Diagnosis by the Bit: A Method for Evaluating the Diagnostic Process*

HORTON A. JOHNSON, M.D.

Departments of Pathology, 
St. Luke's-Roosevelt Hospital Center 
and 
Columbia University, 
College of Physicians and Surgeons, 
New York, NY 10019

ABSTRACT

The extent to which a diagnostic observation or test contributes to a diagnosis is usually represented by a conditional, or Bayesian, probability. According to information theory, the contribution of a diagnostic observation can also be measured in bits of information. This offers a representation of the Bayes rule which may be useful in comparing and evaluating diagnostic sequences, since the diagnostic contributions of the various tests, expressed as bits of information, are additive, providing a simple and graphic representation of the diagnostic process.

“Medicine is a science of uncertainty and an art of probability.”

—Sir William Osler

In Osler's day playing the diagnostic odds was mostly intuitive and constituted part of the art of medicine. Recently, driven by a need to estimate cost-effectiveness and also perhaps by our technological infatuation with numbers, there has been a growing interest in quantifying Osler's odds. The usual approach to quantifying the utility of a test or observation is a probabilistic one. One estimates the increase in the probability of a given disease in the light of a new observation. This Bayesian probability, often called the predictive value of the test or observation, was introduced into medicine in 1966 by Vecchio, but the method of calculation has been in use for many years. It is based upon an algebra devised by Thomas Bayes,† and pub-

† This famous essay was found among Bayes's papers after his death. Perhaps the Reverend Thomas Bayes, Nonconformist minister, took his work less seriously than we do today. He claimed that "So far as Mathematics do not tend to make men more sober and rational thinkers, wiser and better men, they are only to be considered as an amusement, which ought not to take us off from serious business."
lished posthumously in 1763.\(^3\) The predictive value, which can be easily estimated by using a nomogram,\(^5\) has become the standard measure of the effectiveness of a diagnostic test or procedure. However, the evaluation of a diagnostic protocol made up of a sequence of tests, requires the compounding of probabilities.

Information theory, which has long been a useful tool in biology,\(^14,15\) offers another way of representing the Bayes rule. By converting probabilities into bits of information, which are additive, it can provide a graphic representation of the diagnostic process that may be useful for comparing alternative diagnostic sequences.

**Quantifying Information**

In the process of diagnosis, information about the patient is accumulated in a stepwise fashion. One begins with a degree of uncertainty about an eventual diagnosis and works toward reducing that uncertainty to zero. Each step, each new piece of information, obliterates an amount of uncertainty until so little uncertainty remains that the physician is willing to accept the diagnosis as a premise for therapy. In information theory, both information and uncertainty can be measured in bits. A negative quantity of information signifies uncertainty, a lack of information. The physician, beginning with an uncertainty of \(-n\) bits of information, must compile nearly \(n\) bits of diagnostic information in order to remove most of those \(-n\) bits of uncertainty. Since bits of information are additive, it is a fairly straightforward matter to estimate the cumulative information contributed by each diagnostic step. The number of bits of information provided by different diagnostic sequences can thus be compared and evaluated.

The microcomputer has made the bit a household unit for the measurement of information, but the probabilistic nature of the bit may not be generally appreciated.

How does one quantify the information provided by a clinical finding or the digital readout of a measuring device or the printout of a laboratory report? It is intuitive that information is somehow related to probability, to the amount of surprise it evokes in the recipient of that information. The improbable cytological diagnosis of bronchogenic carcinoma in an apparently healthy individual would seem to provide more bits of information than the very same report on a patient already known to have the disease. The amount of information in a message is somehow related to the probability of that message. In 1927, the physicist Leo Szilard\(^26\) showed that information should, like entropy, be expressed as the logarithm of probability. Szilard used the natural logarithm, which allowed him to measure information in entropy units.

The bit or binary unit of information is due to Norbert Wiener, who, in 1948 in his book *Cybernetics*, proposed that the most useful unit of information would be the “recording of a choice between two equally probable simple alternatives as, for example, between heads and tails in the tossing of a coin.”\(^28\) By definition each new bit of diagnostic information must double the probability of a given diagnosis and, conversely, each bit of missing information halves the probability of a given diagnosis. If an initial probability, \(p'\), represents an initial deficit of \(-I\) bits of information, then \(p = (\frac{1}{2})^{-1}\) and the initial uncertainty, \(I'\), is given by

\[
I' = \log_2 p \quad \text{Eq. 1}
\]

Thus the logarithm to the base two of the probability of a diagnosis gives its uncertainty measured in bits of information. If, in a defined population, the prevalence of a certain disease is 12.5 percent, the a priori probability of a patient’s having that disease is one out of eight, and
I' = log₂ (pₐ) = -3 bits

The uncertainty represents an information deficit of -3 bits, which means that three bits of new diagnostic information must be added in order to reduce the uncertainty to zero bits, at which point the diagnosis will be certain.

The information provided by a diagnostic observation or test, ΔI, is the difference between the diagnostic information on hand after the test, I", and that available before the test, I', so that

ΔI = I" - I'

If the test result is at all useful, ΔI will be a positive number, signifying an increase in information and a reduction in uncertainty. If the uncertainty of a given diagnosis was -6 bits before a diagnostic observation and is -2 bits after the observation, the net change is (-2) - (-6) = +4 bits of information. The procedure has reduced the uncertainty by four bits and, correspondingly, has increased the information about the patient's diagnosis by four bits.

It is also possible to estimate the informational value of a diagnostic observation or test from the sensitivity and specificity of that test. However, it must be understood from the outset that any evaluation of clinical testing based upon sensitivity and specificity is subject to two serious constraints. First, the measurements of sensitivity and specificity depend upon the definition of "positivity," which can be rather arbitrary at times. Secondly, sensitivity and specificity are based upon a binary system of measurement: the result of a test is either positive or negative. Most of the results we deal with lie on a continuum. In a system based upon sensitivity and specificity, a blood pressure measurement is either elevated or not elevated, and all elevated measurements will make the same diagnostic contribution. But in reality, two elevated results may contribute very different amounts of information toward a diagnosis. Systolic arterial pressures of 210 and 165 mm Hg may be considered equally "positive" for measuring sensitivity and specificity, even though according to information theory the more improbable pressure of 210 must provide more information than the more probable pressure of 165.

From Bayesian algebra it can be shown that if the prior uncertainty was I', the amount of incremental information, ΔI, provided by a positive test result is

ΔI = log₂ \left( \frac{Sensitivity}{(2^r)} + \frac{1-Specificity}{(1-2^r)} \right) \text{ bits.} \quad \text{Eq. 2}

The numerical value of the informational contribution, ΔI, like the predictive value, gives an estimate of the usefulness of a test in a particular clinical context but cannot be used for evaluating or comparing tests per se, since it depends not only upon the effectiveness of the test but also upon the prior probability of the disease. One can, however, derive from ΔI a measure of test efficiency that is independent of the prior probability. The informational value of a positive test result increases as the initial uncertainty increases. As I' approaches minus infinity, 2^r approaches zero, and, from Eq. 2, ΔI approaches its maximum value of

I_{max} = \log₂ \frac{Sensitivity}{1 - Specificity} \text{ bits.} \quad \text{Eq. 3}

I_{max} is the maximum information obtainable from a positive result of a test, given the sensitivity and specificity of that test. This measure, which turns out to be the information content of the so-called "likelihood ratio," is suitable for comparing the usefulness of tests, since it depends only upon the sensitivity and specificity of the test indepen-
dently of any clinical context. In the extreme case, such as finding a malarial parasite in an erythrocyte, which is pathognomonic of the disease, there are no false positives. The specificity approaches unity, the denominator of Eq. 3 approaches zero, and the value of $I_{\text{max}}$ even though the sensitivity may be poor, approaches infinity. This confirms what we already knew: that a pathognomonic observation can erase an infinite amount of uncertainty.

It is important to notice that according to Eq. 2 the value of $\Delta I$ decreases as $I'$, the prior uncertainty, decreases. This points out the law of diminishing returns in diagnosis: the closer one approaches certainty of diagnosis the less the information provided by a given test. It demonstrates the futility of diagnostic overkill.

Some Clinical Applications

Measured values of sensitivity and specificity are available for many clinical observations and test procedures. In this discussion, the values of the sensitivity and specificity will appear in that order within parentheses following the name of the test or observation. They will be given in decimal notation rather than in percentages. Again, one must remember that these values are not absolute but depend upon the definition of positivity, which may vary with the observer, the method, and the cutoff point at which a test result is considered to be positive.\(^{12}\)

The informational contributions of diagnostic procedures have been calculated from Eq. 2, using a programmable pocket calculator.

Heart Disease

One of the earliest studies of the sensitivity and specificity of clinical observations was that of Kurlander and others who, in 1952, evaluated four commonly used screening tests for heart disease and hypertension.\(^{17}\) In the population they screened, 233 of 690 were ultimately shown to have either hypertension, heart disease, or both. The probability of hypertension and/or heart disease in any member of that population was 0.338. From Eq. 1, the initial uncertainty of heart disease and/or hypertension in any random member of that population was $\log_2 0.338 = -1.57$ bits. (With a pocket calculator, the sequence $[n], [\ln], [] \div [2], [\ln], [=]$ gives the logarithm of $n$ to the base 2.)

If the blood pressure (0.863, 0.740) were used as the initial screening test and found to be elevated (as defined by the authors), this observation would contribute 0.898 bits of information toward the diagnosis, reducing the uncertainty to $-0.668$. The maximum information obtainable from this observation, if the initial uncertainty were infinite, would be $I_{\text{max}} = 1.73$ bits, but starting with an uncertainty of only $-1.57$ bits, little more than half of the potential information contribution can be realized, illustrating the law of diminishing returns as shown earlier.

If the history, electrocardiogram, and chest x-ray were then done in that order, the history (0.541, 0.580) would contribute 0.125 bits, reducing the uncertainty to $-0.545$ bits. The EKG (0.506, 0.766) would then contribute 0.267 bits, reducing the uncertainty to $-0.278$. The chest film (0.476, 0.921) would contribute 0.228 bits out of a maximum possible 2.59 bits, reducing the uncertainty to $-0.0499$ bits. This sequential reduction in diagnostic uncertainty is shown graphically in figure 1. An uncertainty of only $-0.0499$ bits is very close to absolute certainty.

Coronary Artery Disease

Even the best of laboratory tests may leave the physician far from diagnostic
certainty if the initial uncertainty is great. For example, an elevated serum concentration of the MB isoenzyme of creatine kinase in the first 48 hours after the onset of a myocardial infarct is 100 percent sensitive\(^7\) and about 95 percent specific.

Using the nomogram of Pryor and others,\(^{20}\) a 33-year-old woman who smokes and has atypical chest pain has a 9.5 percent probability of having significant coronary artery disease. This is equivalent to \(-3.40\) bits of uncertainty. An elevated MB isoenzyme (1.0, .95) can, in this circumstance, contribute 2.84 bits of information out of a possible 4.32 bits, reducing the uncertainty to \(-0.563\) bits (figure 2a), leaving the diagnosis still far from certain.

If the same woman not only smokes but has typical chest pain and a history of diabetes mellitus, her probability of having ischemic heart disease is 55 percent, equivalent to \(-0.863\) bits of uncertainty. Following the law of diminishing returns, an elevated MB isoenzyme will now contribute only 0.805 out of a possible 4.32 bits, but this is still enough information to reduce the uncertainty to \(-0.0579\), a nearly certain diagnosis of myocardial infarction (figure 2b).

Even in the latter case, a serum concentration of total creatine kinase of over 120 mIU/L (0.55, 0.65)\(^{21}\) is of little value without the isoenzyme measurement. It would contribute only 0.258 bits of information, leaving the clinician far from certainty (figure 2c).
The Thyroid Nodule

Ashcraft has measured the sensitivities and specificities of several diagnostic tests for thyroid carcinoma.1 The incidence of malignancy among solitary thyroid nodules discovered during a routine physical examination is about 0.18. Taking the logarithm to the base 2 gives an uncertainty of $-2.47$ bits at the time of discovery. The finding of a “cold” nodule after iodine uptake $(0.87, 0.30)$ will provide 0.25 bits of information, decreasing the uncertainty from $-2.47$ to $-2.22$ bits. If, following this, the nodule continues to grow in spite of attempted growth suppression by thyroid hormone $(0.33, 0.96)$ 1.69 bits will be added, still leaving $-0.530$ bits of uncertainty (figure 2d).

In another sequence (figure 2e), an initial ultrasound study showing a solid tumor $(0.85, 0.33)$ will contribute 0.27 bits of diagnostic information, decreasing the uncertainty from $-2.47$ to $-2.20$ bits. A subsequent iodine uptake study showing a cold nodule will further reduce the uncertainty to $-1.96$ bits.

An initial fine needle aspiration diagnosed as malignant $(0.73, 0.99)$ will contribute 2.38 bits, leaving a residual uncertainty of $-0.087$ bits, which is very close to certainty (figure 2f).

The previous procedures can be compared independently of any diagnostic sequence by comparing the maximum contribution of which each is capable, $I_{\text{max}}$. By dividing the dollar cost of a test by its $I_{\text{max}}$, one can compare the cost-effectiveness of tests in terms of dollars per bit of useful information (table I).

Genetic Disease

The frequency of cystic fibrosis among white children is about 1 in 2000 live births. From Eq. 1, the initial uncertainty of the diagnosis is $\log_2 0.0005 = -11$ bits. A definite positive diagnosis will require 11 bits of information.

In making a diagnosis of cystic fibrosis, a positive family history is all-important (a). An elevated sweat chloride, used alone as a screening test, is worthless (b). Similarly, using the data of Guinan,11 an elevated serum level of acid phosphatase, used as a screening test and without a history, is of no value (c and d).

The data of Griner9 show the importance of a positive lung scan in making a diagnosis of pulmonary embolism in a hospitalized patient (e). A negative scan, on the other hand, increases uncertainty and nearly cancels out the information given by a positive physical examination and a decrease in arterial oxygen pressure (f).
If the family history reveals that a sibling has the disease but neither of the parents has it, then the probability of the infant's having the disease is increased to 0.25. Taking the logarithm to the base 2 of 0.25, the uncertainty of child's having cystic fibrosis is now —2 bits. The uncertainty has decreased from —11 to —2 bits, and the family history has provided about 9 bits of information.

If at the age of two months the sweat chloride is over 60 mmol per L (0.95, 0.97), this is additional evidence of the disease. The test provides an additional 1.87 bits of information (out of a possible $I_{\text{max}}$ of 5 bits), decreasing the uncertainty from —2 bits to —0.13 bits, which is very close to absolute certainty (figure 3a).

What if the sweat test were done as a screening test before the family history was known? From Eq. 2, the $\Delta I$ is now 5 bits. The test decreases the uncertainty of the diagnosis from —11 to —6 bits. Although the test is used efficiently, providing nearly the maximum information possible, it still leaves the diagnosis far from certain (figure 3b).

### Prostate Cancer

The following example further illustrates the difficulty in trying to use a chemical screening test without a history.

Guinan and others have measured the sensitivities and specificities of several procedures in the diagnosis of prostatic carcinoma. The prevalence of this neoplasm among all men over 50 years of age is about one percent, equivalent to an uncertainty of —6.64 bits, whereas the prevalence among all men over 50 with a history of symptoms referable to the prostate may be as high as 25 percent (—2.00 bits of uncertainty). A positive clinical history contributes 4.64 bits of information toward a diagnosis of prostatic cancer. If this is followed by a positive rectal examination (0.69, 0.89), the uncertainty is decreased to —0.564 bits. An elevated serum concentration of acid phosphatase (0.56, 0.94) will diminish the uncertainty to only —0.0721, at which point the diagnosis of prostatic cancer is very nearly certain (figure 3c).

If the serum acid phosphatase concentration were used as a screening test for men over 50, an elevation, as defined by the authors, would contribute 3.11 bits of information out of a possible 3.22 bits. Although the test is used efficiently, it will only decrease uncertainty to —3.53, still far from certainty. Even if a positive test is followed by a positive rectal examination, the uncertainty will still be —1.42, an inconclusive result in the absence of a history (figure 3d). The difference between 3c and 3d in the value of a positive rectal examination is another example of the law of diminishing returns as one approaches greater certainty.

### Pulmonary Embolism

Just as a positive test result decreases the uncertainty about a given diagnosis,
an unexpected negative result increases that uncertainty. To calculate the contribution of a negative result to a positive diagnosis, one simply substitutes the complements of the specificity and specificity in Eq. 2. This is illustrated by the following example.

Griner and others have collected from several sources the measured sensitivities and specificities of common symptoms and laboratory tests in the diagnosis of pulmonary embolism. The prevalence of pulmonary embolism among all hospitalized patients over 20 years of age is about 1.31 percent, so that the initial uncertainty of this diagnosis in any random hospital patient over 20 is about $-6.25$. Let us assume that data are gathered in the following sequence. A history of predisposing factors (0.94, 0.60), again, as defined by the authors, will contribute 1.2 bits of information. The presence of pleural pain (0.57, 0.60), cough (0.48, 0.60), and dyspnea (0.80, 0.40) will contribute 0.493, 0.250, and 0.391 bits, respectively. Adding these, the history will contribute 2.34 bits of information, decreasing the uncertainty to from $-6.25$ to $-3.91$. Fever over $38^\circ{\text{C}}$ (0.48, 0.60) and tachypnea (0.65, 0.60) will contribute 0.244 and 0.631 bits, respectively, lowering the uncertainty to $-3.04$ bits. An arterial oxygen tension of less than 90 mm Hg (0.95, 0.40) will add another 0.564 bits, lowering the uncertainty to $-2.48$.

A ventilation-perfusion lung scan showing multiple segmental V/P mismatches (0.64, 0.98) is a highly specific finding. It can provide a maximum of 5 bits of information, and if positive in this example, would add 2.29 bits of information, lowering the uncertainty to $-0.193$ bits (figure 3f).

If, using the previous definition of positivity, the lung scan should turn out to be negative, then the informational contribution to a positive diagnosis of pulmonary embolism is calculated using the complements of sensitivity and specificity (0.36, 0.02). This gives a contribution of $-1.27$ bits, increasing the uncertainty of the diagnosis to from $-2.47$ to $-3.74$ bits (figure 3f).

**Conclusion**

The foregoing examples demonstrate graphically how a clinical diagnosis is developed by adding bits of information derived from a sequence of clinical observations and laboratory results. This quantitative representation of the diagnostic process by information theory offers a relatively easy way for the clinician to visualize, evaluate, and compare diagnostic strategies.

However, this method of analysis is useful only in comparing the relative effectiveness of alternative sequences of diagnostic observations. It has been seen that most strategies approach but do not attain certainty. The next and more difficult decision is just how much residual uncertainty is acceptable. Given the law of diminishing returns, as shown before, when does one stop diagnosing and start treating? This decision involves many more and complex variables and may be weighted by what Feinstein has called the “chagrin factor,” which cannot be expressed numerically.

**References**


