer deaths to the value of preventing deaths by building bike lanes was 0.85. The median respondent thought we should be willing to spend only 85% as much to prevent cancer deaths as we should be willing to spend to prevent deaths to cyclists. In contrast, the median ratio comparing cancer deaths to deaths due to roadside obstacles was 1.38, implying that the median individual thought that the value of preventing the cancer deaths was 38% greater. In only 8 of these 112 comparisons did the median ratio show a difference of two-fold or greater (i.e., greater than 2.0 or less than 0.5). The largest difference is 3.5-fold (this was in the UCSD sample for crib deaths over pollution deaths).

Table IV shows the cut-off points for the upper and lower quartiles. For the bike lane comparison, one quarter of the SDSU respondents thought that the prevention of cancer deaths justified spending no more than half as much as bike deaths. However, another quarter thought that cancer deaths justified spending at least 50% more than bike deaths.

A review of these quartiles shows that a major reason why there is little variation among aggregated ratings

Table III. Mean and Median Ratings in Four Samples[a]

Program	STAFF Mean	STAFF Median	ELDERLY Mean	ELDERLY Median	UCSD Mean	UCSD Median	SDSU Mean	SDSU Median	Average rank on means
Workplace	50.6^{2} (57.0)	30	27.3^{3} (20.0)	18	25.6^{7} (23.1)	15	47.0^{4} (80.9)	20	4
Bike lanes	26.9^{7} (19.8)	20	21.0^{8} (10.7)	20	64.9^{3} (110.6)	30	48.7^{2} (63.2)	30	5
Cribs	53.6^{1} (98.1)	25.5	25.4^{4} (14.0)	20	108.6^{1} (195.3)	50	55.8^{1} (83.7)	31	1.75
Roadside obstacles	20.3^{8} (30.8)	12	23.0^{5} (20.1)	15	38.3^{6} (59.7)	20	20.0^{8} (11.6)	18	6.75
Paramedic	49.7^{3} (101.7)	20	22.3^{6} (10.7)	20	51.5^{4} (96.1)	22	48.0^{3} (46.6)	25	4
Pollution	45.7^{4} (48.1)	30	28.3^{2} (25.3)	20	24.3^{8} (26.3)	11	40.1^{5} (75.0)	25	4.75
Median barriers	41.3^{5} (67.2)	20	36.4^{1} (24.6)	29	81.7^{2} (121.6)	40	38.4^{6} (40.8)	20	3.75
Construction falls	27.4^{6} (36.7)	18	21.4^{7} (11.0)	20	46.5^{5} (69.8)	22	34.1^{7} (55.4)	20	6.25

[a]Standard deviations are in parentheses. Superscripts are the rankings of each program within the sample. The last column is calculated by adding the rankings in the four samples and dividing by 4.

is that individuals often disagree about which programs deserve the higher ranking. When 50% think that Program X ranks higher than Program Y and the other 50% think the opposite, then the aggregate judgment will tend to show them valued almost equally. In only three cases out of 112 is the 1st quartile ratio greater than 1.00. In only five cases out of 112 is the 4th quartile ratio less than 1.00. Thus in only eight cases out of 112 did even three quarters or more of the respondents in a sample agree whether a particular program rated higher than another one. If we use a slightly less stringent measure, whether three quarters or more agreed that a particular program rated *as high or higher* than another, an additional 22 cases meet that test. The comparison between roadside obstacles (D) and median barriers (B) was the only one to meet that criterion in all four samples. The comparison between roadside obstacles and bike lanes (B) met it in three samples.

3.3. Substantive Differences in Ratings

Table III shows that the deaths prevented by the crib standard received the highest mean rating in all but the ELDERLY sample. If we take the rankings of its mean ratings, add them across the four samples, and divide by four, we get an average overall ranking of 1.75 [(1+1+1+4)/4], easily the highest of any program. At the low end, roadside obstacles (6.75) and construction falls (6.25) had the lowest overall average rankings. The five other programs all clustered between 3.75 and 5.00.

The two major differences in ratings among the samples pertained to cancer and pollution and to bike lanes. The STAFF sample gave high ratings to cancer and pollution deaths; the UCSD sample gave them very low ratings, and the SDSU and ELDERLY samples were in-between. Although the difference was less striking, the two student samples gave relatively high rankings to deaths prevented by bike lanes compared to the ELDERLY and STAFF samples.

The differences in the mean ratings of programs within a sample were often not statistically significant. Table V presents the results only for those differences that were significant at the 0.10 level (two-tailed). In only 1 of the 28 program comparisons was there a significant difference in all four samples: deaths prevented by median barriers were always valued more highly than deaths prevented by removal of roadside obstacles. In two other comparisons—cribs rated over roadside obstacles and over construction falls—the difference was significant in three of the samples. (In the ELDERLY sample, crib deaths were rated more highly, but the differences were not statistically significant.) In 10 comparisons, the

Table IV. Medians and Quartiles for the Ratios of Program Ratings[a]

(*N* = 51) Programs compared	SDSU 1st quartile	SDSU Median	SDSU 4th quartile	UCSD 1st quartile	UCSD Median	UCSD 4th quartile
A/B	0.50	0.85	1.50	0.405	0.60	1.20
A/C	0.40	0.89	1.90	0.20	0.46	0.94
A/D	0.83	1.38	2.75	0.40	1.00	2.00
A/E	0.55	0.87	1.20	0.50	0.86	1.25
A/F	0.91	1.13	1.54	0.91	1.10	1.70
A/G	0.54	0.88	1.76	0.20	0.67	1.00
A/H	0.72	1.20	2.53	0.38	0.86	1.17
B/C	0.55	1.00	1.60	0.45	0.79	1.00
B/D	1.09	1.50	3.17	1.00	1.20	2.33
B/E	0.51	1.00	1.45	0.60	1.17	3.00
B/F	0.77	1.42	2.50	1.00	2.00	4.50
B/G	1.00	1.08	1.25	0.70	1.00	1.20
B/H	1.00	1.33	2.00	0.75	1.29	2.00
C/D	0.92	1.61	4.63	1.25	1.91	3.33
C/E	0.62	1.00	1.85	1.00	1.75	2.92
C/F	0.71	1.36	3.10	1.14	3.50	5.30
C/G	0.71	1.00	1.96	0.88	1.25	2.00
C/H	0.50	0.91	1.30	0.50	0.91	1.50
D/E	0.32	0.50	1.00	0.50	0.83	1.25
D/F	0.46	1.00	1.58	0.73	1.43	2.50
D/G	0.40	0.66	1.00	0.50	0.78	1.00
D/H	0.50	0.91	1.37	0.50	0.91	1.50
E/F	0.95	1.33	2.19	1.00	1.60	2.20
E/G	0.62	1.04	2.00	0.33	0.73	1.20
E/H	0.92	1.39	2.50	0.67	1.00	1.43
F/G	0.50	0.78	1.40	0.22	0.35	0.92
F/H	0.65	0.91	2.00	0.33	0.67	1.17
G/H	1.00	1.21	1.67	1.00	1.20	2.33

differences were statistically significant in two samples; however, in four of these cases the samples disagreed about which program rated higher. Thus, two or more samples agreed about statistically significant differences in nine comparisons, five of them showing the low rating for roadside obstacles.

In three comparisons, the differences in ratings between programs were never significant in any of the samples. These three comparisons were: (1) workplace cancer vs. lung disease from air pollution, (2) workplace cancer vs. paramedic programs for heart attack victims; and (3) bike deaths vs. heart attacks.

3.4. Substantive Differences Among Methods

When SDSU subjects were asked to rate programs from 100-down rather than from 10-up, the rank order remained very similar in the two groups, However, the ratings were more compressed in the 100-down responses. In the samples where subjects were given the "Discussion of Issued," the median overall ratings for the eight programs ranged from 50–10 in the "10-up" group and from 95–53 in the "100-down" group (see Fig. 1). In the samples where that discussion was not presented, the medians ranged from 37–13 and from 90–55, respectively. The only program where the ranking of the median shifted by more than one place was the crib program, which dropped from third highest to fifth highest when the "Discussion of Issues" was provided in the "10-up" sample.

More dramatic changes were produced by providing subjects with information about the average age of victims in the different programs and the latency period before the preventive effects would be felt in addition to the "Discussion of Issues." The median rating for workplace cancer fell from first to fifth place, and for median barriers, from second to fifth place. The ratings for the

Table IV. Continued

Programs Compared	Sample					
	ELDERLY			STAFF		
	1st quartile	Median	4th quartile	1st quartile	Median	4th quartile
A/B	0.63	0.86	1.75	0.80	1.39	2.73
A/C	0.55	0.88	1.33	0.75	1.00	2.00
A/D	0.71	1.10	2.75	1.00	2.00	4.00
A/E	0.69	0.89	1.58	0.83	1.45	3.00
A/F	0.64	1.00	1.50	1.00	1.00	1.33
A/G	0.45	0.63	1.07	1.00	1.23	2.50
A/H	0.60	1.00	2.00	1.00	1.55	3.00
B/C	0.60	0.89	1.31	0.50	0.96	1.15
B/D	0.79	1.00	1.67	1.00	1.67	2.25
B/E	0.67	1.00	1.50	0.67	1.00	2.00
B/F	0.58	0.90	1.75	0.40	0.96	1.50
B/G	0.35	0.75	1.17	0.50	1.00	1.67
B/H	0.79	1.00	1.34	0.61	1.00	2.50
C/D	0.75	1.29	1.64	1.00	2.00	3.33
C/E	0.65	1.20	1.69	0.67	1.00	2.17
C/F	0.48	1.00	2.00	0.44	1.00	1.67
C/G	0.58	1.00	1.15	0.67	1.27	2.00
C/H	0.73	1.00	1.62	0.83	1.38	2.60
D/E	0.50	0.97	0.42	0.33	0.92	1.33
D/F	0.48	1.00	1.54	0.30	0.50	1.00
D/G	0.46	0.78	0.90	0.50	0.83	1.00
D/H	0.71	0.92	1.24	0.50	1.00	1.40
E/F	0.65	1.08	1.58	0.38	0.83	1.10
E/G	0.48	0.67	0.96	0.61	0.96	1.50
E/H	0.63	1.14	1.55	0.72	1.10	2.00
F/G	0.39	0.70	1.36	0.75	1.13	2.67
F/H	0.58	1.25	2.00	1.00	1.58	3.33
G/H	0.98	1.30	2.63	0.75	1.20	1.67

[a]Program A is workplace cancer; B is bike lanes; C is crib deaths; D is roadside obstacles; E is paramedics for heart attack victims; F is pollution deaths; G is median barriers; H is falls in constuction.

The first row gives the ratio of the rating for workplace cancer to the rating for bike lanes. Thus in the SDSU samples, one fourth of the repsondents thought the value of preventing 10 deaths in the first program was only half as great or less than the value of preventing 10 deaths in the second. Half of the respondents thought that the value of the deaths prevented in the first program was 85% or less than the value of the deaths prevented in the second. Another one quarter of the respondents valued the prevention of cancer deaths at least 50% more than the prevention of deaths in the bike-lane program.

crib standard moved from fifth to first and bicycle lanes moved from sixth to second when age and latency information were given to the subjects.

3.5. Findings on Independent Variables

Table VI shows ratings from the four samples on the scales for "Self," "Other," and "Blame." The table suggests that there was considerable agreement on these scales. The "Self" variable measured whether the respondent thought that he or she personally faced the risk. The "Other" variable measured whether family or friends faced the risk. The "Blame" variable measured the extent to which they thought that those who died bore responsibility for their fate. Respondents gave their answers on a scale from 1 (least) to 7 (most.)

On "blame," all four samples rated roadside obstacles highest, three of the four groups rated pollution lowest. The fourth group rated pollution next to lowest. The only sharp difference was that the SDSU sample gave paramedics the second highest score while the others rated it sixth, seventh, or eighth. In addition, both student samples rated crib deaths fifth highest on blame while the older samples rated it second.

On "self", all four samples gave the highest rating to median barriers, the lowest to construction falls, and

Table V. Statistically Significant Differences in Mean Ratings of Programs[a]

Program comparison	ELDERLY	STAFF	UCSD	SDSU	No. of samples with statistical significance and the same sign
A vs. B (bikes)		0.026	−0.025	0.026	2
C (cribs)			−0.013		1
D (roadside)		0.00		0.025	2
E (paramedics)					0
F (pollution)					0
G (medians)			−0.006		1
H (falls)		0.043	−0.066		1
B vs. C		−0.076	−0.060		2
D				0.002	1
E					0
F		−0.039		0.020	1
G	−0.031				1
H				0.013	1
C vs. D		0.020	0.045	0.004	3
E			0.087		1
F			0.012		1
G	−0.085			0.094	1
H		0.058	0.022	0.000	3
D vs. E		−0.050		−0.000	2
F		−0.000		−0.072	2
G	−0.007	−0.004	−0.035	−0.001	4
H				−0.084	1
E vs. F			0.087		1
G	−0.037				1
H				0.062	1
F vs. G			−0.004		1
H		0.076	−0.040		1
G vs. H	0.025		0.026		2
N = 28	5	11	15	11	

In 1 case, the difference is significant in all four samples: D vs. G.
In 2 cases, the difference is significant in only three samples: C vs. D, C vs. H.
In 6 cases, the difference is significant in only two samples with the same sign.
In 16 cases, the difference is significant in only one sample with the same sign.
In 3 cases the differences are never significant: A vs. E, A vs. F, B vs. E.

[a]Only differences which were statistically significant at the 0.10 level (2 tails) are shown. This analysis is based on "t-tests" of the mean ratings in Table III.

the next to lowest to crib deaths. The only substantial differences were, as expected, that younger samples gave a relatively high rating to bike lanes, and the ELDERLY sample gave a relatively high rating to paramedics. Bike lanes ranked second for the two student samples, third for the STAFF, and sixth for the ELDERLY samples. Paramedics ranked second for the ELDERLY sample, but fifth or sixth for the others.

Rankings of the "Other" variable again show considerable agreement among samples. There are really no sharp differences. Indeed, the ratings for "Other" were almost exactly the same as for the "Self" variable. The only exception is that the rating for paramedics is higher in the three nonelderly samples than it is for the "Self" variable.

Table VII shows the correlation between the "Self" and "Other" variables in the four samples. The two programs addressing auto safety have very high correlations in all samples, reflecting the universality of auto use among the respondents; the pollution program is next highest. The ELDERLY sample also showed a high correlation for the paramedics program.

Table VII also shows the correlations between ratings on "Self" and "Blame." Except for the EL-

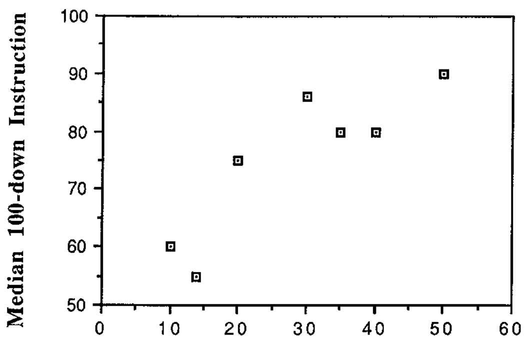

Fig. 1. Rating of programs using 10-up and 100-down instruction.

DERLY sample, three quarters of the correlations are negative. In these cases, the greater the extent to which a person faces a risk, the less likely he or she is to blame the victims. The main exception is the crib death program where the correlation is positive in all four samples. The program to prevent construction deaths is the only other with positive correlations in at least two samples. It should be noted, however, that the correlations between "Self" and "Blame" tend to be modest.

3.6. Findings: Explaining the Ratings

Several different approaches to explaining the ratings were attempted. These included:

1. A regression analysis of the "Self," "Other," and "Blame" variables on the levels of the ratings.
2. Because "Self" and "Other" were typically highly correlated, we ran regressions omitting one or the other.
3. We used each person's ratio of the rating for program X to the rating for program Y as the dependent variable, where the independent variables were the differences between the scores of the independent variables for those two programs. We used ratios for the dependent variable, rather than differences, to partially overcome

Table VI. Mean Scores for Ratings of "Blame," Affects "Self," and Affects "Others" in Four Subject Samples

ELDERLY	UCSD		SDSU		STAFF	
Blame						
5.53 D (roadside)	5.80	D	5.50	D	4.76	D
5.42 C (cribs)	4.77	B	5.16	E	4.53	C
5.26 B (bikes)	4.35	H	5.14	B	4.47	B
4.69 H (falls)	3.94	G	4.76	G	4.26	H
4.47 G (median)	3.89	C	4.62	C	3.45	G
4.32 A (cancer)	3.68	E	4.48	H	3.26	A
3.79 E (pollution)	3.26	A	3.96	A	3.24	F
3.74 F (paramedic)	3.15	F	3.62	F	3.05	E
Self						
5.26 G	5.15	G	4.58	G	5.13	G
4.53 E	4.54	B	4.08	B	4.03	D
3.74 D	4.03	D	3.98	D	3.58	B
3.00 F	2.56	F	3.34	G	3.16	F
1.47 A	2.34	A	3.20	E	2.74	E
1.42 B	2.26	E	2.08	A	2.34	A
1.31 C	1.34	C	2.06	C	1.29	C
1.00 H	1.21	H	1.42	H	1.18	H
Other						
5.32 G	5.18	G	5.16	E	5.24	G
4.84 E	4.83	B	4.74	G	4.55	E
4.00 D	4.74	E	4.70	G	4.47	B
3.89 F	4.23	D	4.60	D	4.45	D
3.42 B	4.11	C	4.28	F	3.84	F
3.32 C	3.38	F	4.18	C	3.29	C
2.21 A	2.88	A	3.08	A	2.66	A
1.63 H	2.21	H	2.52	H	2.61	H

Table VII. Correlations Between Independent Variables

		ELDERLY ($N=19$)	UCSD ($N=35$)	SDSU ($N=50$)	STAFF ($N=38$)
		"Self" and "Other"			
Cancer	A	0.68[a]	0.44[a]	0.58[a]	0.14
Bikes	B	0.46[b]	0.38[b]	0.45[a]	0.73[a]
Cribs	C	0.38[c]	0.12	0.47[a]	0.34[b]
Roadside	D	0.95[a]	0.94[a]	0.83[a]	0.86[a]
Paramedic	E	0.87[a]	0.21	0.55[a]	0.27[c]
Pollution	F	0.56[b]	0.82[a]	0.60[a]	0.79[a]
Medians	G	0.99[a]	0.96[a]	0.85[a]	0.89[a]
Falls	H	undefined	0.15	0.50[a]	0.36[b]
		"Self" and "Blame"			
Cancer	A	−0.24	−0.36[b]	+0.17	−0.05
Bikes	B	−0.21	−0.18	−0.22[c]	−0.25[c]
Cribs	C	+0.21	+0.12	+0.21[c]	+0.15[c]
Roadside	D	−0.01	−0.36[b]	−0.06	−0.24[c]
Paramedic	E	+0.26	−0.37[b]	−0.05	−0.08
Pollution	F	+0.12	−0.08	−0.05	−0.03
Medians	G	+0.24	−0.23[c]	−0.15	−0.13
Falls	H	undefined	+0.36[b]	+0.05	−0.05

[a]Statistically significant at 0.01 level (2 tails).
[b]Statistically significant at 0.05 level.
[c]Statistically significant at 0.1 level.

the problem posed by respondents' use of different maximums. Thus a rating of 80 for program X and 40 for program Y will have the same ratio as a rating of 20 for program X and 10 for program Y.

4. Finally, we performed regression analyses to explain the variation in ratings for individual subjects, rather than the variation in ratios across individuals within a sample. Here, there were only eight observations in each regression—the number of programs each individual rated. These analyses were directed toward understanding individual rather than aggregate variability.

The models used to explain variations across individuals were only modestly successful. The adjusted R^2 were rarely above 0.30 and the percentage of coefficients that were statistically significant at the 0.10 level never exceeded 50%. Space does not permit a detailed report of the findings; however, Table VIII reports the number of statistically significant coefficients for each of the variables in each of these different regressions. Several points are worth noting:

1. Because of the presence of multicollinearity between the "Self" and "Other" ratings, the statistical significance of those variables increase when one or the other was ommitted.
2. The "Other" variable tended to explain more of the variance than the "Self" variable, although the results varied with the sample and the specification.
3. With the strong exception of the STAFF sample, the "Blame" variable performed least well. It played no role in explaining ratings in the ELDERLY sample, and almost no role in most of the SDSU regressions.
4. The regression models did a better job of explaining ratings in some programs than in others. For example, the F-tests for bike paths and paramedics were significant at the 0.10 level in three of the four samples. In contrast, the F-tests for construction falls were never significant at the level.
5. The results for individuals do not differ too sharply from those for the samples as a whole. For example, Table IX shows the results in the ELDERLY sample of a regression of "Other" and "Blame" on a subject's ratings for the eight programs. It shows that 8 out of 17 individuals had significant coefficients for the "Other" variable and 2 for the "Blame" variable. With

Table VIII. Regressions Explaining Ratings: The Number of Statistically Significant Coefficients (at 0.10, 1-tail), Number of Statistically Significant Coefficients, and Number of Coefficients Estimated[a]

STAFF	SDSU	UCSD	ELDERLY	STAFF
a. Regression on ratings with self, blame, and other	6/24 Other 2/8 Blame 3/8 Self 1/8	2/24 Other 2/8 Blame 0/8 Self 0/8	5/24 Other 2/8 Blame 2/8 Self 1/8	5/24 Other 4/8 Self 1/8 Blame 0/8
b. Regression on ratings with self and blame	6/16 Self 2/8 Blame 4/8	4/16 Self 4/8 Blame 0/8	5/16 Self 3/8 Blame 2/8	4/16 Self 4/8 Blame 0/8
c. Regression on ratings with other and blame	5/16 Other 3/8 Blame 2/8	5/16 Other 4/8 Blame 1/8	7/16 Other 4/8 Blame 3/8	6/16 Other 6/8 Blame 0/8
d. Regressions on ratio of ratings with differences in self, other and blame	24/84 Other 3/28 Self 8/28 Blame 13/28	24/84 Other 8/28 Self 8/28 Blame 8/28	18/84 Other 10/28 Self 3/28 Blame 5/28	17/84 Other 12/28 Self 3/28 Blame 2/28

[a]Excluding constant.

Table IX. Regressions on Individuals in ELDERLY Sample[a]

	Other		Self	
	Coefficient	"t"	Coefficient	"t"
1	−0.43	(0.54)	−0.09	(0.11)
2	21.2	(1.965)[b]	13.53	(3.55)[c]
3	4.17	(2.72)[c]	9.61	(1.98)
4	3.86	(3.43)[c]	3.86	(3.43)[c]
5	4.74	(2.31)[c]	5.26	(2.76)[c]
6	3.20	(0.57)	1.35	(1.21)
7	3.14	(1.21)	3.14	(1.21)
8	Undefined		Undefined	
9	−0.67	(0.28)	1.02	(1.17)
10	9.18	(1.77)[b]	−5.15	(1.04)
11	2.26	(3.18)[c]	−1.06	(0.81)
12	0.96	(1.75)[b]	0.58	(1.12)
13	−0.04	(0.05)	−0.95	(0.30)
14	−4.04	(0.75)	−0.45	(0.17)
15	2.18	(1.25)	−1.20	(0.52)
16	1.67	(5.39)[d]	1.52	(3.66)[c]
17	3.42	(1.05)	1.13	(0.28)
18	0.15	(0.04)	−1.29	(0.53)

[a]This table presents results for each of the 18 individuals in the ELDERLY sample. To explain their ratings of the 8 programs, a regression was first run with the "other" and "blame" variables, then with the "self" and "blame" variables. Results are shown here only for the "other" and "self" variables. The "blame" variable was not significant for more than 2 individuals in either of the regressions.
[b]Indicates significance at the 0.10 level in a 1-tailed test.
[c]Indicates significance at the .05 level.
[d]Indicates significance at the 0.01 level.

"Self" instead of "Other", there were four individuals with significant coefficients

4. DISCUSSION

Two chief issues merit discussion: the validity of these findings, and the implications they have for policymaking. Normative guidance depends on several factors, including the representativeness of the sample and the thoughtfulness of the responses. If citizen preferences are to guide policymakers, it is important that these responses really do reflect preferences, and that some degree of deliberation underlies them. First, we review the strengths and weaknesses of our findings, suggesting several areas for improvements in future work. Then, we explore the normative implications if we assume that the basic findings are valid.

4.1. What Do the Findings Mean?

The most important finding of this research is when ratings of how much to spend to prevent a death in different programs are aggregated across the individuals within a sample, the differences between the top-rated and bottom-rated programs are rarely more than twofold. The credibility of this outcome is enhanced by its consistency with the two other studies that have asked similar types of questions.

In a 1979 study,[18] Slovic *et al.* investigated whether respondents believed that some deaths should be valued less than others. They found that the differences explained only a small part of the variation in the levels of risk that people judged "acceptable" for each activity. For their student sample, prevention of the least valued of 34 causes of death (alcoholic beverages) was valued about 2.5 times less than the highest valued (nuclear power). For their League of Women Voters sample, the difference was 4.4-fold (smoking vs. homicide).

Beggs[19] asked another small League of Women Voters sample (15 respondents) to rank programs. The subjects were to assume that the most worthwhile programs prevented 10 deaths, and then to choose how many additional deaths the less worthwhile programs would have to prevent to make them equivelant. For 61 programs, the median ratios varied by somewhat less than a factor of 6. The lowest rated programs included deaths due to "drug overdoses by drug addicts." For programs similar to the ones used in this paper, the variability was less than twofold.

Our results also suggest that the respondents did take the task seriously. Although the model we use to explain the ratings has only limited success, it suggests that the factors predicted to influence the ratings do frequently have the expected effect. Also, consistent with at least a modicum of thoughtfulness are the ratings on the independent variables. The samples usually agree on these factors, and when they disagree, the differences are generally in the expected direction. For example, the student samples say they face greater risks from bike riding than do the older samples.

Many studies[20,21] have shown that different methods of asking questions about preferences can elicit inconsistent responses. Because people may not fully understand the policy implications of the values they state, if they are given policy options to choose among, they may choose options that are inconsistent with their stated values for lifesaving. We believe that our introduction set the policy context clearly enough and our rating method was unambiguous enough, so that the prospect of major inconsistencies is limited. However, no tests were conducted to see if this optimistic conclusion is warranted.

Despite our optimism on this score, we nevertheless believe that our results fall well short of allowing confidence that the preferences of the respondents were accurately assessed. Fischhoff and colleagues[22] have pointed out that:

> . . . People do not have orderly preferences regarding many sets of alternatives that are both diverse and important. Rather, it seems that they have pieces of preferences, in the form of basic values that are strongly held, but not integrated into a coherent perspective. If they are forced to make an evaluative judgment, they must engage in an exercise in inference, deducing the implications of their related beliefs. The inferential process is most likely to produce reliable conclusions when individuals have the opportunity for thoughtful rumination over the issues, consultation with informed others, and hands-on experience with the alternatives (and their consequences), to serve as a check for the conclusions that they derive intellectually.

In this regard, it is striking that the one difference in ratings that was statistically significant in every sample was that between the program for removing roadside obstacles and the program for installing median barriers. Among the eight programs, these two differed in the fewest dimensions. And, because people are familiar with these risks, their assessments of those dimensions will tend to be informed by a common core of knowledge. The chief reason why the differences were statistically significant in every sample was that no more than 12% of the respondents in any sample thought that removing roadside obstacles should rank higher than adding median barriers.

This finding suggests that the lack of agreement characteristic of so many rankings occurred, in part, because individuals were responding to quite different "pieces of preferences" about each program. If that is true, then a more deliberative process should lead to a more comprehensive elicitation. People would not simply address those concerns which occurred to them first, but would be forced to try to put the "pieces" together.

It is plausible that such a process would generate more agreement about which features are the most appropriate to consider, and, therefore, to more agreement in rankings and ultimately to more variation in ratings. Even with full awareness of all of the issues raised by these ratings decisions, difference in ratings would, of course, remain because of differences in value placed on different factors (e.g., responsibility and latency).

The problems caused by lack of deliberation are exacerbated by programs that are especially ambiguous. For example, in the case of the paramedic program for heart attack victims, it was clear from discussions after the survey was administered that some respondents viewed the victims as blameless, while other assumed that the victims were to blame for an unhealthy lifestyle. The STAFF sample viewed heart attack victims as the least responsible of all, while the SDSU sample viewed them as the second most responsible. In two of the four samples (STAFF and UCSD), the ratings for this program had the highest coefficient of variation. In both of those samples, the "Blame" variable was statistically significant in explaining variations in the ratings of that program. We recognize that the issue of whether a set of victims is blameworthy is an ethical as well as a factual question. However, we think that further discussion could remove some sources of disagreement.

The limited ability of our models to explain the variations in ratings attests at least as much to the inadequacy of our understanding as it does to the possible irrationality of the respondents' responses. Nevertheless, we would have more confidence in the deliberativeness of the responses if we were able to explain them better with variables that seem as if they should play a role in decisions. The explanatory function would have been better served if our instrument had included enough programs to allow more satisfactory regressions on each individual's ratings. However, we thought that adding programs would threaten the individual's ability to keep them in his or her head and compare them meaningfully. We placed a higher priority on trying to elicit meaningful ratings than on contributing to an explanation of the ratings.

In addition to concerns a bout the lability of preferences, the choice of a metric is itself problematic. Our data suggest that ratios in relation to a top value of 100 were different than those rated in relation to a bottom value of 10. This lack of symmetry indicates that the responses cannot be characterized by a true ratio scale. If the ratios change with the scale, then we obviously can say little about whether program X is truly valued three times as much as program Y.

Another point deserving mention is the choice of our independent variables. Beggs used two variables to explain the valuation in his sample: how responsible the victim was for his fate and whether the respondent, himself, faced the risk.[19] In order to allow more explicitly for concern for third parties, we added a variable ("Other") for whether family or friends faced the risk. The "Self" and "Other" variables were often so highly correlated that it was not possible to distinguish their independent contributions. When the correlation was low (e.g., in the student samples for the paramedic program to prevent heart attack deaths), the "Other" variable usually accounted for more of the variance in ratings. In general, the "Other" variable slightly outperformed the "Self" variable. Even if this difference were much greater, however, we recognize that there is a difference between altruism that extends to family and friends and altruism that extends to total strangers. The issue of self-interested behavior is quite important for our results because, to the extent that people give the highest ratings to reducing the risks that they personally face (or indeed that their friends and relations face), two consequences appear: (1) the aggregated ratings will tend to vary with the number of people facing a risk; and (2) the results will change if we pick samples in which different risks are more prevalent.

Thus, for example, it is plausible that the relative rating for preventing deaths in construction would be higher in a sample of construction workers. Moreover, for most of the programs we found a negative correlation between the ratings on "Self" and the ratings on "Blame". This finding suggests that people are less likely to view the victim as responsible for his fate when they face the same hazard themselves. If true, then the impact of the "Self" variable is larger than its coefficient suggests: part of its impact is being picked up in the "Blame" variable. However, we should note that the magnitude of these correlations is not large and they are only sometimes statistically significant. Overall, we cannot draw a firm conclusion about the relative role of altruism and self-regard in these ratings.

The exceptions to the negative correlation between "Self" and "Blame"—the crib program and the two workplace programs—are worth noting. With cribs, we can infer that those who, in fact, do have responsibility for caring for infants are more likely to believe that they would be blameworthy if something happened.

That interpretation is consistent with the findings across samples, where the two student samples, whose members are much less likely to have had children, gave the crib program their fifth highest "Blame" score, compared to the second highest for the two older samples. The positive correlations in the job cancer and construction falls programs suggest that those individuals who actually take such risky jobs perceive these risks as more voluntarily undertaken. Such a conclusion is perhaps not surprising since so few of the respondents in our white-collar, largely middle-class samples do face them.

Although our research does not support very large differences in spending per death prevented, it also indicates that not all deaths are valued the same. Thus a policy of maximizing the valuation that citizens place on preventing deaths could lead to different priorities than a policy of maximizing the number of deaths prevented. For example, California highway safety engineers report(23) frequent public demands for installing median barriers while their own calculation show that removing roadside obstacles would be more cost-effective in preventing fatalities. The findings here give support for weighting deaths prevented by median barriers more heavily.

The policy implications of our findings are also limited by the recognition that social judgments about the value of a risk-reduction program are not a function solely of the expected health effects and sometimes not of the health effects at all. This distinction is made clearly by Slovic *et al.*(18), who demonstrated that differences in valuations on the deaths prevented accounted for little of the variation in levels of acceptable risk. Public fears about nuclear reactors, for example, appear to be driven by perceptions of the worst possible outcomes rather than by any assessment of the expected number of deaths resulting from their use. From the nuclear industry's viewpoint as well, the value placed on preventing a radioactive release that killed someone could be enormous, not because that death is somehow more worth preventing, but rather because that event could help to trigger a shutdown of the entire industry(24). To the extent that policymakers desire to maximize the value placed upon risk reduction, rather than the public health impact alone, they might choose to spend many times more to prevent a death from this cause than from others.

Despite these caveats, we believe that a clearer sense of the value citizens place upon the health consequences is a useful datum for policymakers to consider. Valuation decisions are inevitably subjective: they need not be arbitrary. The methodology employed here appears to be a fruitful one for exploring the relative value placed on reducing different risks.

5. APPENDIX

5.1. Discussion of Issues

(Presented orally to respondents before they answered the survey.) Some people may feel that the prevention of all deaths should be valued equally. Others may not. First, I am going to present some arguments for valuing the prevention of some deaths differently than others. Then I am going to make some arguments for valuing all deaths the same.

The argument for valuing deaths differently is that characteristics of the individual, or of the situation in which the hazard is encountered, affect us differently.

For example, regarding the individual, the average age of the victims may be a factor. Some programs may prevent the deaths of children or young adults, while others may affect people who are very old. If you believe that public policy should be concerned with the number of years of life added rather than merely the number of deaths prevented, then age should be considered and programs might be valued differently. One reason to favor the young is the notion that everyone should get at least a certain minimum number of years and that there is a basic unfairness about early death as well as a tragic quality due to the cutting off of a life in full bloom.

Another possible factor affecting how people value preventing a death may be the immediacy of the death. Some hazards, like pollution, often cause disease only after many years of exposure. Thus the deaths that the current spending will prevent may not occur for 10, 20, or even 30 years. We can argue that it is better to spend the money to prevent deaths now. Who knows? In 30 years they may have found a cure to the diseases that pollution causes.

The nature of the situation in which the risk is encountered may matter too. For example, some risks (e.g., in playing sports) are accepted voluntarily. Of course, if you are not even aware that a risk exists, you don't even have the choice of accepting it. Some risks are part of the requirements of everyday living (e.g., driving a car), but some of those risks are probably more controllable than others. In some situations the people who get hurt are ones who were careless; in other cases, there is really nothing you can do to protect yourself. We might choose to make greater efforts to protect people from risks that they are uninformed about, didn't accept voluntarily, or are unable to control through their own actions. The reason is that we want to help people who are conscientious more than we want to help people who are careless, both because they deserve it more and because we want to encourage people to be conscientious.

What is the argument *against* different values for the deaths in the different programs? The chief one is that a death is a death. Just because one person is old and another young does not mean that society should value the second one more. Each person's life should be considered equally valuable regardless of age or other characteristics. The same goes for the issue of *when* the deaths occur. Just because the deaths prevented by one program occur a number of years in the future is no reason to treat them as less valuable than deaths prevented now. People in the future are just as valuable as people now.

As for the different contexts—like whether the person knew about the hazard and whether they were partly at fault for their own death—it is important to keep in mind that all of us make careless mistakes at some times. We should also note that if we decide to spend a lot more money to prevent some types of deaths than others, the result will be that we won't be able to prevent as many deaths as we could if we spent equal amounts for each death prevented.

ACKNOWLEDGMENTS

We would like to acknowledge the helpful advice received from Paul Slovic and the excellent research assistance given by Tim Speth, Debra Cruet, and Eduardo Silva. This research was supported by NSF Grant SES-8420410. Dr. Kaplan was also supported by NIH Grant K04 HL 00809.

REFERENCES

1. M. J. Bailey, *Reducing Risks to Life: Measurement of the Benefits*, (American Enterprise Institute, Washington, D.C.,1980).
2. J. D. Graham and J. W. Vaupel. "The Value of Life: What Difference Does It Make?" *Risk Analysis* **1**, 89–95 (1981).
3. J. F. Morrall III, "A Review of the Record," *Regulation*, **10**, 25–34 (1986).
4. C. C., Travis, S. R., Pack, and A. Fisher, "Cost-Effectiveness as a Factor in Cancer Risk Management," *Environmental International* **13**, 469–474 (1987).
5. R. Zeckhauser and D. Shepard, "Where Now for Valuing Lives?" *Law and Contemporary Problems* **40** 6–35 (1976).
6. R. Dardis, "The Value of Life: New Evidence from the Labor Market," *American Economic Review* **50**, 1077–1082 (1980).
7. D. Ghosh, D. Lees, and W. Seal, "Optimal Motorway Speed and Some Valuations of Time and Life," *Manchester School Economic and Social Studies* **43**, 134–143 (1975).
8. A. Marin and G. Psacharopoulos, "The Reward for Risk in the Labor Market: Evidence from the United Kingdom and Reconciliation with Other Studies," *Journal of Political Economy* **90**, 827–853 (1982).
9. W. Kip Viscusi, "Labor Market Valuations of Life and Limb: Empirical Evidence and Policy Implications," *Public Policy* **26**, 339–386 (1978).
10. J. P. Acton, *Evaluating Public Programs to Save Lives: The Case of Heart Attacks*, (The Rand Corporation, Santa Monica, 1973).
11. M. W. Jones-Lee, M. Hammerton, and P. R. Philips, "The Value of Safety: Results of a National Sample Survey," *Economic Journal* **95**, 49–72 (1985).
12. W. Kip Viscusi and C. O'Connor, "Adaptive Responses to Chemical Labeling: Are Workers Bayesian Decision Makers?" *American Economic Review* **74**, 942–956 (1984).
13. W. Kip Viscusi and W. A. Magat, *Learning about Risk: Consumer and Worker Responses to Hazard Information* (Harvard University Press, Cambridge, 1987).
14. W. Kip Viscusi, W. A. Magat, and A. Forest, "Altruistic and Private Valuations of Risk Reduction," *Journal of Policy Analysis and Management* **7**, 227–245 (1988).
15. A. Muller and T. J. Reutzel, "Willingness to Pay for Reduction in Fatality Risk," *American Journal of Public Health*, **74**, 808–812 (1984).
16. B. Fischhoff, P. Slovic, and S. Lichtenstein, "How Safe is Safe Enough? A Psychometric Study of Attitudes Toward Technical Risks and Benefits," *Policy Sciences* **9**, 127–152 (1978).
17. B. Fischhoff, S. Lichtenstein, P. Slovic, S. L. Derby, and R. L. Keeney, *Acceptable Risk* (Cambridge University Press, Cambridge, 1981).
18. P. Slovic, S. Lichtenstein, and B. Fischhoff, "Images of Disaster: Perception and Acceptance of Risks from Nuclear Power," in G. Goodman and W. Rowe (eds.), *Energy Risk Management* (Academic Press, London, 1979), pp. 223–245.
19. S. D. Beggs, "Diverse Risks and the Relative Worth of Government Health and Safety Programs: An Experimental Survey," *U.S. Environmental Protection Agency, Environmental Benefits Analysis Series* (EPA Report No. 230-04-85-005, June 1984).
20. A. Tversky, S. Sattath, and P. Slovic, "Contingent Weighting in Judgment and Choice," *Psychological Review* **95**, 371–384 (1988).
21. D. Kahneman and A. Tversky, "Choices, Values and frames," *American Psychologist* **39**, 341–350 (1983).
22. B. Fischhoff, R. Gregory, and L. Furby, "Evaluating Voluntary Risks of Injury," prepared for *Conference on Integrating Perspectives on Injury Prevention*, sponsored by the Houston-Galveston Injury Prevention Group and the School of Public Health, University of Texas Health Science Center at Houston, October 1–3, 1984.
23. J. Mendeloff, "Measuring Elusive Benefits: On the Value of Health," *Journal of Health Politics, Policy, and Law* (Fall 1983).
24. P. Slovic, "Perception of Risk," *Science* **236**, 280–285 (1987).