

SEEKING AND APPLYING DIAGNOSTIC INFORMATION IN A HEALTH CARE SETTING *

Shawn P. CURLEY

University of Minnesota, USA

J. Frank YATES

University of Michigan, USA

Mark J. YOUNG

Medical College of Wisconsin, USA

Accepted June 1989

Many studies have shown that people have difficulty judging the diagnostic value of conditional probability information with respect to one or more hypotheses. The present research addressed two aspects of performing the diagnostic task in a health care decision: (a) recognition of the information's importance, and (b) correct usage of that information. In experiment 1, health care providers, who are trained in, and regularly exposed, to conditional probabilities imparting diagnostic information, exhibited at least a rudimentary recognition of the need for this information in assessing diagnosticity. Experiment 2 indicated that health care and layperson subjects had difficulty in actually applying the information, however. This difficulty prompts a need for judgment aids and caution in using diagnostic information.

An ongoing area of concern to researchers is the updating of judgments in light of diagnostic evidence. Bayes' Theorem has been pivotal to this research. In odds form, Bayes' Theorem states:

$$\frac{P(H|E)}{P(\bar{H}|E)} = \frac{P(H)}{P(\bar{H})} \times \frac{P(E|H)}{P(E|\bar{H})}. \quad (1)$$

Alternately:

$$\Omega_{\text{posterior}} = \Omega_{\text{prior}} \times L.$$

* The authors wish to thank Donald Bartlett and Bong-Oh Kim for their assistance with the data analysis.

Requests for reprints should be sent to S.P. Curley, Dept. of Information & Decision Sciences, University of Minnesota, 271 19th Ave. S., Minneapolis, MN 55455, USA.

The first term, the posterior odds, expresses the likelihood of the hypothesis H relative to its complement, after receiving the diagnostic evidence E . The second term in the equation, the prior odds or base rate, expresses the same judgment prior to the diagnostic evidence. The final term, the likelihood ratio, expresses the diagnosticity of the evidence with respect to the hypothesis H .

Early studies showed 'conservatism' in the use of diagnostic sample information (Edwards 1968; Phillips and Edwards 1966; reviews in Rapoport and Wallsten 1972; Slovic and Lichtenstein 1971). Subsequently, studies have consistently demonstrated that people do not intuitively use Bayes' Theorem in applying the diagnosticity of available information to the evaluation of a hypothesis (Fischhoff and Beyth-Marom 1983; Kahneman and Tversky 1972; Slovic et al. 1977). Attention has subsequently focused upon usage of the two terms on the right side of eq. (1). Most of this attention has been directed toward use of prior odds, or base rate, information (Ajzen 1977; Bar-Hillel 1980; Medin and Edelson 1988; Tversky and Kahneman 1980). The present paper addresses the diagnosticity portion of the right side of eq. (1), embodied in the likelihood ratio.

Two questions are of interest with respect to diagnostic conditional probability information. First, do individuals correctly recognize what specific diagnosticity information is needed in evaluating the hypotheses, and to what extent? Second, whether or not they recognize the need for the essential information, can they correctly use that information if it is supplied?

An exemplar diagnosticity judgment task is illustrated by the display in table 1. Subjects were to decide which of the two diagnoses, disease A or B, was more likely. They were provided with one piece of information. In table 1, the percentage of those with disease A who show the symptom of fever is 66%; 34% of those with disease A do not have fever. To help the subjects in their selection, one additional item

Table 1
Exemplar diagnosticity task.

	Disease A	Disease B
Fever	66%	X1
Rash	X2	X3

of information – $X1$, $X2$, or $X3$ – was offered. The normative response is $X1$, the percentage of those with disease B who show the symptom of fever. This item would allow determination of the diagnosticity of the sign ‘fever’ for the two diseases. Typically, subjects do not select this item (Wolf et al. 1985). They tend to prefer the information in cell $X2$, and then use it as if it were diagnostic, behavior labelled ‘pseudodiagnosticity’ (Doherty et al. 1979). Einhorn and Hogarth (1978; Einhorn 1980) have suggested that this behavior might arise from the general lack of cell $X1$ information in the environment.

One difficulty with the pseudodiagnosticity studies is their restriction of subjects to only one additional piece of information. Natural environments are not usually so constrained. However, Beyth-Marom and Fischhoff (1983) allowed any number of requests, and their results corroborated the findings: subjects were reluctant to select $P(E|\bar{H})$ information. Further, when offered reasons for selecting the different pieces of information, even subjects who selected $P(E|\bar{H})$ had difficulty in choosing the appropriate justification from among those provided.

We report two experiments here which extend our knowledge of subjects’ ability to understand and use conditional probability indicators as capturing diagnostic, or contingency, information. The experiments address the two main questions with respect to diagnostic indicators: do subjects recognize the need for them? Can subjects correctly use them?

Experiment 1 generalized the results of Beyth-Marom and Fischhoff (1983) with respect to the first of these two questions to physician subjects. The medical setting is one in which conditional probabilities are readily available. Physicians in medical settings are trained in, and regularly see, conditional probability information for clinical tests. The information is communicated as test sensitivity, $P(T^+ | D)$, and test specificity, $P(T^- | \bar{D})$. Those with the disease are D ; those without the disease are \bar{D} . The result of test T is the diagnostic evidence; those who test positive are T^+ ; those who test negative are T^- . Unlike the subjects used in prior research, these subjects might well be expected to be familiar with the necessity of having this information.

Subjects in this setting were interrogated with an open request format comparable to that used by Beyth-Marom and Fischhoff. Subjects were allowed to request all pieces of evidence deemed relevant. In addition, the subjects provided justifications for their selections. They

did this in their own words, rather than being restricted to experimenter-provided responses. Thus, experiment 1 was designed to explore the robustness of Beyth-Marom and Fischhoff's findings by reducing various barriers to subjects' expressions of the importance of diagnostic evidence. Experiment 2 addressed the second question of whether subjects sufficiently understand the conditional probability information and can put it to proper use.

Experiment 1

Method

Subjects

Eighteen fourth-year medical students and 34 fellowship physicians from the University of Michigan Medical Center served as subjects. The physicians were from a variety of medical specialties, comprising a cross-section. The subjects completed the task individually, along with several other unrelated tasks, and were paid a fixed fee for their participation.

Procedure

The instructions to the subjects appear in the appendix. The subjects were informed that they were to gather information in their hospital regarding a disease and a diagnostic test for that disease. They were offered seven pieces of information for which to search, and were to avoid irrelevant and redundant information. For each item, subjects indicated whether or not they would order a search for that item, and why or why not. The items of information were described in words, and corresponded to the following: $P(D)$, $P(\bar{D})$, $P(T^+ | D)$, $P(T^- | D)$, $P(T^+ | \bar{D})$, $P(T^- | \bar{D})$, $P(T)$. T represents those receiving the test, regardless of test outcome. This probability notation was not presented to the subjects.

The first two items are base rate information, and one is redundant given the other. Similar redundancies exist for the conditional probabilities. To assess the diagnosticity of the test T , the subject needed two of the four conditional probabilities, e.g., $P(T^+ | D)$ and $P(T^+ | \bar{D})$. The probabilities $P(T^+ | D)$ and $P(T^- | D)$ are redundant, as are the probabilities $P(T^+ | \bar{D})$ and $P(T^- | \bar{D})$. Redundant items were used as a manipulation check for the instruction that items were costly.

The subjects' justifications for their selections were coded by two independent, blind coders. The coders were instructed to be highly sensitive to the justifications as indicative of a general understanding of the proper use of the information. That is, the coders were to settle in the subject's favor in rating the justifications. This provided a conservative estimate of the proportion of subjects who were unable to express the need for the diagnostic information. The coders agreed upon 131 of the 138 responses presented to them. The inter-rater reliability of 94.9% was taken to be acceptable. Responses for which there was no agreement were not included in the analyses as described below.

Table 2
Requests and justifications for available diagnostic information (experiment 1).

	Information pairs		
	Base rate $P(D); P(\bar{D})$	Sensitivity $P(T^+ D); P(T^- D)$	Specificity $P(T^+ \bar{D}); P(T^- \bar{D})$
Number (%) requesting at least one of this pair	41/52 (78.8%)	51/52 (98.1%)	52/52 (100%)
Number (%) of these with justification consistent with proper use	34/36 ^a (94.4%)	46/47 ^b (97.9%)	46/48 ^b (95.8%)

Note: D = those with the disease. \bar{D} = those without the disease. T^+ = those who test positive. T^- = those who test negative.

^a Two subjects did not provide justifications, and the two raters disagreed on three subjects' justifications. These five are omitted.

^b Two subjects did not provide justifications, and the two raters disagreed on two subjects' justifications. These four are omitted.

Results and Discussion

The manipulation check indicated that subjects heeded the instruction that there was a search cost. Only 2/52 (3.8%) of the health care providers requested all seven pieces of information; they did discriminate. However, their discrimination was not uniform. Subjects had particular difficulty acknowledging the redundancy of the conditional probability information. Of the remaining 50 subjects, 5 (10%) requested both base rate items, 26 (52%) requested both $P(T^+ | D)$ and $P(T^- | D)$, 23 (46%) requested both $P(T^+ | \bar{D})$ and $P(T^- | \bar{D})$. Out of 50, 22 (44%) subjects correctly recognized that they needed exactly one from each pair [$p(T^+ | D)$, $p(T^- | D)$] and [$p(T^+ | \bar{D})$, $p(T^- | \bar{D})$].

The item requests, shown in the upper half of table 2, indicated that the health care subjects recognized that conditional probability information was needed to determine diagnosticity. At least one of the base rate items was selected by 78.8% of the subjects; 98.1% selected at least one of $P(T^+ | D)$ or $P(T^- | D)$; and, all 52 subjects selected at least one of $P(T^+ | \bar{D})$ or $P(T^- | \bar{D})$. The percentage selecting conditional likelihood information significantly differed from the percentage seeking base rate information (using Cochran's (1950) test of proportions from matched samples, $Q(1) = 10.0$, $p < 0.01$), underscoring the subjects' recognition that likelihood data were useful.

Further, of the subjects who selected the items, nearly all of them gave a justification which was at least broadly consistent with the proper usage of that information within a Bayesian updating procedure, as determined by the coders (lower half of table 2). Examples of subjects' correct justifications included:

[$P(D)$, $P(\bar{D})$]:

This gives the incidence of disease in your population.

Prevalence of disease important in a screening situation.

$[P(T^+ | D), P(T^- | D):]$

This information is needed to establish the sensitivity of the test.

Need to get true positive rate of test.

To evaluate the diagnostic test for the disease.

$[P(T^+ | \bar{D}), P(T^- | \bar{D}):]$

To determine the false positive rate.

Must know specificity of the test in order not to rely too heavily on results.

How useful is the test--if all test positive it is not useful.

As these examples indicate, the health care subjects had at least a minimal awareness of what the information represented. They typically communicated this understanding using the 'sensitivity' and 'specificity' labels for the described conditional probabilities.

In summary, there was an indication that the health care subjects had some understanding that the conditional probability information was needed to evaluate a diagnostic test. They could generally correctly tie the conditional probability statements to the clinical constructs of sensitivity and specificity in recognition of the information's usefulness. The redundancy of requests, however, suggested some difficulty in understanding the information. If this misunderstanding existed, we would also expect it to manifest itself in attempts to use the information. The issue of how subjects use the information was addressed in experiment 2.

Experiment 2

Method

Subjects

The subjects described in experiment 1, plus two additional fellowship physicians, completed the task individually. Also from the Medical Center, 29 chief medical residents gathered for a short course were enlisted. Extending beyond health care providers, 208 undergraduates from the University of Michigan took part in small groups. All but the chief residents were paid a set fee for participating.

Procedure

Each subject received a base rate, $P(D)$, and diagnostic test information, $P(T^+ | D)$ and $P(T^+ | \bar{D})$, all described in words. The subjects did not see probability notation. The subject's task was to judge the probability of disease after receiving a positive test result, by placing a slash through a scale anchored at 0 and 100. The instructions for experiment 2 are included in the appendix.

There were two considerations in our design. First, Beyth-Marom and Fischhoff (1983) suggested that subjects are able to use diagnostic information properly. However, these investigators used a limited class of diagnostic values. Only probabilities opposite and equidistant from $P = 0.50$ were included, e.g., $P(T^+ | D) = 0.70$ and $P(T^+ | \bar{D}) = 0.30$. We wanted to extend their conclusions to a wider class of values. For example, it may be that subjects implicitly, and incorrectly, assume that $P(T^+ | \bar{D}) = 1 - P(T^+ | D)$, and behave accordingly.

Table 3
Experimental conditions and 'rationally' predicted direction for revising the probability of disease for each condition (experiment 2).

Experimental condition	$P(T^+ D)$	$P(T^+ \bar{D})$	Predicted revision
30/-	0.30	(omitted)	?
30/70	0.30	0.70	Down
30/50	0.30	0.50	Down
30/10	0.30	0.10	Up
70/-	0.70	(omitted)	?
70/30	0.70	0.30	Up
70/50	0.70	0.50	Up
70/90	0.70	0.90	Down

A complementary consideration was to test the hypothesis that subjects with incomplete diagnosticity information (for example, in the design described by table 1) implicitly fill in the missing value. Thus, subjects who only have $P(T^+ | D)$ would assume a value for $P(T^+ | \bar{D})$ and respond on this basis. In particular, a pilot study suggested the complement of $P(T^+ | D)$ as a likely assumption for the missing value of $P(T^+ | \bar{D})$. The present experiment contained two conditions in which subjects received only partial diagnostic evidence.

Eight combinations of the conditional probabilities were used in the present study. In all cases, $P(D) = 0.50$ was given. As indicated by table 3, four groups saw $P(T^+ | D) = 0.70$, with one group each receiving $P(T^+ | \bar{D})$ values of 0.90, 0.50, 0.30, and one group for whom this second conditional probability was not given. Also, four groups saw $P(T^+ | D) = 0.30$, with one group each receiving $P(T^+ | \bar{D})$ values of 0.10, 0.50, 0.70, and one group for whom this conditional probability was missing. Thus, there were eight conditions of the study in a between-subjects design.

Results and Discussion

Subjects' use of diagnostic information can be compared to that prescribed by Bayes' Theorem (table 5). However, given the time and complexity entailed in using eq. (1), we should not expect subjects to be accurate intuitive Bayesians in the present task. Instead, a more basic property of rational information usage was deemed more meaningful. If $P(T^+ | D) > P(T^+ | \bar{D})$, then an individual should revise $P(D)$ upward upon receiving a positive test result, since T^+ is more likely to occur if the patient has the disease than if the patient does not have the disease. Similarly, if $P(T^+ | D) < P(T^+ | \bar{D})$, then an individual should revise $P(D)$ downward. These predictions are indicated in tables 3 and 4 for each condition. As a corollary, the greater the discrepancy between $P(T^+ | D)$ and $P(T^+ | \bar{D})$, the greater should be the revision. These principles for the use of diagnostic evidence have normative appeal independent of Bayes' Theorem. If subjects' diagnostic judgments were rational, they should reflect these principles. Did they?

Table 4

Nonparametric analyses (experiment 2): Response frequencies and direction of revision for the probability of disease by each group.

Condition	<i>n</i>	# Revising upward	# Revising downward	Predicted revision	Predominant revision
30/-	29	6	19 ^a	?	Down
30/70	24	4	16 ^a	Down	Down
30/50	29	2	20 ^b	Down	Down
30/10	29	6	20 ^a	Up	Down
70/-	45	35	7 ^b	?	Up
70/30	45	31	9 ^b	Up	Up
70/50	46	31	7 ^b	Up	Up
70/90	44	19	14	Down	None

Note: The binomial distribution was used for H_0 : Proportion shifting Up = Proportion shifting Down.

^a $p < 0.02$; ^b $p < 0.001$.

No reliable differences were observed among any of the subject groups; therefore, the data from all 291 subjects were pooled and analyzed together. Table 4 shows the numbers of subjects who revised the probability of disease either upward or downward in response to the diagnostic evidence for each of the eight conditions of the study. The nonparametric analyses in table 4 illustrate that the modal direction of revision was consistently in the direction of $P(T^+ | D)$. Table 5 shows the mean posterior judgments for each condition. The parametric analyses were corroborative. Columns 4 and 5 indicate that subjects typically were shifting from the base rate of $P(D) = 0.50$ in the

Table 5

Parametric analyses (experiment 2): Mean revised probability of disease for each group.

Condition	<i>n</i>	$P(D T^+)$ using Bayes' Theorem	Mean $P(D T^+)$ Judgment	vs. 0.50 ^a	vs. Bayes ^b
30/-	29	?	0.399	^c	np
30/70	24	0.30	0.409	^d	^d
30/50	29	0.375	0.385	^f	
30/10	29	0.75	0.378	^d	^f
70/-	45	?	0.654	^f	np
70/30	45	0.70	0.589	^e	^f
70/50	46	0.625	0.556	^d	^e
70/90	44	0.25	0.544		^f

^a The $t(n-1)$ statistic was used for H_0 : mean judged $P(D|T^+) = P(D) = 0.50$.

^b The $t(n-1)$ statistic was used for H_0 : mean judged $P(D|T^+) = P(D|T^+)$ using Bayes' Theorem ('np' signifies that test was not performed).

^c $p < 0.05$; ^d $p < 0.01$; ^e $p < 0.001$; ^f $p < 0.0001$.

direction of $P(T^+ | D)$, and column 6 supports that the shift was not generally in accord with the revision prescribed by Bayes' Theorem.

Overall, the most striking observation is that the subjects' responses were relatively insensitive to the value of $P(T^+ | \bar{D})$. In fact, there was no difference across conditions 70/30, 70/50, and 70/90 ($F(2, 132) = 0.99$, $p > 0.10$), or across conditions 30/70, 30/50, and 30/10 ($F(2, 79) = 0.27$, $p > 0.10$). (Nonparametric Kruskal-Wallis tests of differences led to the same results with $p > 0.30$ and $p > 0.20$, respectively.) Contrary to the performance of subjects in the Beyth-Marom and Fischhoff (1983) study, subjects had difficulty using diagnostic information in the present task. Subjects did not distinguish differences in $P(T^+ | \bar{D})$, despite the fact that some of these conditions should have differed qualitatively and sharply. Because no differences were observed across levels of $P(T^+ | \bar{D})$, the attempt to identify implicitly assumed values of this quantity when missing was not relevant. Indeed, the subjects' judgments could be well-described as arising from a simple strategy whereby, if $P(T^+ | D) > 0.50$, then the revision was upward; and, if $P(T^+ | D) < 0.50$, then the revision was downward.

General discussion

Individuals have been shown to have difficulty in correctly using diagnostic information. Einhorn and Hogarth (1978; Einhorn 1980) have implicated the absence of this information in subjects' experience as underlying difficulty in its use. Unlike previous results using less experienced subjects, we found that health care subjects did have some recognition of the need for the appropriate conditional probability information for judging diagnosticity. Physicians are schooled in the use of, and regularly see, this information in the form of the sensitivity and specificity of diagnostic tests; and they have at least a rudimentary understanding of its necessity. However, knowledge of how to apply this information is less established. Difficulty with the use of the information was suggested by the high number of redundant requests in experiment 1, and verified more directly in experiment 2.

Both health care and layperson subjects in experiment 2 were insensitive to one of the conditional probabilities, $P(T^+ | \bar{D})$, in making diagnosticity judgments. This insensitivity led to violations of a basic principle, which states that a test is diagnostic in favor of the hypothesis for which the result is more likely, either D or \bar{D} . This principle has prescriptive force independent of any particular rule for transforming the conditional probabilities into a single judgment of diagnostic impact, e.g., Bayes' Theorem.

Table 6
Levels of increasing abstraction and structuring (top to bottom) of contingency information.

-
1. *Data instances*
 - A. Case-by-case/sequential
 - B. Block/concurrent
 2. *Count /tally*
 - A. Frequency list
 - B. Organized presentation
 3. *Indicators*
 - A. Conditional probabilities
 - B. Single number/contingency index
-

For a wider perspective on these findings, note that this and other studies of individuals' judgments with diagnostic information are conceptually related to a series of contingency judgment experiments. In the most basic contingency judgment task, the co-occurrence of levels of two dichotomous variables is the covariation of interest. The subject receives data concerning the frequencies with which instances of each of the four combinations of the two variables occur. A judgment of the existence and strength of the relationship between the two variables is requested. Subjects have been found to have difficulty with this task, as well (Arkes and Harkness 1983; Beyth-Marom 1982; Crocker 1981; Curley et al. 1988). In the diagnosticity task, a similar contingency judgment is required, for example, between the diseases and signs in table 1. The task differs only in the level of structure in the data.

Table 6 shows successive levels of increasing abstraction and structure. The data instances themselves are the most concrete and least structured form of contingency information. In this form (form 1), the data may be observed sequentially, or all at once. The physician's everyday practice exemplifies receiving information in this form, sequentially, case-by-case. More abstract is a count or tally of the data instances, which may be more or less organized (form 2). The tabular form in table 1 exemplifies an organized frequency presentation format.

More compactly, the data can be described using indicators. It is with data of this form that diagnosticity studies are concerned. In form 3A, the data are described as pairs of conditional probabilities. The sensitivity and specificity of a diagnostic test are data at this level of abstraction. In form 3B, the data are structured even further into a

single contingency indicator, for example, as a correlation coefficient or posterior probability. In the diagnosticity context of the present experiments, the subjects had difficulty in transforming contingency information from form 3A, information concerning $P(T^+ | D)$ and $P(T^+ | \bar{D})$, to the single indicator form 3B, a posterior probability in light of test information.

It is therefore indicated that individuals would benefit from prescriptive aids in using diagnostic information at this level of abstraction. Recent attention to the likelihood ratio, a single indicator capturing the essential implications of both conditional probabilities, as useful for understanding clinical findings is justified by the current findings (cf. Sackett et al. 1985). For example, a test relative to a particular disease could be summarized as having 80% sensitivity $P(T^+ | D)$, 60% specificity $P(T^- | \bar{D})$, and a likelihood ratio of 2. That is, in the sample, a positive test result was twice as likely to be found among diseased patients than among patients without the disease. We need to explore more understandable summaries of diagnosticity, rather than relying on judgment and intuitive understanding.

The results also highlight particular situations in which the misuse of diagnostic evidence, e.g., from a laboratory test, is expected to be most pronounced. In selecting among tests, an insensitivity to values of $P(T^+ | \bar{D})$ would adversely influence the relative evaluations of tests having differing sensitivities and specificities. With respect to individual tests, it is true that cases in which there is high sensitivity ($P(T^+ | D)$ is large) but low specificity ($P(T^+ | \bar{D})$ is also large) rarely occur, if at all. However, cases at the other end are common, i.e., those in which the diagnostic evidence has low sensitivity ($P(T^+ | D)$ is small) but high specificity ($P(T^+ | \bar{D})$ is also small). A test having these properties would be particularly open to misinterpretation and misuse.

Appendix

Question text: Experiment 1

Suppose you are attempting to gather information in your hospital regarding a particular disease and a diagnostic test for that disease. The information must be gathered by a chart review of cases in your hospital. The process is both costly and time-consuming, so you should be sure that the information to be gathered is informative. Specifically, you want to avoid irrelevant or redundant information.

Many patients are admitted to your hospital from suspicion of the disease of interest. You can order a search for each of the following items of information in this group of patients. However, each of the items would involve conducting a separate search. For which of the following items would you request such a search in this group of patients?

- (1) The percentage of these patients who have the disease?
- (2) The percentage of these patients who do not have the disease?
- (3) The percentage of those with the disease who test positive?
- (4) The percentage of those with the disease who test negative?
- (5) The percentage of those without the disease who test positive?
- (6) The percentage of those without the disease who test negative?
- (7) The percentage of these patients who receive the test?

Question Text: Experiment 2

You are seeing a patient in whom you suspect coronary disease. After the history and initial examination, you determine that: If you were to see 100 patients like this one, you would expect 50 of them to have significant coronary obstruction.

You are now evaluating the results of a new diagnostic test for coronary disease. You have the following information about this new test with respect to patients with suspected coronary disease.

The percentage of those with significant coronary obstruction who test positive is # # out of 100.

The percentage of those without significant coronary obstruction who test positive is # # out of 100.

Suppose your patient receives a positive test. Given this result, if you were to see 100 patients like this one, in how many of them would you expect to find significant coronary obstruction?

References

- Ajzen, I., 1977. Intuitive theories of events and the effects of base rate information on prediction. *Journal of Personality and Social Psychology* 35, 303–314.
- Arkes, H.R. and A.R. Harkness, 1983. Estimates of contingency between two dichotomous variables. *Journal of Experimental Psychology: General* 112, 117–135.
- Bar-Hillel, M., 1980. The base-rate fallacy in probability judgments. *Acta Psychologica* 44, 211–233.
- Beyth-Marom, R., 1982. Perception of correlation reexamined. *Memory and Cognition* 10, 511–519.
- Beyth-Marom, R. and B. Fischhoff, 1983. Diagnosticity and pseudodiagnosticity. *Journal of Personality and Social Psychology* 45, 1185–1195.

- Cochran, W.G., 1950. The comparison of percentages in matched samples. *Biometrika* 37, 256–266.
- Crocker, J., 1981. Judgment of covariation by social perceivers. *Psychological Bulletin* 90, 272–292.
- Curley, S.P., M.J. Young, M.J. Kingry and J.F. Yates, 1988. Primacy effects in clinical judgments of contingency. *Medical Decision Making* 8, 216–222.
- Doherty, M.E., C.R. Mynatt, R.D. Tweney and M.D. Schiavo, 1979. Pseudodiagnosticity. *Acta Psychologica* 43, 111–121.
- Edwards, W. 1968. 'Conservatism in human information processing'. In: B. Kleinmuntz (ed.), *Formal representation of human judgment*. New York: Wiley. pp. 17–52.
- Einhorn, H.J., 1980. 'Learning from experience and suboptimal rules in decision making'. In: T.S. Wallsten (ed.), *Cognitive processes in choice and decision behavior*. Hillsdale, NJ: Erlbaum. pp. 1–20.
- Einhorn, H.J. and R.M. Hogarth, 1978. Confidence in judgment: Persistence of the illusion of validity. *Psychological Review* 85, 395–416.
- Fischhoff, B. and R. Beyth-Marom, 1983. Hypothesis evaluation from a Bayesian perspective. *Psychological Review* 90, 239–260.
- Kahneman, D. and A. Tversky, 1972. Subjective probability: A judgment of representativeness. *Cognitive Psychology* 3, 430–454.
- Medin, D.L. and S.M. Edelson, 1988. Problem structure and the use of base-rate information from experience. *Journal of Experimental Psychology: General* 117, 68–85.
- Phillips, L.D. and W. Edwards, 1966. Conservatism in a simple probability inference task. *Journal of Experimental Psychology* 72, 346–354.
- Rapoport, A. and T.S. Wallsten, 1972. Individual decision behavior. *Annual Review of Psychology* 23, 131–175.
- Sackett, D.L., R.B. Haynes and P. Tugwell, 1985. *Clinical epidemiology: A basic science for clinical medicine*. Boston, MA: Little, Brown and Co.
- Slovic, P. and S. Lichtenstein, 1971. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance* 6, 649–744.
- Slovic, P., B. Fischhoff and S. Lichtenstein, 1971. Behavioral decision theory. *Annual Review of Psychology* 28, 1–39.
- Tversky, A. and D. Kahneman, 1980. 'Causal schemas in judgments under uncertainty'. In: M. Fishbein (ed.), *Progress in social psychology*. Hillsdale, NJ: Erlbaum.
- Wolf, F.M., L.D. Gruppen and J.E. Billi, 1985. Differential diagnosis and the competing-hypotheses heuristic: A practical approach to judgment under uncertainty and Bayesian probability. *Journal of the American Medical Association* 253, 2858–2862.