

# Probabilistic Connotations of Carcinogen Hazard Classifications: Analysis of Survey Data for Anchoring Effects

Sarah E. Spedden<sup>1,2</sup> and P. Barry Ryan<sup>1</sup>

Received July 25, 1991; revised March 30, 1992

---

A study was undertaken to test for anchoring effects when numerical probabilities were elicited for carcinogen hazard designations. Subjects were asked to assign probabilities to both *probable* and *possible*. The sequence was randomly varied so that half of the subjects evaluated *probable* first and half evaluated *possible* first. While there was no consensus on the numerical probabilistic meanings assigned to these verbal expressions, in general, *probable* was assigned a higher probability than *possible* and there were specific values that were assigned frequently, indicating some consistency in interpretation. There appeared to be a fairly constant scaling factor between the probabilities assigned to the designations. Anchoring was manifested in two ways: a smaller difference between the designations when they were evaluated in sequence than when they were evaluated in isolation, and assignment of readily accessible "benchmark" values such as 50% and 75%.

---

**KEY WORDS:** Hazard identification; carcinogen classification; verbal probability expression; probable human carcinogen; possible human carcinogen.

## 1. INTRODUCTION

When research and regulatory agencies evaluate chemicals for human carcinogenic potential, these chemicals are classified based on the degree to which the evidence is judged to be conclusive. The classifications are designated by terms such as *known*, *probable*, and *possible*, which denote levels of uncertainty.<sup>(1-3)</sup> This process is referred to as hazard identification and is one of four steps in the carcinogen risk assessment process. Since hazard designations are more accessible and less technical than either carcinogenic potency or exposure data, they play an important role in risk communication. A preliminary survey was conducted to elicit probabilistic meanings and to test for anchoring effects when re-

spondents are asked to assess both *probable* and *possible* as designations of carcinogen hazard.

## 2. PRIOR LITERATURE AND HYPOTHESES

Teigen<sup>(4)</sup> has identified five dimensions that are critical to the way statements of uncertainty are interpreted. The first of these dimensions is the probability level. In the context of communicating carcinogen hazard, the probability level may reflect either the likelihood that cancer *would* result from exposure or the mechanistic plausibility that cancer *could*, under the right conditions, be induced by the chemical. The second of Teigen's dimensions is the seriousness of the outcome. That the outcome is cancer rather than something less dreaded can result in a context effect on the interpretation of carcinogen hazard designations. The effect of context is demonstrated by Kraus *et al.*<sup>(5)</sup> in their study

<sup>1</sup> Exposure Assessment and Engineering Program, Department of Environmental Health, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115.

<sup>2</sup> To whom all correspondence should be addressed.

of intuitive toxicology in which they show that the public's concern about exposure to toxic and carcinogenic substances is insensitive to dose. Teigen's third dimension is whether the source of the uncertainty is incomplete knowledge or stochastic variability. In classifying animal carcinogens for human carcinogenicity, uncertainty results from incomplete knowledge of how animal and other data can be extrapolated to humans when human data are lacking. However, there are also elements of stochastic variability in carcinogenesis. The fourth dimension Teigen describes is the degree of noncommitment. *Probable* and *possible* can be used by communicators to convey vagueness and lack of consensus on other dimensions such as probability level and degree of uncertainty. The final of Teigen's dimensions is whether the term draws attention to a *positive possibility* or to the *lack of complete certainty*. To illustrate this distinction, terms such as *it is not quite certain* or *it is somewhat doubtful* (that a particular chemical is a human carcinogen) would draw attention to the lack of complete certainty. Conversely, terms such as *possible* and *probable* draw attention to the positive possibility and thus impart a specific tone to the communication of uncertainty.

*Probable* and *possible* are used by the U.S. Environmental Protection Agency (EPA) and the International Agency for Research on Cancer (IARC) to distinguish agents based on qualities of the evidence pertinent to human carcinogenicity. Approximately 80% of the agents for which there are sufficient data relevant to carcinogenicity are labeled *probable* or *possible*; therefore, these two designations are key to the communication of carcinogen hazard. However, in this application, numerical probabilistic meanings are not specified. A classification of *probable* is supported by what is judged by agency scientists to be more conclusive evidence for human carcinogenic potential than that resulting in a *possible* classification. Thus, *probable* is used to represent greater hazard than *possible*.<sup>(2,3)</sup> A numerical interpretation of a designation does not translate to an individual risk of cancer given exposure or to a population frequency of cancer but rather to a subjective uncertainty about whether or not the compound could cause cancer in humans. From decision theory, if a perfectly calibrated judge were to assign to each of a group of chemicals an identical subjective numerical probability of its being a human carcinogen, the fraction of that group of chemicals that actually proved to be carcinogenic would equal the probability assigned. In other words, the perfectly calibrated judge would be right the percentage of time that equalled his subjective probability assignment.<sup>(6)</sup> Subjective uncertainty resulting from incom-

plete knowledge is one of Teigen's five dimensions of uncertainty. As with other studies,<sup>(7-11)</sup> the study described in this paper is concerned with eliciting probability levels, but as Teigen points out, the interpretation of the terms is influenced by other important dimensions of uncertainty.

*A priori*, it was expected from common usage that if point probability estimates were elicited for the designations, *possible* would be assigned a probability lower than *probable*, *probable* would be assigned a probability greater than 50%, and *possible* would be assigned a probability less than or equal to 50%. However, *possible*, by reflecting an event with something greater than a 0% chance of occurrence could be used to describe any event that has a probability of occurrence greater than 0% and less than 100%. This results in a more nebulous probabilistic connotation for *possible* than for *probable*. These relationships are confirmed by various studies. For example, Mosteller and Youtz<sup>(8)</sup> report mean and median probability assignments between 62% and 80% for *probable* in 10 different studies and between 27% and 55% for *possible* in 9 studies (Table I). In Mosteller and Youtz' study of science writers, *possible* has a larger range of probabilities assigned than most of the other expressions evaluated. However, the authors observe that *possible* seems to have distinct meanings for different groups and propose that the range of its probability assignments is not due to a general connotation of vagueness but instead, to a general diversity of interpretations. In a study of expressions of frequency, Pepper and Prytulak<sup>(12)</sup> showed that the numerical interpretation of a frequency expression is strongly influenced by the general frequency of the events it describes. That is, *probable* when it describes a high probability outcome from a relatively rare event could be expected to be assigned a lower numerical probability than *probable* when it describes a high probability outcome from a relatively common event. Thus, the context in which verbal probability phrases are used can affect numerical interpretations of the phrases.<sup>(13,14)</sup> Based on the average values reported in Graham's<sup>(11)</sup> study, where he examined seven descriptors of carcinogenic hazard and in the studies cited by Mosteller and Youtz, *probable* consistently communicated more certainty than *possible*. There does not appear to be a difference in magnitude between the probabilities assigned in Graham's study using the context of carcinogenicity and the other studies that used either no context or a context other than carcinogenicity.

In studies eliciting numerical meanings for verbal expressions, the words are often presented in arrays and the assessment of any given word can be influenced by the values ascribed to the words surrounding it. An an-

**Table I.** Mean and Median Probabilities Assigned to *Possible* and *Probable* in this Study, Graham's Study<sup>(11)</sup>, and Multiple Studies Reported by Mosteller and Youtz<sup>(6)</sup>

Studies listed by author	Sample size	Possible	Probable
S.E. Spedden and P.B. Ryan, <i>probable</i> -first (1991)	38	43	60
S.E. Spedden and P.B. Ryan, <i>possible</i> -first (1991)	38	43	70
J.D. Graham (1989)	202	41	67
J.D. Graham, scored alone (1989)	100	58	
E.M. Johnson (1973)	28	51	62
A. Kong, G.O. Barnett, F. Mosteller, and C. Youtz, uniform scale <sup>a</sup> (1986)	170	27	65
A. Kong, G.O. Barnett, F. Mosteller, and C. Youtz, free choice <sup>b</sup> (1986)	144	—	64
R.T. Reagan, F. Mosteller, and C. Youtz (1989)	115	40 <sup>c</sup>	70 <sup>c</sup>
F. Mosteller and C. Youtz (1990)	230	33	70
S. Lichtenstein and J.R. Newman (1967)	186	37	71
J.M. Levine and D. Eldredge (1970)	20	55	71
D.V. Budescu and T.S. Wallsten (1985)	32	38	72
G.D. Bryant and G.R. Norman (1980)	32	43	77
V.F. Reyna (1981)	41	55 <sup>c</sup>	80 <sup>c</sup>

<sup>a</sup> In the uniform scale format, choices for answers were integer increments from 0–10 out of 10 (e.g., 2 of 10).

<sup>b</sup> In the free choice format, any number between 0 and 100 was an acceptable answer.

<sup>c</sup> Median.

**Table II.** Questions Used to Elicit Probability Assignments for *Probable* and *Possible*

Please answer question 1 before opening the flap and reading question 2.

1. We are going to ask you questions about chemicals that may or may not cause cancer. We will be asking you to use a scale that goes from 0–100%. On this scale, a chemical that definitely causes cancer in humans would be rated 100%. A chemical that definitely does not cause cancer in humans would be rated 0. Suppose you hear that scientists have classified a particular chemical as a *possible* (*probable*) human carcinogen. On this scale, how would you rate this chemical?

Answer: \_\_\_\_\_

2. Suppose you hear that scientists have classified a particular chemical as a *probable* (*possible*) human carcinogen. How would you rate this chemical?

Answer: \_\_\_\_\_

choring effect<sup>(15,16)</sup> occurs when numerical responses are influenced by the value chosen as the starting point for formulating an assessment. Anchoring can occur when subjects are asked to assign values to two or more quantities and the magnitude of adjustment they make after assigning the first value to assigning succeeding values is insufficient. The starting value is typically an acces-

sible number such as a known quantity, a best guess, or an estimate of the median of a distribution. In the study described in this paper, anchoring can occur in two ways: (1) subjects can anchor the probability they assign to the second designation on the probability they assign to the first; or (2) they can anchor one or both probability assignments on readily accessible values. Anchoring can result in a smaller difference between the assignments than if the designations are evaluated independently or can be manifested by consistent assignments at frequently used values.

The study described in this paper was designed to test for an anchoring effect when only two phrases are evaluated, thus, the sequential presentations, (1) *probable* then *possible* and (2) *possible* then *probable* are pertinent. However, since the respondents were not allowed to change their first answers after reading the second question, the study can also be used to examine the meanings ascribed to each word in isolation.

### 3. SURVEY ADMINISTRATION

In this investigation, the survey questions (Table II) were administered to two classes of graduate-level health management students. Half of the respondents were asked to assign a probability to *possible* first and half to *prob-*

able first.<sup>3</sup> The questions were typed on pieces of paper that were folded and stapled so that the second question was not read before the first was answered. For each of the two sequences, there were 38 usable point-estimate responses, resulting in a total of 76 usable responses.

#### 4. RESULTS AND DISCUSSION

##### 4.1. Comparison of Numerical Probabilistic Meanings for *Probable* and *Possible*

Most of the respondents assigned point probabilities to the designations. Since the examples given were point estimates (Table II), the question wording may have discouraged an interval answer. Alternatively, since subjective probability assignments already incorporate vagueness and uncertainty,<sup>(17)</sup> a range estimate may be redundant. This study limited analysis to point estimate responses to provide comparability with other studies, ease in calculating mean values, and clarity in detecting anchoring effects.

Almost all of the respondents followed the basic assumptions and assigned *probable* a probability higher than *possible*, *possible* a value of 50% or lower, and *probable* a value greater than 50%. However, some of the respondents violated these assumptions. In 11 of the 76 usable point estimate responses, *probable* was assigned a score that was the same as or lower than the score that was assigned to *possible*. Some respondents may have had difficulty differentiating *probable* and *possible* because only two expressions instead of several were used, both expressions commonly represent more central than extreme likelihoods, and they were presented sequentially instead of simultaneously. Some investigators note that on a probability scale, the relative positions of probabilistic phrases can vary among individuals<sup>(17)</sup> while others<sup>(11,18)</sup> report that, in general, the relative ranks are fairly constant and unaffected by context. The specific expressions evaluated and the format in which they are presented could determine which of these effects is observed.

The mean probabilities assigned to *possible* and *probable* (Table III) are comparable to values cited by Mosteller and Youtz (Table I) and to values observed by Graham.<sup>(11)</sup> Thus, the summary measures seem to be fairly constant from study to study. The ranges of probabilities assigned to *possible* and *probable* overlap. Thus,

Table III. Measures of Central Tendency for Probabilities Assigned to *Probable* and *Possible* used to Describe Carcinogenic Hazard

	<i>Possible</i> -first		<i>Probable</i> -first	
	<i>Possible</i>	<i>Probable</i>	<i>Possible</i>	<i>Probable</i>
Sample size	38	38	38	38
Median	50	75	50	62.5
Mode	50	75	50	75
Mean	43	70	43	60
SD	22.4	17.9	26.7	26.0

even though most subjects ordered the designations as expected and assigned them different scores, there may be no numerical interpretations that unambiguously differentiate them.

The mean probability score is not necessarily representative of the probability assigned by most individuals. For example, the mean probability assigned to *possible* by both groups was 43%, but only three of the 76 respondents assigned a score in the range from 31–49%. *Probable* was assigned its mean value by only six of the 76 respondents. In this study, the mode represents a sizeable segment, about one third, of both the *probable* and *possible* responses and captures the common interpretation of the phrase more effectively than the mean. However, risk communication efforts based on one of the summary central value statistics would not communicate accurately to respondents at both extremes who might interpret the designations much differently from what was intended by the communicator.

Scatter plots of the data indicate that there may be a linear relationship between the value individuals assign to *possible* and the value they assign to *probable* (Figs. 1 and 2). The sample correlation coefficient,  $r$ , and the 95% confidence interval on the population correlation coefficient,  $\rho$ , are .46 (.17–.64) for the *possible*-first case and .65 (.40–.81) for the *probable*-first case. Thus, a linear relationship between these two values is a reasonable conclusion. For each question sequence, the line through the points that minimizes the perpendicular distance from the point to the line was calculated. These least distance lines are more appropriate for comparisons of these data than least-squares regression lines because it is not necessary to designate one of the variables as independent and the other as dependent. The fitted lines are shown in Figs. 1 and 2. In addition, the ratio *probable*/*possible* was calculated for both the *possible*-first and *probable*-first groups. The peak of the frequency distributions for the ratios is in the range having a midpoint of 1.55 (range: 1.2–1.9). The linear fit and the

<sup>3</sup> These questions were designed with the help of Dr. Nancy Neil, a cognitive psychologist and decision scientist.

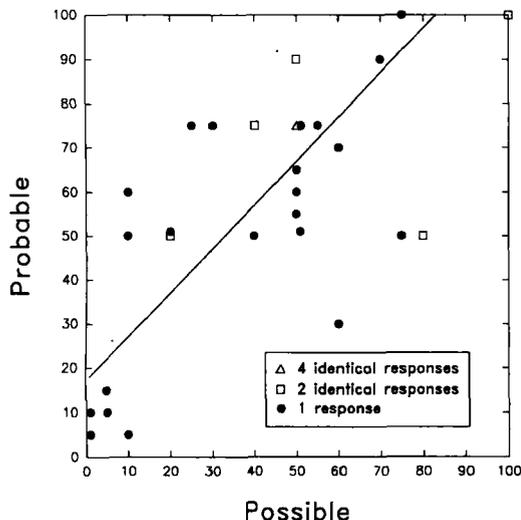


Fig. 1. Paired *probable* and *possible* probability assignments and fitted least-distance line where *probable* was evaluated first.

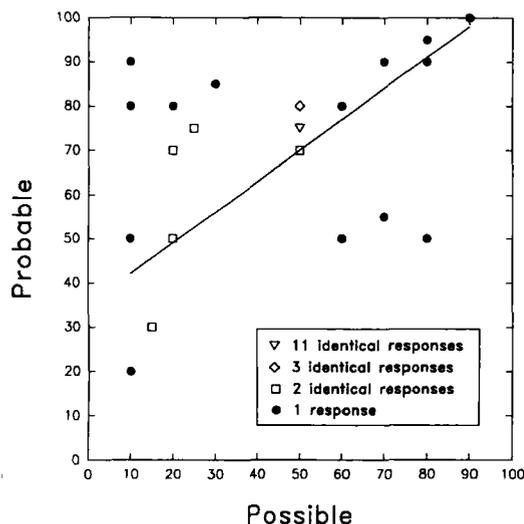


Fig. 2. Paired *probable* and *possible* probability assignments and fitted least-distance line where *possible* was evaluated first.

near-constant ratio may indicate that for many people, there is a consistent scaling from *possible* to *probable* or from *probable* to *possible* that could reflect an internal calibration. If a value were specified for one of the two designations, respondents should be able to scale the other designation proportionately. Alternatively, subjects may choose the values they assign more for their accessibility than for their numerical relationship to each other, they may find it difficult to accept a new definition

for a probabilistic phrase that to them already has a specific numerical meaning, and they may be using readily accessible values such as 50% to indicate approximations and might be unwilling to change their assignments to values that seem more precise (e.g., 46% or 53%).

#### 4.2. Anchoring and Effects of Sequence

Based on the averages and ranges of the probabilities assigned, sequence seems to affect the values ascribed to *probable* but not to *possible*. The descriptive statistics are summarized in Table III and illustrated by box plots in Fig. 3. All three measures of central tendency for *possible* are the same regardless of whether it is evaluated first or follows *probable*. For *probable*, the means and medians differ between the two sequences, but the modes are identical. The ranges of probabilities assigned to both designations are large. Beyth-Marom<sup>(7)</sup> uses the interquartile range ( $C_{25}$ - $C_{75}$ ) to highlight the most common numerical assignments and the 80% range ( $C_{10}$ - $C_{90}$ ) to highlight the variability in numerical assignments. These measures illustrate greater variability in probability assignments when *probable* is assessed first than when it follows *possible*. In both sequences, *probable* has a narrower interquartile range than *possible* indicating better agreement about numerical assignments for *probable* than for *possible*. Mosteller and Youtz<sup>(8)</sup> also report that *possible* has a broader range of numerical interpretations than *probable*. The small variability for assignments to *possible* reported by Beyth-Marom<sup>(7)</sup> could

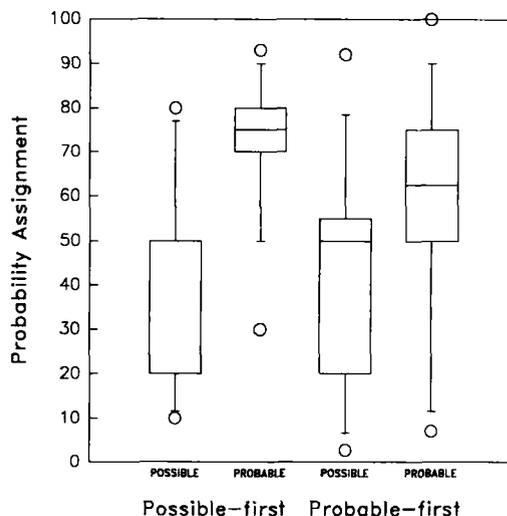


Fig. 3. Box plots of *possible* and *probable* by *possible*-first sequence and *probable*-first sequence.

be due to an effect of translation since in her study the probabilistic phrases were given in Hebrew and have been translated.

An appreciable fraction of the respondents assigned a score around 50% to the designation they evaluated first. The numerical meanings of *possible* and *probable* may be sufficiently obscure that when people are confronted with either in isolation, they start by assigning an accessible central value. Others have reported that probabilistic phrases can be interpreted as more central and less specific than intended by the communicator.<sup>(18)</sup> Since 50% is the mode probability assignment for *possible* in both sequences, this assignment may reflect a common probabilistic interpretation unaffected by anchoring. Most of the subjects who scored *possible* first and assigned it a value near 50%, assigned *probable* a value near 75%. This has a stabilizing effect on the distribution of probabilities assigned to *probable* demonstrated by a narrow interquartile range and can be attributed to anchoring.

Two cases at the tails of the distribution of scores assigned to *probable*, among those evaluating *probable* first, appear to demonstrate an anchoring effect, since the adjustments are fairly small. Of the respondents who scored *probable* first, the five who assigned *probable* a score of 15% or lower, assigned *possible* a score of 10% or lower. The three who assigned *probable* a score of 100%, assigned *possible* scores of 75–100%. However, the assignments could reflect other issues, such as degree of confidence in the scientists who determine carcinogen classifications, overall beliefs about risk from chemicals, and interpretation of how probabilistic expressions are used in a regulatory context. Beyth-Marom<sup>(7)</sup> characterizes these effects as: (1) personal evaluations in contrast to pure numerical translations; and (2) the influence of the desirability of the event on probability assessments. These effects could be manifestations of the influence of some of Teigen's<sup>(4)</sup> other dimensions. For example, the source of the uncertainty is incomplete knowledge rather than statistical variation, the communicators could be using these terms to communicate vagueness or lack of consensus, and the terms communicate a positive possibility instead of a lack of complete certainty.

The inclination might be to attribute to anchoring the difference in means between when *probable* is evaluated in isolation and when it is evaluated following *possible*. However, it is significant that the modes, which represent a sizeable segment of the respondents, show no sequence effect. The difference in means between the two sequences can likely be explained by the fact that the words were not evaluated simultaneously. More than half of the respondents assigned *probable* a score lower

than 70% when they scored it first, but less than one fourth scored it lower than 70% when they evaluated *possible* first. As discussed earlier, this effect occurs because many respondents assigned the phrase they evaluated in isolation a value near 50%; thus, *probable* was more likely to be assigned a lower value when it was evaluated in isolation than when it was evaluated following *possible*.

## 5. CONCLUSIONS

Likely anchoring effects were detected in this study. For example, accessible values were assigned to both designations by a large fraction of the respondents, *probable* had a less variable interpretation when it followed *possible* than when it occurred first, and respondents who assigned the first designation they evaluated an extreme value tended to assign the second designation a value close to the same extreme. An exact universal numerical interpretation does not exist for either *possible* or *probable*. In fact, the range of probabilities assigned is large, and even the 80% ranges are not mutually exclusive and cannot distinguish the two terms. However, despite sources of ambiguity in the interpretation of these verbal expressions of uncertainty, there are some points of consistency. Most of the respondents in this preliminary survey ascribed to *probable* a meaning of more likely than not (greater than an even chance) and to *possible* a meaning of no more than as likely as not (at most an even chance) and the interquartile ranges are nearly mutually exclusive. The mean probability scores do not reflect most individuals' numerical interpretations of either of these terms. However, a substantial fraction of the respondents assigned *probable* a value about 75% and *possible* a value about 50%, which are the modes of the distributions. Various factors can modify the numerical interpretations. Those people who respond to the effect of the hazard (cancer) or to the source of the communication (e.g., a government agency) rather than to the different levels of uncertainty represented by the designations might be expected to ascribe nearly identical probabilities to both designations. The proximity of other expressions of uncertainty, the sequence in which the expressions are presented, and the context in which the phrases are used can also influence the probabilistic interpretations.

The agencies' use of designations for carcinogen hazard that can have implicit numerical meanings and important nonnumerical dimensions is key to risk communication. Though some common interpretations for *probable* and *possible* can be identified, there are ex-

ceptions to these interpretations that must be considered. When these words are used to describe uncertainty surrounding the potential for human carcinogenicity, they should represent levels of uncertainty compatible with their common interpretations and their intended meanings should be stated.

Hazard designations are used to describe categories that differentiate agents by their likelihood of being carcinogenic to humans. Level and pattern of exposure as well as carcinogenic potency are not part of IARC's assessment and are considered separately from hazard identification by EPA. Since it is described with words that are in common usage, an agent's hazard classification is more accessible than either its exposure or potency information and can be applied more readily to risk communication. However, assessments that include considerations of elements such as exposure level and potency characterize the level of risk and should form the basis for risk communication. Thus, designations that implicitly convey probability of adverse effects should be reserved for classifications that are based on assessments of risk rather than only on assessments of hazards.

#### ACKNOWLEDGMENTS

Sarah Spedden was supported in part by the Occupational Safety and Health Educational Resource Center NIOSH grant award 5 T15 OH07096 at the Harvard School of Public Health and in part by a cooperative agreement (CR812-699-03) with the U.S. Environmental Protection Agency for the program Environmental Health and Public Policy in the Harvard Interdisciplinary Programs in Health. Barry Ryan acknowledges support by a grant from the U.S. Environmental Protection Agency through the Societal Institute of the Mathematical Sciences (SIMS) entitled Studies of Statistical and Environmental Factors in Health (Human Toxicology) and a Fellowship from the Winkler Foundation.

We value the effort our colleagues devoted to reviewing and commenting on this work and thank particularly John Evans, John Graham, and Nancy Neil.

#### REFERENCES

1. National Research Council, *Risk Assessment in the Federal Government: Managing the Process* (National Academy Press, Washington, D.C., 1983).
2. U.S. Environmental Protection Agency, "Guidelines for Carcinogen Risk Assessment," 51FR33992, *US Federal Register* (September 24, 1986).
3. International Agency for Research on Cancer, "Preamble," in *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans, Occupational Exposures in Petroleum Refining; Crude Oil and Major Petroleum Fuels*, Vol 45 (World Health Organization, 1989).
4. K. H. Teigen, "The Language of Uncertainty," *Acta Psychol* **68**, 27-38 (1988).
5. N. Kraus, T. Malmfors, and P. Slovic, "Intuitive Toxicology: Expert and Lay Judgments of Chemical Risks," *Risk Anal.* (1992).
6. S. Lichtenstein, B. Fischhoff, and L. D. Phillips, "Calibration of Probabilities: The State of the Art to 1980," in D. Kahneman, P. Slovic, and A. Tversky (eds.), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge University Press, Cambridge, U.K. 1982), Ch. 22.
7. R. Beyth-Marom, "How Probable is Probable? A Numerical Translation of Verbal Probability Expressions," *J Forecasting* **1**, 257-269 (1982).
8. F. Mosteller and C. Youtz, "Quantifying Probabilistic Expressions," *Statist. Sci.* **5**, 2-34 (1990).
9. S. Lichtenstein and J. R. Newman, "Empirical Scaling of Common Verbal Phrases Associated with Numerical Probabilities," *Psychon. Sci.* **9**, 563-564 (1967).
10. T. S. Wallsten, D. V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth, "Measuring the Vague Meanings of Probability Terms," *J. Exp. Psychol [General]* **115**, 348-365 (1986).
11. J. D. Graham, "Communicating About Chemical Hazards," *J. Policy Anal. Manag.* **8**, 307-313 (1989).
12. S. Pepper and L. S. Prytulak, "Sometimes Frequently Means Seldom: Context Effects in the Interpretation of Quantitative Expressions," *J. Res. Pers.* **8**, 95-101 (1974).
13. H. H. Clark, "Comment on Quantifying Probabilistic Expressions," *Statist. Sci.* **5**, 12-16 (1990).
14. J. N. Tanur, "Comment: On the Possible Dangers of Isolation," comment on "Quantifying Probabilistic Expressions," *Statist. Sci.* **5**, 21-22 (1990).
15. A. Tversky and D. Kahneman, "Judgment Under Uncertainty: Heuristics and Biases," *Science* **185**, 1124-1131 (1974).
16. B. Fischhoff, "Risk: A Guide to Controversy," Appendix C in *Improving Risk Communication*, (National Academy Press, Washington, D.C., 1989).
17. T. S. Wallsten and D. V. Budescu, "Comment on Quantifying Probabilistic Expressions," *Statist. Sci.* **5**, 23-26 (1990).
18. A. Kong, G. O. Barnett, F. Mosteller, and C. Youtz, "How Medical Professionals Evaluate Expressions of Probability," *N. Engl. J. Med.* **315**, 740-744 (1986).