we have described. He could then examine closely the size and source of the illusory correlations he experiences and thereby, one hopes, learn to guard against such errors in his clinical practice.

The experience would also remind him that his senses are fallible, that his clinical judgments must be checked continually against objective measures, and that his professional task is one of the most difficult and complex in all of psychology.

18. Probabilistic reasoning in clinical medicine: Problems and opportunities

David M. Eddy

To a great extent, the quality and cost of health care are determined by the decisions made by physicians whose ultimate objective is to design and administer a treatment program to improve a patient's condition. Most of the decisions involve many factors, great uncertainty, and difficult value questions.

This chapter examines one aspect of how these decisions are made, studying the use of probabilistic reasoning to analyze a particular problem: whether to perform a biopsy on a woman who has a breast mass that might be malignant. Specifically, we shall study how physicians process information about the results of a mammogram, an X-ray test used to diagnose breast cancer. The evidence presented shows that physicians do not manage uncertainty very well, that many physicians make major errors in probabilistic reasoning, and that these errors threaten the quality of medical care.

The problem

A breast biopsy is not a trivial procedure. The most common type (around 80%) is the excisional biopsy, in which the suspicious mass is removed surgically for microscopic examination and histological diagnosis by a pathologist. Usually the patient is admitted to a hospital and given a full set of preoperative diagnostic tests. The biopsy is almost always done under general anesthesia (with a probability of approximately 2 out of 10,000 of an anesthetic death). A small (1- to 2-in.) incision is made, and tissue the size of a pecan to a plum is removed. In many cases (perhaps 1 in

The preparation of this paper was supported by a grant from The Henry J. Kaiser Family Foundation.

2) the loss of tissue is barely noticeable; in others there is an indentation remaining. In an occasional case (perhaps 1 in 200) there is an infection or drainage that can persist for several weeks. The charge is approximately $700. This procedure can be done on an outpatient basis and under local anesthesia. As an alternative to the excisional biopsy, some surgeons prefer in some cases to obtain tissue by using a needle. This can be done on an outpatient basis, leaves no scar or other residual effects, and is far less expensive. However, it is thought by many physicians to be less reliable in that an existing malignant lesion may be missed.

An important factor that affects the need for biopsy is the possibility that the breast mass is a cancer. To estimate this possibility, a physician can list the possible diseases, assess the frequencies with which various signs and symptoms occur with each disease, compare this information with the findings in the patient, estimate the chance that she has each of the diseases on the list, and perform a biopsy if the probability of cancer or another treatable lesion is high enough. To help the physician, many textbooks describe how non-malignant diseases can be differentiated from cancer. For example, the following passage describes one such benign disease – chronic cystic disease.

> Chronic cystic disease is often confused with carcinoma of the breast. It *usually* occurs in parous women with small breasts. It is present *most commonly* in the upper outer quadrant but *may* occur in other parts and eventually involve the entire breast. It is *often* painful, particularly in the premenstrual period, and accompanying menstrual disturbances are *common*. Nipple discharge, *usually* serous, occurs in *approximately* 15% of the cases, but there are no changes in the nipple itself. The lesion *is* diffuse without sharp demarcation and without fixation to the overlying skin. Multiple cysts are firm, round, and fluctuant and *may* transilluminate *if* they contain clear fluid. A large cyst in an area of chronic cystic disease *feels* like a tumor, but it is *usually* smoother and well delimited. The axillary lymph nodes are *usually* not enlarged. Chronic cystic disease *infrequently* shows large bluish cysts. More often, the cysts are multiple and small.[1] (del Regato, 1970, pp. 860–861)

Similar descriptions are available for fibroadenomas, fat necrosis, trauma, and a half dozen other breast conditions, as well as for cancer.

This type of probabilistic information can be used to help a physician analyze the possible causes of a patient's breast mass. With assessments of the values of the possible outcomes (e.g., properly diagnosing a cancer, doing an unnecessary biopsy of a non-malignant lesion, not biopsying and missing a malignant lesion, and properly deciding not to biopsy a benign lesion), the physician can assess the chance that the patient, with her particular signs and symptoms, has cancer, and the physician can select an action.

[1] In this and all subsequent quotations, the italics are added.

The case of mammography

Other diagnostic tests are available to help the physician estimate the chance that a particular woman's breast lesion is malignant. Perhaps the most important and commonly used is mammography. The value of this test rests on the fact that the components of malignant cells absorb X rays differently from the components of non-malignant cells. By studying the mammograms, a radiologist may be able to see certain signs that occur with different frequencies in different lesions, and from this information a judgment can be made about the nature of the lesion in question. Typically, mammograms are classified as positive or negative for cancer. Occasionally an expanded classification scheme is used, such as one containing the three classes: malignant, suspicious, and benign.

The test is not perfect, in that some malignant lesions are incorrectly classified as benign and some benign lesions are called malignant. Thus, one factor that is very important to the clinician is the accuracy of the test.

Probabilistic reasoning

Let us develop this notion more precisely. The purpose of a diagnostic test is to provide information to a clinician about the condition of a patient. The physician uses this information to revise the estimate of the patient's condition and to select an action based on that new estimate. The action may be an order for further diagnostic tests, or if the physician is sufficiently confident of the patient's condition, a therapeutic action may be taken. The essential point is that the physician can have degrees of certainty about the patient's condition. The physician will gather evidence to refine this certainty that the patient does or does not have cancer, and when that certainty becomes sufficiently strong (in the context of the severity of the disease and the change in prognosis with treatment), action will be taken.

We can associate a probability, the physician's subjective probability that the patient has cancer, with this degree of certainty. The impact on patient care of a diagnostic test such as mammography, therefore, lies in its power to change the physician's certainty or subjective probability that the patient has cancer.

The notion of a subjective probability or degree of certainty appears in many different forms in the medical vernacular. For example, one author writes that "because the older age group has the greatest proportion of malignant lesions, there is heightened *index of suspicion* of cancer in the mind of a clinician who faces an older patient" (Gold, 1969, p. 162). Another author states that the mammogram can reduce the number of breast biopsies "in many instances when the examining physician's *rather firm opinion* of benign disease is supported by a *firm mammographic diagnosis* of benignancy" (Wolfe, 1964, p. 253). A third describes it this way: "If the *subjective impression* of the clinician gives enough reason for suspicion of carcinoma, the clinician will be compelled to biopsy despite a negative mammogram" (Clark, et al., 1965, p. 133). Other expressions that reflect this notion include, "*confidence level*" (Byrne, 1974, p. 37). "*impression of malignancy*" (Wolfe, 1967, p. 138), "a *more positive diagnosis*" (Egan, 1972, p. 392), and so forth. These statements are not precise because few physicians are formally acquainted with the concepts of subjective probability and decision analysis. Nonetheless, there is ample evidence that the notions of degrees of certainty are natural to physicians and are used by them to help select a course of action.

Interpreting the accuracy of mammography

Now consider a patient with a breast mass that the physician thinks is probably benign. Let this probability be 99 out of 100. You can interpret the phrase "that the physician thinks is probably [99 out of 100] benign" as follows. Suppose the physician has had experience with a number of women who, in all important aspects such as age, symptoms, family history, and physical findings are similar to this particular patient. And suppose the physician knows from this experience that the frequency of cancer in this group is, say, 1 out of 100. Lacking any other information, the physician will therefore assign (perhaps subconsciously) a subjective probability of 1% to the event that this patient has cancer.

Now let the physician order a mammogram and receive a report that in the radiologist's opinion the lesion is malignant. This is new information and the actions taken will obviously depend on the physician's new estimate of the probability that the patient has cancer. A physician who turns to the literature can find innumerable helpful statements, such as the following: "The accuracy of mammography is approximately 90 percent" (Wolfe, 1966, p. 214); "In [a patient with a breast mass] a positive [mammogram] report of carcinoma is highly accurate" (Rosato, Thomas, & Rosato, 1973, p. 491); and "The accuracy of mammography in correctly diagnosing malignant lesions of the breast averages 80 to 85 percent" (Cohn, 1972, p. 98). If more detail is desired, the physician can find many statements like "The results showed 79.2 per cent of 475 malignant lesions were correctly diagnosed and 90.4 per cent of 1,105 benign lesions were correctly diagnosed, for an overall accuracy of 87 per cent" (Snyder, 1966, p. 217).

At this point you can increase your appreciation of the physician's problem by estimating for yourself the new probability that this patient has cancer: The physician thinks the lump is probably (99%) *benign*, but the radiologist has produced a *positive* X-ray report with the accuracy just given.

Table 1. *Accuracy of mammography in diagnosing benign and malignant lesions*

Results of X ray	Malignant lesion (cancer)	Benign lesion (no cancer)
Positive	.792	.096
Negative	.208	.904

Source: The numbers are from Snyder (1966).

Bayes' formula can be applied to assess the probability. This formula tells us that

$$P(\text{ca} \mid \text{pos}) = \frac{P(\text{pos} \mid \text{ca})\, P(\text{ca})}{P(\text{pos} \mid \text{ca})\, P(\text{ca}) + P(\text{pos} \mid \text{benign})\, P(\text{benign})}$$

where

$P(\text{ca} \mid \text{pos})$ is the probability that the patient has cancer, given that she has a positive X-ray report (the posterior probability)

$P(\text{pos} \mid \text{ca})$ is the probability that, if the patient has cancer, the radiologist will correctly diagnose it (the true-positive rate, or sensitivity)

$P(\text{ca})$ is the probability that the patient has cancer (prior probability)

$P(\text{benign})$ is the prior probability that the patient has benign disease [$P(\text{benign}) = 1 - P(\text{ca})$]

$P(\text{pos} \mid \text{benign})$ is the probability that, if the patient has a benign lesion, the radiologist will incorrectly diagnose it as cancer (the false-positive rate)

Table 1 summarizes the numbers given by Snyder. The entries in the cells are the appropriate probabilities (e.g., $P(\text{pos} \mid \text{ca}) = .792$).

Using 1% as the physician's estimate of the prior probability that the mass is malignant and taking into account the new information provided by the test, we obtain

$$P(\text{ca} \mid \text{pos}) = \frac{(0.792)\,(0.01)}{(0.792)\,(0.01) + (0.096)\,(0.99)} = 0.077$$

Thus, the physician should estimate that there is approximately an 8% chance that the patient has cancer.

Incorrect probabilistic reasoning

Unfortunately, most physicians (approximately 95 out of 100 in an informal sample taken by the author) misinterpret the statements about the accuracy of the test and estimate $P(\text{ca} \mid \text{pos})$ to be about 75%. Other investigators have observed similar results (Casscells, Schoenberger, & Grayboys, 1978). When asked about this, the erring physicians usually report that they assumed that the probability of cancer given that the patient has a positive X ray [$P(\text{ca}\,|\,\text{pos})$] was approximately equal to the probability of a positive X ray in a patient with cancer [$P(\text{pos}\,|\,\text{ca})$]. The latter probability is the one measured in clinical research programs and is very familiar, but it is the former probability that is needed for clinical decision making. It seems that many if not most physicians confuse the two.

There are really two types of accuracy for any test designed to determine whether or not a specific disease is present. The *retrospective accuracy* concerns $P(\text{pos}\,|\,\text{ca})$ and $P(\text{neg}\,|\,\text{no ca})$. (The abbreviation "no ca" refers to the event the patient does not have cancer. This can occur because she either has a benign disease or she has no disease at all.) This accuracy, the one usually referred to in the literature on mammography, is determined by looking back at the X-ray diagnosis after the true (histological) diagnosis is known. Let us use the term *predictive accuracy* to describe $P(\text{ca}\,|\,\text{pos})$ and $P(\text{benign}\,|\,\text{neg})$, the accuracy important to the clinician who has an X-ray report of an as yet undiagnosed patient and who wants to predict that patient's disease state.

Confusing retrospective accuracy versus predictive accuracy. A review of the medical literature on mammography reveals a strong tendency to equate the predictive accuracy of a positive report with the retrospective accuracy of an X-ray report; that is, to equate $P(\text{ca}\,|\,\text{pos}) = P(\text{pos}\,|\,\text{ca})$. There are many reasons to suspect that this error is being made. First, the wordings of many of the statements in the literature strongly suggest that the authors believe that the predictive accuracy [$P(\text{ca}\,|\,\text{pos})$] equals the retrospective accuracy [$P(\text{pos}\,|\,\text{ca})$] that they report in their studies. For example, a 1964 article in *Radiology* stated, "*the total correctness* of the X-ray diagnosis was 674 out of 759, or 89 percent" (vol. 84, p. 254). A contributor to *Clinical Obstetrics and Gynecology* in 1966 said, "Asch found a 90 percent *correlation* of mammography with the pathologic findings in 500 patients" (vol. 9, p. 217). "The *agreement* in radiologic and pathologic diagnosis was 91.6 percent" (Egan, 1972, p. 379). All of these statements imply that if the patient has a positive test the test will be correct and the patient will have cancer 90% of the time. This is not true.

Second, some authors make the error explicitly. The following appeared in a 1972 issue of *Surgery, Gynecology and Obstetrics* in an article entitled "Mammography in its Proper Perspective" and was intended to rectify some confusion that existed in the literature: "In women with proved carcinoma of the breast, in whom mammograms are performed, there is no X-ray evidence of malignant disease in approximately one out of five patients examined. If then on the basis of a negative mammogram, we are to defer biopsy of a solid lesion of the breast, then there is a one in five chance that we are deferring biopsy of a malignant lesion" (vol. 134, p. 98). The author has incorrectly stated that $P(\text{neg}\,|\,\text{ca}) = .2$ implies $P(\text{ca}\,|\,\text{neg}) = .2$. His error becomes very serious when he concludes that "to defer biopsy of a *clinically benign* solid lesion of the breast that has been called *benign* on mammography is to take a step backward in the eradication of carcinoma of the breast in our female population." The chance that such a patient has cancer depends on the prior probability, but is less than 1 in 100. His analysis is in error by more than a factor of 20.

Surgery, Gynecology and Obstetrics published in 1970 (vol. 131, pp. 93–98) the findings of another research group, who computed the "correlation of radiographic diagnosis with pathologic diagnosis" as follows. They took all the patients with histologically proven diagnoses and separated them into three groups on the basis of the X-ray diagnosis – "benign," "carcinoma," and "suspected carcinoma." In the "X-ray benign" ("negative" in our terminology) group, the tally showed that 84% in fact had benign lesions. It was also noted that 87.5% of the "X-ray carcinoma" (or "positive") group had biopsy-proven malignant lesions. Thus, $P(\text{ca}\,|\,\text{pos}) = 87.5\%$ and $P(\text{benign}\,|\,\text{neg}) = 84\%$. But the authors mistook this predictive accuracy for the retrospective accuracy. They stated that "A correct mammographic diagnosis was made in 84 percent *of those with benign lesions* and in 87.5 percent *of those with carcinoma*." In fact, the true-positive rate [$P(\text{pos}\,|\,\text{ca})$] in this study was actually 66% and the true-negative rate [$P(\text{neg}\,|\,\text{benign})$] was 54%.

In a letter to the editor in the September 11, 1976, issue of the *National Observer*, a physician presented five "observations and facts" to support his opinion that "routine [i.e., screening] mammography is not in the best interest of the population at large at any age." Here is the first set of observations.

> (1) The accuracy of the examination of mammography is reported to be between 80 percent and 90 percent, depending on such factors as the age of the patient, whether or not she has fibrocystic disease, the type of radiographic equipment, the experience of the radiologist, and what our definition of "accurate" is. . . . Even if we conclude that accuracy is 85 percent generally (and I am sure that not every radiologist in the nation can approach that figure in his own practice), then that means that *15 percent of the women X-rayed will wind up with incorrect interpretations of the findings, or more likely, their mammograms will simply fail to demonstrate the disease. This means that 15 percent of the women will be given a false sense of security if they are told their X-rays are normal, if indeed they already have cancer.* It is difficult to assess the harm done to this group, for they would obviously be better off with no information rather than with incorrect information. Told that her mammogram is normal and she need not come back for one more year, a woman with breast cancer may well ignore a lump in her breast which might otherwise send her to the doctor immediately.

There are several errors in this author's reasoning. First, the "accuracy" of mammography cannot be expressed as a single number. Assume the author means that the true-positive and true-negative rates both equal 85%.

Second, these rates (of 85%) are observed when mammography is used to make a differential diagnosis of known signs and symptoms. Such lesions are generally more advanced than the lesions being sought in a screening examination, which is the situation the author is addressing. More reasonable estimates for the true-positive and true-negative rates in screening programs are 60% and 98%, respectively.

Third, even using 85%, we find several inaccuracies in the reasoning. Consider the second sentence. There are two ways an incorrect interpretation can occur: (a) the patient can have cancer and a negative examination, $P(\text{ca,neg})$; or (b) she can have a positive examination but not have cancer, $P(\text{no ca,pos})$.[2] From elementary probability theory we know that

$$P(\text{ca,neg}) = P(\text{neg} \mid \text{ca})\, P(\text{ca})$$

$P(\text{neg} \mid \text{ca})$ is the complement of $P(\text{pos} \mid \text{ca})$ and therefore equals .15 in this case. We do not know $P(\text{ca})$ precisely, but for a screening population we are reasonably certain that it is less than .005. That is, fewer than 5 out of 1,000 women harbor an asymptomatic but mammogram-detectable cancer of the breast.

Thus,

$$P(\text{ca,neg}) \le (.15)\,(.005) = .00075$$

Also,

$$P(\text{no ca,pos}) = P(\text{pos} \mid \text{no ca})\, P(\text{no ca}) \ge (.15)\,(.995) = .14925$$

The total probability of an incorrect interpretation [i.e., $P(\text{ca,neg}) + P(\text{no ca,pos})$] is the sum of these two numbers, which is 15%, as the author states. However, this does not mean that "more likely, their mammograms will simply fail to demonstrate the disease." $P(\text{ca,neg}) = .00075$ is *not* more likely than $P(\text{no ca,pos}) = .14925$. It is about 200 times *less* likely.

Another problem is that 85% "accurate" does not mean that "15 percent of the women will be given a false sense of security if they are told their X-rays are normal." The author appears to be trying to estimate $P(\text{ca} \mid \text{neg})$. Now by Bayes' formula,

$$P(\text{ca} \mid \text{neg}) = \frac{P(\text{neg} \mid \text{ca})\, P(\text{ca})}{P(\text{neg} \mid \text{ca})\, P(\text{ca}) + P(\text{neg} \mid \text{no ca})\, P(\text{no ca})}$$

$$= \frac{(.15)(.005)}{(.15)(.005) + (.85)(.995)} = .00089$$

That is, if 10,000 asymptomatic women are screened, and if we use the author's misestimate of the accuracy, 8,458 of them will leave with a

[2] $P(A,B)$ is the joint probability that both events A and B occur.

Table 2. *Presence of cancer and results of X rays in 1000 women who have abnormal physical examinations*

	Women with cancer	Women with no cancer	Total
Women with positive X rays	74	110	184
Women with negative X rays	6	810	816
Total	80	920	1,000

Note: A true-positive rate of .92 ($P(\text{pos}|\text{ca}) = 0.92$) implies that of 80 women who have cancer, 74 will have positive X rays and 6 will have negative X rays. Of all the women with positive X rays, 74/184 have cancer, or $P(\text{ca}|\text{pos}) = 74/184 = 40\%$.
Source: The numbers are from Wolfe (1964).

negative examination. The author thinks that about 1,269 of them will have a false sense of security. In fact, only about 9 will. This number has been overestimated by a factor of about 150.

Finally, adding the phrase, "if indeed they already have cancer" further confuses the meaning of the sentence. The phrases "a false sense of security," "if [given] they are told their X-rays are normal," and "if they already have cancer" translate symbolically into $P(\text{ca} \mid \text{neg,ca})$. This probability is 1, not .15.

The importance of P(ca). In addition to confusing the two accuracies, many authors do not seem to understand that, for a test of constant retrospective accuracy, the meaning to the physician of the test results (the predictive accuracy) depends on the initial risk of cancer in the patient being mammogrammed. Even if it is assumed that the true-positive and true-negative rates are constant for all studies, the proper interpretation of the test results – the chance that a patient with a positive (or negative) mammogram has cancer – will depend on the prevalence of cancer in the population from which the patient was selected, on the pretest probability that a patient has cancer. This can be extremely important when one compares the use of the test in a diagnostic clinic (where women have signs and symptoms of breast disease) with its use in a screening clinic for asymptomatic women.

The importance of this is shown by an example. Suppose a clinician's practice is to mammogram women who have an abnormal physical examination. The frequency of cancer in such women has been found in one study to be approximately 8% (Wolfe, 1964). In one series of mammograms in this population, a true-positive rate of 92% and a true-negative rate of 88% was obtained (Wolfe, 1964). Let the physician now face a patient who he feels is representative of this sample population (i.e., let $P(\text{ca}) = 8\%$). Suppose he orders a mammogram and receives a positive result from the radiologist. His decision to order a biopsy should be based on the new probability that the patient has cancer. That probability can be calculated to be 40% (see Table 2). Would a negative report have ruled out cancer? The probability that this woman, given a negative report, still has cancer is slightly less than 1%. The logic for this estimate is shown in Table 2.

Now, suppose the clinician orders the test to screen for cancer in a woman who has no symptoms and a negative physical examination. The prevalence of mammography-detectable cancer in such women is about .10% (e.g., Shapiro, Strax, & Venet, 1967). For the purposes of this example, let the retrospective accuracy of the radiologist be unchanged – that is, in this population of patients let him again have a true-positive rate of 92% and a true-negative rate (for the diagnosis of benign lesions) of 88%.[3] The literature provides data only on the retrospective accuracy of the test in women who have cancer and benign diseases. In one study about 60% of these women had no disease at all (Wolfe, 1965). Thus, in this case,

$$P(\text{ca} \mid \text{pos}) = [P(\text{pos} \mid \text{ca})P(\text{ca})] \div [P(\text{pos} \mid \text{ca})P(\text{ca}) + P(\text{pos} \mid \text{benign})P(\text{benign}) + P(\text{pos} \mid \text{no disease})P(\text{no disease})]$$

$P(\text{benign})$, $P(\text{no disease})$, and $P(\text{pos} \mid \text{no disease})$ are not discussed explicitly in the literature. This is instructive and it leads us to suspect that their importance in the analysis of these problems is not understood. For this example, we shall use the data presented by Wolfe (1965) and assume that $P(\text{no disease})$ is about 60% and $P(\text{benign})$ is about 40%. We shall also make an assumption favorable to mammography and let $P(\text{pos} \mid \text{no disease})$ be 0%.

To continue with this example, say the radiologist reports that the mammogram in this asymptomatic woman is positive. Given the positive mammography report, the probability that the patient has cancer ($P(\text{ca} \mid \text{pos})$) is about 1 out of 49, or about 2.0% (Table 3). In the previous example that involved women with symptoms, $P(\text{ca} \mid \text{pos})$ was 40%. Thus, depending on who is being examined, there can be about a *twentyfold difference* in the chance that a woman with a positive mammogram has cancer.

This raises a major question about medical reasoning – when trying to evaluate a patient's signs and symptoms, how should a physician use information about the basic frequency of the possible diseases in the

[3] This is not a good assumption, since the "accuracy" changes as the population being examined changes. For example, the true-positive rate is lower when one is using the test in an asymptomatic population because the cancers tend to be much smaller and harder to detect. The assumption is made only to demonstrate the importance of $P(\text{ca})$.

Table 3. *Presence of cancer and results of X ray in 1,000 women who have no symptoms*

	Women with cancer	Women with benign lesions	Women with no cancer	Total
Women with positive X rays	1	48	0	49
Women with negative X rays	0	352	599	951
Total	1	400	599	1,000

Note: A true-positive rate of 0.92 implies that the X ray will detect cancer in the one woman who has the disease. A true-negative rate of 0.88 for benign disease implies that of 400 women with benign disease, 352 will have negative X rays, whereas in 48 the X ray will be positive. Thus, 49 women will have positive X rays, but only one has cancer, or $P(\text{ca} \mid \text{pos}) = 1/49 = 2\%$.

population at large? The profession appears to be confused about this issue. On the one hand, physicians make statements that the relative commonness of a disease should *not* affect the estimate of the probability that a particular patient has the disease. This notion appears in several maxims, such as, "The patient is a case of one" and, "Statistics are for dead men." In discussions of specific problems, the idea is sometimes expressed subtly as in the statement, "The younger women obviously have a fewer number of the malignancies which, however, should exert very little influence on the individual case" (Wolfe, 1967, p. 138). It can also be stated explicitly and presented as a rule to be obeyed. For example, the following appeared in a textbook on clinical diagnosis: "When a patient consults his physician with an undiagnosed disease, neither he nor the doctor knows whether it is rare until the diagnosis is finally made. Statistical methods can only be applied to a population of thousands. The individual either has a rare disease or doesn't have it; the relative incidence of two diseases is completely irrelevant to the problem of making his diagnosis" (DeGowin & DeGowin, 1969, p. 6).

On the other hand, these statements are often inconsistent with the behavior of physicians who try, however imperfectly, to use this diagnostic information. Witness the following maxims that are passed on in medical schools: "When you hear hoofbeats, think of horses not of zebras," "Common things occur most commonly," "Follow Sutton's law: go where the money is," and so forth. It appears that many physicians sense the value of information on the prior probability of a disease but that the formal lessons of probability theory are not at all well understood. Without a formal theory, physicians tend to make the same kinds of errors in probabilistic reasoning that have been observed in other contexts (Kahneman & Tversky, 1973, 4; Lyon & Slovic, 1976).

Implications: Mammograms and biopsies

These problems can have important practical implications. For instance, in the examples just cited two authors based their conclusions on incorrect probabilistic reasoning. One incorrectly argued that a woman with a breast mass that appears benign on physical examination and benign on the X ray still has a 20% chance of having cancer and recommended that she be biopsied. Another author based a recommendation against screening on a gross misestimate of the frequency with which women would have a false sense of security (i.e., have a cancer missed by the mammogram). Both authors may have come to the same conclusion with correct reasoning, but they may not have.

The value of diagnostic information. The value of mammography in women who have symptoms and signs of breast disease lies in its ability to provide diagnostic information that will affect the clinician's decision to biopsy. More precisely, the outcome of the test should change a clinician's estimate of the probability that the patient has cancer. As one author puts it:

> Mammography can assist the clinician in differentiating between benign and malignant lesions. . . . Some lesions, especially the small ones, may lack the characteristics that give the clinician an index of suspicion high enough to justify biopsy. It is here that the . . . mammogram may provide additional objective evidence. Thus, in the case of an indeterminate lesion of the breast, mammography can aid the physician in deciding whether to perform a biopsy study (Clark & Robbins, 1965, p. 125).

For any diagnostic test to be useful it must provide information that can potentially change a decision about how the patient should be managed – to call for a biopsy in some patients who would otherwise not be biopsied, and, we should hope, to obviate biopsies in some women who would otherwise receive them. This notion is developed formally in statistical decision theory and has been used to analyze some medical problems in a research setting (e.g., Lusted et al., 1977).

Many physicians recognize that the X-ray report carries useful information that should help in patient management, but precisely how the information should be used is ordinarily not stated. The explanations given by most authors contain few specific directions. "Mammography is not designed to dictate treatment procedures but may provide, in certain cases, just that bit of more precise information, so that undesirable sequelae are avoided" (Egan, 1972, p. 392). "Mammography is a valuable adjunctive to the surgeon in the diagnosis and treatment of breast lesions" (Lyons, 1975, p. 231). "Mammography may assist in clarifying confusing palpable findings" (Egan, 1969, p. 146). It "plays a supportive or auxiliary role . . ." (Block & Reynolds, 1974, p. 589). The precise nature and degree of the support is usually left to the clinician's judgment.

Mammograms and biopsies: The practice. It seems that the role of mammography in such cases is only partially understood. To understand this, let us examine the impact that clinical investigators predict mammography will have on the need to biopsy diseased breasts. While the statements quoted above imply that the use of X rays should help select patients for biopsy, an equal number of statements suggest that mammography cannot, indeed should not, perform this function. "Any palpable lesion requires verification by excision and biopsy regardless of the X-ray findings" (Lesnick, 1966, p. 2007). "While mammography is usually definitive it is not a substitute for biopsy" (Egan, 1969, p. 148). "In no way would this procedure detract from the importance of biopsy. As a matter of fact, the use of routine mammography will reaffirm the importance of biopsy, since X-ray evidence of a malignant lesion requires biopsy for confirmation. . . . It in no way detracts from the importance of the biopsy. . . . [B]iopsy is as much a necessity for the confirmation of X-ray findings as it is for the confirmation of physical signs" (Gershon-Cohen, & Borden, 1964, pp. 2753, 2754). "It is apparent that mammography is not a substitute for surgery" (DeLuca, 1974, p. 318). "Let us state emphatically that mammography is not a substitute for biopsy" (McClow & Williams, 1973, p. 618).

One of the most precise policy statements on how mammography should be used to help select patients for biopsy appeared in *Archives of Surgery* in 1966 (vol. 93, pp. 853–856). A careful examination of the directions reveals that only half of the test's potential is used. The scheme for using mammography "to determine the treatment or disposition of each patient" involves three categories of patients:

Category A: "The patients with a 'lump' or 'dominant lesion' in the breast are primarily surgical problems and there should be no delay in obtaining a biopsy. Mammography, in this instance, is strictly *complementary*. . . . It may disclose occult tumors" (p. 854).

Category B: "The patients have symptoms referable to the breast but no discrete mass or 'dominant lesion'. . . . In this category, the surgeon and clinician will find the greatest yield from mammography because here the modality is *confirmatory*." Here the mammogram will give confirmation and encouragement, "if the clinical impression is benign. It should not, however, dissuade him from a prior opinion to biopsy" (p. 855).

Category C: These patients have no signs or symptoms, there are no clinical indications for biopsy, and a mammogram can only increase the number of biopsies.

Thus, the author has outlined a plan that nullifies the value of mammographic information in selecting patients in whom a biopsy can be avoided. Only the added bit of information that implies biopsy is used. The information that might eliminate a biopsy is ignored.

Mammograms and biopsies: The potential. To appreciate how problems in probabilistic reasoning can affect the actual delivery of medical care, let us now examine the role that mammography might play in differential diagnosis and in the selection of patients for biopsy. As described above, the purpose of the test is to change the decision maker's subjective estimate of the chance that a patient has cancer. If that probability is high enough (as determined by the physician and patient), biopsy is recommended. Call this probability the *biopsy threshold*.[4] Now consider the impact of the test on the management of two groups of patients.

The first group consists of those patients who, on the basis of a history and physical examination, are thought by the clinician to have clinically obvious cancer. Using data published by Friedman et al. (1966), let the prior probability (the frequency) of cancer in this group be 90%. If a mammogram were performed on such a patient, a positive result would increase the probability of cancer [$P(\text{ca}\mid\text{pos})$] to perhaps 95%. A negative mammogram would still leave the patient with a 71% chance of having cancer. This high probability is the motivation of such statements as: "If the subjective impression of the clinician gives enough reason for suspicion of cancer, the clinician will be compelled to biopsy" (Clark et al. 1965, p. 133). A 71% chance of malignancy is still high enough that almost anyone would want to be biopsied.

Now consider a second group of patients who have a dominant mass that is not obviously carcinoma. In one study the probability that such a mass is malignant was 14% (Friedman et al., 1966). In the absence of further information, the clinical policy in such cases is to biopsy the lesion: "If a dominant lump develops, it should be removed and examined microscopically" (del Regato, 1970, p. 861). Using this as a guideline, let us suppose that the patient's biopsy threshold is 10%. That is, if, to the best of the physician's knowledge, the probability that his patient has cancer is above 10% then the patient and physician agree that a biopsy should be done.[5] Using a biopsy threshold of 10%, we can determine the

4 Anyone needing to be convinced of the existence of a biopsy threshold can reason as follows. Can we agree that no one is willing to be biopsied if the chance of cancer is 1 in 30 trillion? And can we agree that virtually everyone wants to confirm the diagnosis and be treated if that chance is 98 in 100? If so, then somewhere between 1 in 30 trillion and 98 in 100 everyone has a biopsy threshold. Of course if a woman refuses biopsy and treatment even when the chance of cancer is certain, then she has no threshold.

5 The biopsy threshold is a very fascinating and important number. Shifting it exerts great leverage on the number of women biopsied, the frequency of unproductive biopsies, the cost of managing a patient, as well as a patient's prognosis. Because of risk aversion and the fact that they are making the decision for someone else, physicians generally set the biopsy threshold quite low. The statement "if there is *any* chance that the lesion is malignant, a biopsy should be done" is typical. "If the physician is not completely satisfied that the lesion is benign, it should be biopsied without delay" (Allen, 1965, p. 640). There is evidence that women themselves generally set the threshold higher than do physicians – although there is wide variation.

For example, we can examine data from a large clinical trial in which mammography and a breast physical examination were used to screen asymptomatic women for breast cancer (Shapiro, Strax, & Venet, 1971). Depending on how the breast lesion was detected (i.e., by which test or combination of tests), the probability that a woman's breast disease was cancer

impact of a mammogram on the management of 1,000 such patients. Without the test, all patients would have to be biopsied, 860 of them unproductively. The approximate fate of the original 1,000 patients with a dominant lesion when mammography is used is presented in Figure 1.[6]

Patients with positive mammograms have a 53% chance of having cancer and, since we have assumed they have a biopsy threshold of 10% they should be biopsied. Because the probability is 34% that a patient with an uncertain mammogram has cancer, these patients should also be biopsied. Patients with a negative mammogram have a 4% chance of having cancer, and, since this is below their assumed biopsy threshold (10%), they would not want to be biopsied but would prefer to be followed closely. The total number of immediate biopsies has been reduced from 1,000 to 240. At least 30 more biopsies will have to be done eventually because 30 of the 760 remaining patients have cancer.

In this way, the expected benefits from having a mammogram (such as a reduction of the chance of an unnecessary biopsy from approximately 86% to a little over 13%) can be compared with the costs (e.g., a radiation hazard and about $75), and the slight decrease in expected survival (there is a 3%

varied from 15% to 54%. On the basis of a positive physical examination, physicians recommended that 545 women who had negative mammograms be biopsied. Despite the fact that the frequency of cancer in this group was 15%, 31% of the women declined the recommended biopsy. The frequency of cancer in women who had a positive mammogram and a negative breast physical examination was 20%, but 29% of the women in this group declined a recommended biopsy. In women who had positive results on both tests, the frequency of cancer was 54% and only 5% of these women preferred not to be biopsied at the recommended time. Thus, from this crude information it appears that about 31% of women had a biopsy threshold greater than 15%, 29% of women had a biopsy threshold greater than 20%, and in 5% of women the threshold exceeded 54%.

6 To sketch the impact of mammography on these patients (and the patients with other signs and symptoms) much information is needed that is not directly available in the literature. It is fortunate that in one study (Friedman et al., 1966) the data on the frequency of cancer and the retrospective accuracy of mammography are presented separately for three groups of patients – those with obvious carcinoma, those with a dominant mass, and patients with other signs and/or symptoms of breast disease. The published data are incomplete, however, and the data on the frequency of an uncertain X-ray diagnosis in benign and malignant lesions are not included. The data available in the Friedman study were used, and for this example the following assumptions were made: (1) Lesions not biopsied were in fact benign, (2) lesions not biopsied were coded negative, (3) half of the benign lesions that were not coded negative were coded positive (the other half being coded uncertain), and (4) half of the malignant lesions that were not coded positive were coded negative. The first two assumptions are the most optimistic interpretation of mammography's accuracy. The third and fourth assumptions are very important and as the false-positive (or false-negative) rate tends toward zero, the power of a positive (negative) X-ray report to rule cancer in (out) increases. Likewise, as the false-positive or false-negative rates increase, the test loses its predictive power. Interpretation of Friedman's data is made even more difficult by its presentation in terms of breasts rather than patients. Nonetheless, there is much information in this report and it is reasonable to use it in this example provided the reader understands that this is an illustration, not a formal analysis. A formal analysis of these questions would require better data. The figures for the accuracy used in the text for the evaluation of the patients in group 2 are as follows: $P(\text{pos}\mid\text{ca}) = .52$, $P(\text{uncertain}\mid\text{ca}) = .24$, $P(\text{neg}\mid\text{ca}) = .24$, $P(\text{pos}\mid\text{benign}) = .075$, $P(\text{uncertain}\mid\text{benign}) = .075$, and $P(\text{neg}\mid\text{benign}) = .85$.

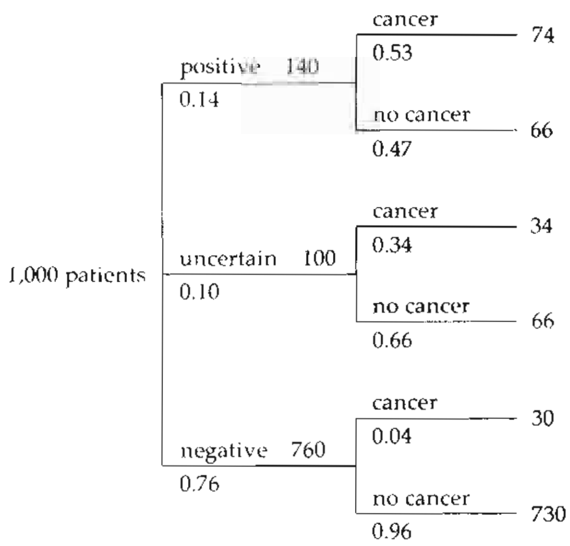

Figure 1. Probability of cancer in women with dominant lesions.

chance that diagnosis of a malignant lesion will be postponed a month or so). If the notion of a biopsy threshold and some simple probability theory were used, many patients in this group who had negative mammograms would be spared a biopsy. In the absence of this type of analysis "the surgical consensus here is that all patients [in this group] should have a biopsy, *regardless of mammographic findings*" (Friedman et al., 1966, p. 889).

The importance of the biopsy threshold in this example should be stressed. If the physician and his patient had set the threshold at 1% – that is, if the patient felt that a 1 in 100 chance of having cancer was sufficient to warrant a biopsy – then a negative mammogram report would not have eliminated the need for the biopsy (a 4% chance of cancer would exceed this threshold). The mammogram may have given the clinician some information but this information would not have contributed to the decision to biopsy. Use of mammography in this case would have to be justified on other grounds.

The practice revisited. This type of analysis helps make clear the potential usefulness of mammography in the differential diagnosis of various lesions. It also helps us evaluate the following policy statements:

1. "Mammography adds little to the management of the clinically [i.e., physically] palpable breast nodule that, on the basis of its own characteristics, requires biopsy" (from *Archives of Surgery*, 1974, vol. 108, p. 589). In the study of the patients with a dominant mass, biopsy was required on clinical grounds alone. The use of mammography split the group into subgroups with frequencies of cancer ranging from 53% to 4%. Biopsy might be avoided in the latter group and the number of biopsies might be reduced 73% (from 1,000 per 1,000 to 270 per 1,000).

2. "For clinical purposes mammography must provide accuracy at approximately the 100 percent level before it alone can direct management" (from *Archives of Surgery*, 1974, vol. 108, p. 589). In a population like the second group discussed above, it might be quite rational to let mammography select patients for biopsy. Recall that the true-positive rate used in that example was 52% and that a more accurate test would be even more valuable.

3. "Mammography is not a substitute for biopsy" (from *Oncology*, 1969, vol. 23, p. 148). The purpose of both mammography and biopsy is to provide information about the state of the patient. Some patients, in the absence of mammography, require biopsy. In some of these patients a negative mammogram would obviate the biopsy, and in these cases the mammogram would replace the biopsy.

4. "Every decision to biopsy should be preceded by a mammogram" (from *Oncology*, 1969, vol. 23, p. 146). Consider clinically obvious carcinoma. The probability of cancer will be above almost anyone's biopsy threshold no matter what the outcome of the mammogram. The primary justification for this policy in such a case must lie in the chance that the clinically obvious is benign (otherwise the patient would have to have a mastectomy [breast removal] anyway) *and* that there is a hidden, nonpalpable, malignant lesion. The probability of this compound event is the product of the probabilities of the two events, which is extremely small (on the order of 1 out of 5,000).

5. "To defer biopsy of a *clinically benign* lesion of the breast which has been called *benign on mammography* is to take a step backward in the eradication of carcinoma of the breast" (from *Surgery, Gynecology and Obstetrics*, 1972, vol. 134, p. 98). Let "clinically benign" be represented by a P(ca) of 5%. After a negative mammogram, the probability that such a patient has cancer is approximately 1%. Out of 100 biopsies, 99 would be unproductive. Is the deferral of biopsy here a step backward or forward? The other point is that if the policy were followed, all lesions from "clinically benign" through clinically obvious carcinoma would require a biopsy no matter what the outcome of the test was. This seems to contradict the author's statement that "when used in its proper perspective, mammography is an excellent adjunct to the physician in the management of carcinoma of the breast" (from *Surgery, Gynecology and Obstetrics*, 1972, vol. 134, p. 98).

6. "Mammography must never be used instead of biopsy when dealing with a 'dominant lesion' of the breast and should never change the basic surgical approach in breast diseases, i.e., a 'lump is a lump' and must be

biopsied either by incision or aspiration" (from *Archives of Surgery*, 1966, vol. 93, p. 854). Patients with dominant lesions and biopsy thresholds over 5% would disagree with this statement.

7. "The fallacy comes in relying on [mammography] in doubtful cases. It is essential after examining and palpating the breast to decide whether you would or would not do a biopsy if X-ray were not available. If you would do a biopsy, then do it. If you are sure there is no indication for surgery or physical examination, then order a mammogram. As soon as one says to himself, and particularly if he says to a patient, 'I am not quite sure about this – let's get an X-ray,' one unconsciously has committed himself to reliance on the negativity of the mammogram, when one should only rely on positivity. This is a psychological trap into which we all tend to fall and is much more serious than a certain number of false-positive diagnoses reached with mammography" (Rhoads, 1969, p. 1182). Not a single biopsy will be avoided by this policy. This is a shame because, as the author of the above statement himself puts it, "there are few areas in which so much surgery is necessitated which could be avoided by better methods of diagnosis than the breast."

We are now in a position to appreciate the following story that appeared in the San Francisco *Chronicle* (Kushner, 1976). A woman reporter had just discovered a mass in her breast and described a consultation with her physician.

"I'd like you to get a xeromammogram. It's a new way to make mammograms – pictures of the breasts."

"Is it accurate?"

He shrugged, "Probably about as accurate as any picture can be. You know," he warned, "even if the reading is negative – which means the lump isn't malignant – the only way to be certain is to cut the thing out and look at it under a microscope."

The woman then discussed the problem with her husband.

"What did the doctor say?"

"He wants to do a xeromammogram. Then, whatever the result is the lump will have to come out."

"So why get the X-ray taken in the first place?"

"It's something to go on, I guess. And our doctor says it's right about 85 percent of the time. . . . So, first I've scheduled an appointment to have a thermogram. If that's either positive or negative, and if it agrees with the Xerox pictures from the mammogram, the statistics say the diagnosis would be 95 percent reliable."

In summary, it would seem reasonable to ask that if the purpose of mammography is to help physicians distinguish benign from malignant breast disease, thereby sparing some patients a more extensive and traumatic procedure such as a biopsy, then we ought to let the test perform that function. If on the other hand the physician should always adhere to a prior biopsy decision and be unmoved by the mammogram outcome, then we ought not to claim that the purpose of the test is to help distinguish benign from malignant disease, since that distinction will be made definitively from a biopsy. Finally, if the purpose of the test is to search for hidden and clinically unsuspected cancer in a different area of the breast (away from a palpable mass that needs biopsy anyway), we ought to recognize explicity that the chances of such an event are extremely small and that the use of the test amounts to screening.

My purpose is not to argue for a specific mammography or biopsy policy – to do so would require better data and a better assessment of patient values. It is to suggest that we have not developed a formal way of reasoning probabilistically about this type of problem, that clinical judgment may be faulty, and that current clinical policies may be inconsistent or incorrect.

Discussion

These examples have been presented to illustrate the complexity of medical decision-making and to demonstrate how some physicians manage one aspect of this complexity – the manipulation of probabilities. The case we have studied is a relatively simple one, the use of a single diagnostic test to sort lesions into two groups, benign and malignant. The data base for this problem is relatively good. The accuracy and diagnostic value of the test has been studied and analyzed in many institutions for many years. As one investigator put it, "I know of no medical procedure that has been more tested and retested than mammography" (Egan, 1971, p. 1555).

The probabilistic tools discussed in this chapter have been available for centuries. In the last two decades they have been applied increasingly to medical problems (e.g., Lusted, 1968), and the use of systematic methods for managing uncertainty has been growing in medical school curricula, journal articles, and postgraduate education programs. At present, however, the application of these techniques has been sporadic and has not yet filtered down to affect the thinking of most practitioners. As illustrated in this case study, medical problems are complex, and the power of formal probabilistic reasoning provides great opportunities for improving the quality and effectiveness of medical care.

Judgment under uncertainty: Heuristics and biases

Edited by

Daniel Kahneman
University of British Columbia

Paul Slovic
Decision Research
A Branch of Perceptronics, Inc.
Eugene, Oregon

Amos Tversky
Stanford University

The right of the University of Cambridge to print and sell all manner of books was granted by Henry VIII in 1534. The University has printed and published continuously since 1584.

Cambridge University Press

Cambridge
London New York New Rochelle
Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia

First published 1982
Reprinted 1982, 1983 (Twice), 1984

Printed in the United States of America

Library of Congress Cataloging in Publication Data
Main entry under title:
Judgment under uncertainty.
Bibliography: p.
Includes index.
1. Judgment. I. Kahneman, Daniel, 1934–
II. Slovic, Paul, 1938– III. Tversky, Amos.
BF441.J8 153.4'6 81-10042
ISBN 0 521 24064 6 hard covers AACR2
ISBN 0 521 28414 7 paperback

Contents