African Journal of Biotechnology Vol. 7 (25), pp. 4807-4818, 29 December, 2008
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB08.050
ISSN 1684–5315 © 2008 Academic Journals

Review

Bayes' theorem: A paradigm research tool in biomedical sciences

Okeh, U. M.[1] and Ugwu, A. C.[2]*

[1]Department of Industrial Mathematics and Applied Statistics, Ebonyi State University, Abakaliki, Nigeria.
[2]Department of Medical Radiography and Radiological Sciences, Nnamdi Azikiwe University, Nnewi Campus, Anambra State, Nigeria.

Accepted 1 December, 2008

One of the most interesting applications of the results of probability theory involves estimating unknown probability and making decisions on the basis of new (sample) information. Biomedical scientists often use the Bayesian decision theory for the purposes of computing diagnostic values such as sensitivity and specificity for a certain diagnostic test and from which positive or negative predictive values are obtained in other to make decisions concerning the well-being of the patient. Often times error rates are encountered and estimated from the results of trials of the screening test with a view to calculating the overall case rate for which an accurate estimate is rarely available. The concept of conditional probability takes into account information about the occurrence of one event to predict the probability of another event. It is on this premise that this article presents Bayes' theorem as a vital tool. A brief intuitive development of this theorem and its application in diagnosis is given with minimum proof and examples.

Key words: Conditional probability, probability, sensitivity, specificity, odd ratio, error rate.

INTRODUCTION

An investigation by Stigler (1983) suggests that Bayes' theorem was discovered by Nicholas Saunderson some time before Bayes. Bayes' theorem is named after the British Mathematician Reverend Thomas Bayes' (1702-1761), who studied how to compute a distribution for the parameter of a binomial distribution (to use modern terminology). In 1763, Bayes' work was edited and presented (in a short paper which has become one of the most memoir in the history of science and one of the most controversial) after his death, as an Essay towards solving a problem in the Doctrine of chances.

Bayes' theorem (also known as Bayes' rule or Bayes' law) is a result in probability theory that relates conditional probabilities. If A and B denote two events, P(A/B) denotes the conditional probability of A occurring, given that B occurs. The two conditional probabilities P(A/B) and P(B/A) are generally different. Bayes' theorem gives a relation between P(A/B) and P(B/A). An important application of Bayes' theorem is that it gives a rule how to update or revise the strengths of evidence-based beliefs in light of new evidence a posterior. As a formal theorem, Bayes' theorem is valid in all interpretation of probability. The use of Bayes' theorem for evaluating laboratory tests and the principles and techniques of medical decision analysis are being introduced in approximately half of the medical schools (Elstein et al., 1985).

Further explanation of Bayes' theorem will be developed with an understanding of its derivation from conditional probabilities and giving some versions of Bayes' theorem. The treatment will be denoted exclusively to an understanding of what the theorem is, what it means and what it can be used for. We shall review some principles and terminology of diagnostic test evaluation and go further to give diagnostic application of Bayes' theorem in biomedical research using examples. Specifically, it will be shown that positive and negative predictive values can be computed from the sensitivity and specificity of diagnostic tests, an important characteristic of such tests. Bayes' theorem will be seen in terms of odd ratio (a measure of association for 2 x 2 tables that are not functions of χ^2 especially in cross-

*Corresponding author. E-mail: tonybullng@yahoo.ca. Tel: +2348076241297.

sectional studies. Most importantly, expected error rates (false positive and false negative rates) will be investigated if a test is to be used in a screening program. The importance of the characteristics of diagnostic tests was indicated in a recent survey (Davson-Saunder et al.,1987) were 89% of the medical schools surveyed include this topic in their curriculum. Recent result on the evaluation of genetic testing with particular reference to the estimations of sensitivity, specificity, PPV and NPV will be made known.

PROBABLITY

For better understanding of Bayes' theorem, it is important to introduce some notation from probability theory. Suppose we wish to study a certain population of individuals and suppose that each of these individuals is either diseased (D) or disease-free (D^c). Because every individual in the population is in one of these two categories, these two categories (or events) are said to be jointly exhaustive. Because, an individual cannot be in both categories, the two categories are said to be mutually excusive. So the two conditions, D and D^c form mutually exclusive and jointly exhaustive events in our population. Now, suppose that we wish to predict the disease status of an individual to be randomly selected from this population, i.e. what is the likelihood that a diseased individual will be selected or that a disease-free individual will be selected? We will use a numerical measure of likelihood on a scale from zero to one called probability. The probability that an individual in the population is diseased, denoted by P(D), is the overall likelihood that the individual is diseased, sometimes called the prevalence rate of the disease; it is computed by dividing the number of diseased individuals in the population by the total of individual in the population:

$$P(D) = \frac{number\ \ of\ \ diseased\ \ people}{population\ \ \ size}$$

P(D) is in other words taken to be the prior probability or marginal probability of D. It is "prior" in the sense that it does not take into account any information about D^C.
Also for $P(D^c)$, the probability that an individual from the population is diseased-free is as follows:

$$P(D^c) = \frac{number\ of\ \ disease-free\ people}{population\ size}$$

So $P(D^c)$ is the prior or marginal probability of B, and acts as a normalizing constant. Note that

$$P(D) + P(D^c) = 1 \qquad (1)$$

Because the sum of these two fractions yield the number of diseased people plus the number of disease-free people divided by the total number of people, which is one (remember that the categories, diseased and disease- free are jointly exhaustive). If an event has a probability of 0, then that event is considered to be nonexistent. So, P(D)= 0 implies that no one in the population has the disease. If an event has a probability of 1, then that event is considered to be certain to occur. Hence, P(D)=1 implies that everyone in the population has the disease. If P(D) =0.5, then a randomly selected individual from the population is just as likely to be diseased as to be disease-free.

Suppose now that a certain number of individuals in this population are male (M). From the discussion above, we can compute the probability that a randomly selected individual from the population is male, P(M), by dividing the number of males by the total number of individuals in the population:

$$P(M) = \frac{number\ of\ \ male}{population\ size}$$

Consider the probability that a male in this population is diseased. Technically, this is referred to as the conditional probability that an individual in this population is diseased given that the individual is male and is denoted by P(D/M). Therefore, this conditional probability is the number of individuals in the population who are diseased and male divided by the total number of males in the population (relative frequency of male who are diseased).

Mathematically,

$$P(D/M) = \frac{P(D)\ and\ P(M)}{P(M)} \quad provided\ P(M) \neq 0$$

or

$$P(D/M) = \frac{P(M/D)\ and\ P(D)}{P(M)}$$

where P(D/M) is the conditional probability of D, given M. It is also called the posterior probability because it is derived from or depends upon the specified value of M. Where

$P(D)$ and $P(M) = P(D) \times P(M)$

Also $P(D\ and\ M) = P(D/M) \times P(M)$

Similarly, the proportion of females (F) who are diseased is as follows:

$$P(D/F) = \frac{number\ of\ \ diseased\ \ female}{number\ of\ \ female}, provided\ P(F) \neq 0$$

Table 1. A 2×2 table of disease status versus sex.

Sex	Disease status		
	Diseased	**Not Diseased**	**Total**
Male	e	f	e+f
Female	g	h	g+h
Total	e+g	f +h	e +f+g +h

Using the notation of Table 1, Where e = number of diseased males; f = number of disease-free males; g = number of diseased females; h = number of disease-free females.

Table 2. A 2×2 table of disease status versus sex.

Sex	Disease status		
	Diseased	**Not Diseased**	**Total**
Male	200	500	700
Female	400	300	700
Total	600	800	1400

Also $P(D/F) = \dfrac{P(F/D)P(D)}{P(F)}$

It is called conditional probability of D given F.

$$P(D/F) = \frac{P(D)\ and\ P(F)}{P(F)},\ provided\ P(F) \neq 0$$

Where $P(D)\ \ and\ P(F) = P(D/F) \times P(F)$; and means intersection of D and F.

Note that in the case of conditional probability, the population can be thought of as being reduced so that it includes only those individuals with the given condition, ie. in computing conditional probabilities attention is restricted to the subpopulation of individuals assuming the given condition. Based on the above explanations, we will conclude that Bayes' theorem relates the conditional and marginal probabilities of stochastic events A and B.

These probabilities and conditional probabilities can easily be computed from a table that shows the population cross- classified according to sex and disease status; such a table is called a contingency table (Table 1).

$$P(M) = \frac{number\ of\ males}{population\ size}$$

= (e+f)/(e+f+g+h)

$$P(D) = \frac{(number\ of\ diseased\ individuals)}{population\ size}$$

=(e+g)/(e+f+g+h)

If we are interested in the disease risk only among males in the population, then we would ignore the g+h females and calculate as follows:

$$P(D/M) = \frac{(number\ of\ diseased\ males)}{(number\ of\ males)}$$

=e/(e+f)

To find proportion of diseased individuals who are females we would ignore the f+h disease-free individuals:

$$P(F/D) = \frac{(number\ of\ diseased\ females)}{number\ of\ diseased}$$

Furthermore, $P(D^c)$=(f+h)/(e+f+g+h)

$P(M/D^c)$=f/f+h and $P(F/D^c)$=h/f+h

Example 1: Consider the population in the example below (Table 2):

From the above table, the prevalence rate is as follows:

P(D)=600/1400=.428

The proportion of diseased individuals who are male is as follows:

P(M/D)=200/600=.3

The proportion of disease-free individuals who are male is as follows:

$P(M/D^c)$=500/800=0.625.

Consequently, one-third of diseased individuals are male; 63% of the disease-free individuals are male. Finally, the proportion of males who are diseased is as follows:

P(D/M)=200/700=0.286.

From the discussion above, it was shown that if a contingency table showing the entire population cross-classified according to sex and disease status is provided, then all conditional probabilities can be computed directly. However, it is not always the case that, that contingency table is available; sometimes only certain probabilities are available. For instances, suppose that the proportion of diseased individuals who are male, P(M/D), the proportion of diseased individuals who are female, P(F/D), and the prevalence rate, P(D) are known. Note that if P(D) is known, then $P(D^c)$ will also be known since $P(D^c)$= 1-P(D) by equation 1.

DERIVING BAYES' RESULT FROM CONDITIONAL PROBABILITIES

Now suppose we wish to know the proportion of males who are diseased, P(D/M). Bayes' theorem solves the problem of finding this probability if we recall the fact that D is an event in form of A_1 and M as another event B, such that

$$P(A_1 / B) = \frac{P(A_1 \ and \ B)}{P(B)}, provided \ P(B) \neq 0$$

and similarly the probability of another events A_2 given B is

$$P(A_2 / B) = \frac{P(A_2 \ and \ B)}{P(B)}, provided \ P(B) \neq 0$$

Where $P(B) = P(A_1 \ and \ B) + P(A_2 \ and \ B)$

$$P(A_1 and \ B) = P(A_1) \times P(B / A_1) \ and \ P(A_2 \ and \ B) = P(A_2) \times P(B / A_2)$$

This lemma is sometimes called the product rule for probabilities. If both sides of this equation is divided by P(B), provided that P(B)≠0 then we obtain Bayes' theorem as

$$P(A / B) = \frac{P(B / A)P(A)}{P(B)} .$$

In terms of P(D/M), this result will be written for only two events as

$$P(D / M) = \frac{P(D)P(M / D)}{P(D)P(M / D) + P(D^c)P(M / D^c)} \quad (2)$$

Bayes' rule is thus a way of computing a certain conditional probabilities, P(D/M), exclusive from certain probabilities [P(D) and $P(D^c)$] and certain conditional probabilities, [P(M/D) and $P(M/D^c)$] that are known. If the prevalence rate [P(D)] and the proportions of the diseased and disease-free subpopulations that are male [P(M/D) and $P(M/D^c)$] are known, then the proportion of males that are diseased [P(M/D)] can be computed using equation 2.

In the same manner, the conditional probability of D^c given M is

$$P(D^c / M) = \frac{P(D^c)P(M / D^c)}{P(D)P(M / D) + P(D^c)P(M / D^c)} \quad (3)$$

In terms of equation 2 it may be helpful to note that the term in the numerator, P(D)P(M/D), also appears in the denominator as part of a sum of two terms, the other term in that sum having exactly the same form but with D^c replacing D.

Example 2: By way of a numerical example, suppose we have

P(D)= 0.4,P(M/D)=0.25,$P(M/D^c)$=0.667.

Then we can compute the proportion of males who are diseased, P(D/M), by using the above probabilities and equation 2,

$$P(D / M) = \frac{(0.40)(0.25)}{(0.4)(0.25) + (0.6)(0.667)} = 0.20$$

Note that this result matches the value that we obtained in example 1 since the numbers were taken from that example.

Bayes' rule can be extended to the case of a set of mutually exclusive and jointly exhaustive events that exceed 2. For instance, the population can be classified into individuals who are married, divorced, or single. The formula for three mutually exclusive and jointly exhaustive events e.g, A,B, and C, and arbitrary event E is as follow:

$$P(A / E) = \frac{P(A)P(E / A)}{P(A)P(E / A) + P(B)P(E / B)P(C)P(E / C)} \quad (4)$$

For computing P(B/E) and P(C/E) using Bayes' rule denominator would be exactly the same as above, the numerator would be P(B) P(E/B) and P(C) P(E/C), respectively. A proof and more formal development of Bayes' rule can be found in most probability texts (Rosner, 1990).

Table 3. Association between birth weight and maternal age: cross sectional study.

Maternal age	Birth weight		
	B	$\overline{B}$	Total
A	10	40	50
$\overline{A}$	15	135	150
Total	25	175	200

Source: Fleiss, 1987.

ALTERNATIVE FORM OF BAYES'THEOREM

Bayes' theorem is often embellished by noting that

$$P(B)=P(A\cap B)+P(A^c\cap B)=P(B/A)P(A)+P(B/A^c)P(A^c)$$

where A^c is the complement-tary event of A (often called "not A"). So the theorem can be restated as

$$P(A/B)=\frac{P(B/A)P(A)}{P(B/A)P(A)+P(B/A^c)P(A^c)} \quad (5)$$

More generally, where $\{A_i\}$ forms a partition of the e vent space,

$$P(A_i/B)=\frac{P(B/A_i)P(A_i)}{\sum_j P(B/A_j)P(A_j)}$$

for any A_i in the partition.

BAYES' THEOREM IN TERMS OF ODDS RATIO

Some scientist presented a great many measures of association for 2 x 2 tables that are not functions of χ^2 (Goodman and Kruskal, 1954, 1959). Here Bayes' theorem will be used in terms of odd ratio.

Frequently, one of the two characteristics being studied is antecedent to birth weight, a measure of the risk of experiencing the outcome under study when the antecedent factor is present is

$$\Omega_A=\frac{P(B/A)}{P(\overline{B}/A)} \quad (6)$$

Where Ω_A =odds that B will occur when A is present. Since P(B/A) may be estimated by

$$P(B/A)=\frac{P_{11}}{P_{1.}} \text{ and } P(\overline{B}/A)=\frac{P_{12}}{P_{1.}}.$$

Therefore Ω_A may be estimated by

$$o_A=\frac{P_{11}/P_{1.}}{P_{12}/P_{1.}}=\frac{P_{11}}{P_{12}}. \quad (7)$$

Example 3: Suppose that we are studying the association, if any between the age of the mother (A represents a maternal age less than or equal to 20 years). $\overline{A}$, a maternal age over 20 years) and the birth weight of her offspring (B represents a birth weight less than or equal to 2500 g; $\overline{B}$, a birth weight over 2500 g). Let us suppose that the sample of 200 records has been selected and the data are shown below.

From the example above, the estimated odds that a mother aged 20 years or less will deliver an offspring weighing 2500 g or less are, from Table 4

$$O_A=\frac{0.05}{0.20}=\frac{1}{4}=0.25 \quad (8)$$

Thus for every four births weighing over 2500 g to mothers aged 20 years or less, there is one-birth weighing 2500 g or less. The information conveyed by these odds is exactly the same as that conveyed by the rate of how birth weight specific to young mothers,

$$P(B/A)=\frac{P_{11}}{P_{1.}}=\frac{0.05}{0.25}=0.2$$

But the emphasis differs. One can imagine attempting to educate prospective mothers aged 20 years or less. The impact of the statement, "One out of every five of you is expected to deliver an infant with a low birth weight," may well be different from the impact of "For every four of you who deliver infants of fairly high weight, one is expected to deliver an infant of low birth weight." When A is absent, the odds of B's occurrence are defined as

$$\Omega_{\overline{A}}=\frac{P(B/\overline{A})}{P(\overline{B}/\overline{A})}. \quad (9)$$

which may be estimated as

Table 4. Joint proportions derived from table 3 above.

Maternal Age	Birth weight		
	B	$\overline{B}$	Total
A	0.05(=P_{11})	0.20(=P_{12})	.25(=$P_{1.}$)
$\overline{A}$	0.075(=P_{21})	0.675(=P_{22})	.75(=$P_{2.}$)
Total	.125(=$P_{.1}$)	0.875(=$P_{.2}$)	1.

Source: Fleiss, 1987.

Table 5. A 2×2 table of disease status versus test result.

Test result	Disease status		
	Diseased	Not Diseased	Total
Test +	e	f	e +f
Result-	g	h	g +h
Total	e +g	f + h	e+f+g+h

$$O_{\bar{A}} = \frac{P_{21} / P_{2.}}{P_{22} / P_{2.}} = \frac{P_{21}}{P_{22}}.$$

For our example, the estimated odds that a mother aged more than 20 years will deliver an offspring weighing 2500 g or less are

$$O_{\bar{A}} = \frac{0.075}{0.675} = 0.11$$

Thus for every nine births weighing over 2500 g to mothers aged more than 20 years (as opposed to every four to younger mothers), there is one birth weighing 2500 g or less.

INDICES OF DIAGNOSTIC TEST PERFORMANCE

The basic notation and principles of diagnostic test evaluation have been delineated (Khamis, 1987). A brief review will be given in the section.

Consider a new test designed to discriminate between those with a particular disease (or condition) and those who are disease-free. Suppose a contingency table can be formed showing the cross-classification of all individuals in a population according to their disease status (sometimes called the gold standard) and their test result (Table 5).

For notational purposes and as defined previously, let D represent the event that an individual is diseased, D^c the event that an individual is disease-free, + the event that an individual tests positive, and – the event that an individual tests negative. Recall also from Table 3 that e represents the number of individuals who are diseased and test positive, f represents the number of individuals who are disease-free and test positive, and so on.

Suppose we take Sonographic Science as a case, the growth of ultrasound can, in large part, be attributed to new clinical applications. Researchers must demonstrate that proposed applications selectively identify patients with specific disease. Statistical parameters which are measures of the diagnostic value of the test (typically sensitivity, specificity, and accuracy) have been developed to judge the efficacy of diagnostic tests. The patient may or may not have disease and the ultrasound examination may or may not have positive findings. Four outcomes are possible:

1. True-positive (TP): the ultrasound findings are positive and the patient has disease.
2. False -positive (FP): the ultrasound findings are positive and the patient does not have disease.
3. True-negative (TN): the ultrasound findings are negative and the patient does not have disease.
4. False-negative (FN): the ultrasound findings are negative and the patient has disease.

The determination of disease is accomplished independently by using an established procedure (a surgical biopsy). The perfect diagnostic test would identify all diseased persons with positive findings and all non diseased persons with negative findings .The sensitivity of the diagnostic test is the percentage of all subjects with disease that yield a positive test result. Mathematically,

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \quad (10)$$

Sensitivity describes how well the diagnostic test identifies subjects with disease. Minimizing false-negative

improves the reliability of the diagnostic test. From the preceding sections, this is seen to simply the conditional probability, P(+/D)=e/(e+g).

This is a measure of how sensitive the test is in detecting the disease. The test is in detecting the disease. The specificity of the test is the proportion of those not having the disease who test negative or how well the diagnostic test excludes non diseased subjects from having a positive test result. Minimizing false-positives also improves the reliability of the diagnostic test.

Mathematically,

$$Specificity = \frac{TN}{TN + FP} \times 100 \quad (11)$$

From the above section, it is seen simply as the conditional probability $P(-/D^c)=h/(f+h)$.

This is a measure of how specific the test is in detecting absence of the disease. Accuracy of the diagnostic test is the percentage of all subjects tested who are correctly assessed as having or not having disease. The equation for accuracy is,

$$Accuracy = \frac{TP + TN}{All\ subjects\ tested} \times 100 \quad (12)$$

Following the explanation in Table 3, Accuracy=(e+h)/(e+f+g+h) while two measures, sensitivity and specificity, are just conditional probabilities, the two measures of particular interest to biomedical scientist and clinicians for the purpose of predicting individuals disease status given that their test results known are the positive predictive value (PPV) and the negative predictive value(NPV) of the test. The PPV indicates the likelihood of disease if the test is positive or the proportion of those with a positive test result who are diseased i.e. the conditional probability P(D/+)=e/(e+f).The NPV is defined as the proportion of those with a negative test result who are disease-free,i.e. the conditional probability $P(D^c/-)=h/(g+h)$.

The defining equations are

$$PPV = \frac{TP}{TP + FP} \times 100 \quad (13)$$

While $$NPV = \frac{TN}{TN + FN} \times 100 \quad (14)$$

Sensitivity, specificity, accuracy, and positive and negative predictive values are also expressed by fractions between 0 and 1 to indicate probability of various outcomes. In this format the factor 100 is eliminated in equations 11 through 14. Applications of these terms are given in the following example.

Example 4: In Hedrick et al. (1995) carried out a study to obtain NPV, PPV, accuracy, sensitivity and specificity. They carried out an ultrasound examination with the following outcomes when compared with the diagnostic standard.

True-positive (TP) =53
False-positive (FP) =18
True-negative (TN) =70
False-negative (FN) =4

Total number of subjects tested = 145. The results indicates that

$$Sensitivity = \frac{53}{57} \times 100 = 93\%$$

$$Specificity = \frac{70}{88} \times 100 = 80\%$$

$$Accuracy = \frac{53 + 70}{145} \times 100 = 85\%$$

$$PPV = \frac{53}{71} \times 100 = 85\%$$

$$NPV = \frac{70}{74} \times 100 = 95\%$$

Example 5: Townsend et al. (1988) reported on a retrospective review of initial sonograms performed on 65 twin gestations to evaluate the ability of sonography to distinguish monochorionic from dichorionic gestations based on the thickness of the membrane separating the fetuses. The results are shown as a contingency table (Table 6). In this table, the indicators of diagnostic accuracy are as follows:

Sensitivity = P(+/D) = 39/42 = 93%, where P(+/D) means the probability of testing positive and having disease and 93% here indicates a very high sensitivity.

Specificity $P(-/D^c)$ = 15/23 = 65%. Also $P(-/D^c)$ means testing negative and having no disease.

The predictive values are as follows:

PPV = P(D/+) = 39/47 = 83%. P(D/+) means having disease and testing positive.

NPV = $P(D^c/-)$ = 15/18 = 83.3%. P(Dc/-) means having no disease and testing negative.

Hence, a thick membrane (+) detected dichorionic gestation (D) with a sensitivity of 93%, and a thin membrane (-) detected monochorionic gestation D^c with a specificity of 65%. The PPV of a thick membrane in predicting dichorionicity of 83%, and the NPV of a thin

Table 6. A 2×2 table for example 3.

Ultrasound Exam. Result	Dichorionicity(D). A condition	Monochorionicity (D^1).A condition	Total
Thick membrane (+)	39	8	47
Thick membrane (-)	3	15	18
Total	42	23	65

Source: Townsend et al. (1988).

membrane in predicting monochorionicity is 83.3%. In other words, the ultrasound examination result (whether it is + or -) is about 83% accurate for prediction purposes.

AN APPLICATION OF BAYES' THEOREM TO DIAGNOSTIC TEST EVALUATION

It is often the case that clinicians can measure how often specific symptoms occur in diseased and disease-free people (for definitions of the sensitivity and specificity, see Yerushalmy,1947), but cannot directly measure the predictive value of a set of symptoms (i.e, the PPV and NPV). Therefore, it would be beneficial if we could find a way to use the sensitivity and specificity, which are quantities that a biomedical researcher can estimate, to compute predictive values, which are quantities that he/she needs to make appropriate diagnoses or decisions. In other words, given the prevalence rate in the population, P(D), the sensitivity, P(+/D), and the specificity, P(-/D^c), how can we compute the PPV, P(D/+), and the, P(D^c/-)?. The answer is Bayes' theorem. A frequent application of Bayes' theorem is in evaluating the performance of a diagnostic test intended for use in a screening program (Fleiss, 1987).

Note that D and D^c are mutually exclusive and jointly exhaustive events. So, by Bayes' theorem (equation 2 , using different symbols,

$$PPV = P(D/+) = \frac{P(D)P(+/D)}{P(D)P(+/D) + P(D^c)P(+/D^c)}$$

$$NPV = P(D^c/-) = \frac{P(D^c)P(-/D^c)}{P(D^c)P(-/D^c) + P(D)P(-/D)}$$

By using an extension of the property in equation 1, it can be shown that P(+/D^c) = 1-P(-/D^c) = 1-specificity and P(-/D) = 1-P(+/D) = 1- sensitivity. Now the preceding formulas for PPV and NPV can be written in a more convenient form by substituting these expressions for P(+/D) = sensitivity, P(-/D^c) = specificity, and P(D^c) = 1-P(D).

$$PPV = \frac{P(D)\times(sensitivity)}{P(D)\times(sensitivity) + (1-P(D))\times(specificity)} \quad (15)$$

$$NPV = \frac{(1-P(D))\times(sensitivity)}{P(D)\times(1-sensitivity) + (1-P(D))\times sensitivity} \quad (16)$$

Equations 15 and 16 express the PPV and NPV of a test exclusively in terms of the prevalence rate and the diagnostic values of the test.

Example 6: As a means of checking these formulas, we can use the data from example 5: P(D) = 42/65 = 0.65, Sensitivity = P(+/D) = 0.93 and Specificity = P(-/D^c) = 0.65

Using equations 10 and 11, we can obtain the PPV and NPV as follows:

$$PPV = \frac{0.65\times0.93}{(0.65)\times(0.93) + (0.35\times0.35)} = 0.83$$

$$NPV = \frac{0.35\times0.65}{(0.65\times0.07) + (0.35\times0.65)} = 0.833$$

These values for PPV and NPV obtained from Bayes' theorem match those from example 3 where they were obtained directly from the contingency table.

Example 7: In a study (Khamis,1990),181 pregnant women were referred for a level II ultrasound examination at or around a gestation of 18 weeks with particular attention devoted to head size, head shape, and ventriculo-hemispheric ratio to diagnostic hydrocephalus and assessment of the cranial vault. Of the 181 pregnancies, 12 fetuses were found to have spinal defects. The goal of this study was to examine the effectiveness of using ultrasound detection of a bullet-shaped head (flattening of the parietal regions with pointing of the frontal bones) to predict spinal defects. According to the article, the sensitivity of the bullet-shaped was 75% and the specificity was 98.3%. Let correspond to the event that a fetus has spinal defects, + correspond to detection of a bullet–shaped head by ultrasound examination, and – correspond to lack of detection of a bullet–shaped head. The above information provides the following values: P(D) = 12/181=0.066, Sensitivity = P(+/D) = 0.75, Specificity = P(-/D^c) = 0.983. If a clinician uses ultrasound examination to detect a bullet-shaped head in the fetus under the conditions described in the article, can the above values be used to estimate the likelihood of spinal defects given a positive

(or negative) test result?. Yes, if the test is positive, the probability of spinal defects is as follows (see also equation 15):

$$PPV = P(D/+) = \frac{0.66 \times 0.75}{(0.066 \times 0.75) + (0.934 \times 0.017)} = 0.757$$

If the test is negative, the probability of no spinal defects is as follows (see also equation 16):

$$NPV = P(D^c/-) = \frac{0.934 \times 0.983}{(0.066 \times 0.25) + (0.934 \times 0.983)} = 0.982$$

A negative test result indicates no spinal defects with high accuracy (98.2%), but the estimated probability of spinal defects in the presence of a positive test result is only 75.7%.

DRUG TESTING

Example 8: In using Bayes' theorem to evaluate the result of drug tests, suppose a certain drug test is 99% sensitive and 99% specific, that is, the test will correctly identify a drug user as testing positive 99% of the time, and will correctly identify a non-user as testing negative 99% of the time. This would seem to be a relatively accurate test, but Bayes' theorem will reveal a potential flaw. Let us assume a corporation decides to test its employees for opium use, and 0.5% of the employees use the drug. We want to know the probability that, given a positive drug test, an employee is actually a drug user. Let "D" be the event of being a drug user and D^c indicate being a non-user. Let "+" be the event of a positive drug test. We need to know the following:

(a) P(D), or the probability that the employee is a drug user, regardless of any other information. This is 0.005, since 0.5% of the employees are drug users. This is the priori probability of D.
(b) $P(D^c)$, or the probability that the employee is not a drug user. This is 1-P(D), or 0.995.
(c) P(+/D), or the probability that the test is positive, given that the employee is a drug user. This is 0.99, since the test is 99% accurate.
(d) $P(+/D^c)$, or the probability that the test is positive, given that the employee is not a drug user. This is 0.01, since the test will produce a false positive for 1% of non-users.
(e) P(+), or the probability of a positive test event, regardless of other information. This is 0.0149 or 1.49%, which is found by adding the probability that the test will produce a true positive result in the event of drug use (= 99% × 0.5% = 0.495%) plus the probability that the test will produce a false positive in the event of non-drug use (= 1% × 99.5% = 0.995%).This is the prior probability of +.

Given this information, we can compute the posterior probability P(D/+) of an employee who tested positive actually being a drug user.

$$P(D/+) = \frac{P(+/D)P(D)}{P(+)} = \frac{P(+/D)P(D)}{P(+/D)P(D) + P(+/D^c)P(D^C)} \quad \ldots.17$$

$$= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} = 0.3322$$

Despite the high accuracy of the test, the probability that an employee who tested positive actually did use drugs is only about 33%, so it is actually more likely that the employee is not a drug user. The rarer the condition for which we are testing, the greater the percentage of positive tests that will be false positives. Details on how to calculate error rates (false positive and false negative) will be discussed under the next subheading.

ERROR RATES

Of greater concern than the test's sensitivity and specificity, however, are the error rates to be expected if the test is actually used in a screening program. If a positive result is taken to indicate the presence, then the false positive rate, say P_{F+}, is the proportion of people among those responding positive, who are actually free of the disease, or $P(D^c/+)$ according to Bayes' theorem

$$P_{F+} = P(D^c/+) = \frac{P(+/D^c)P(D^c)}{P(+)} \quad (18)$$

It can also be written in the form,

$$P_{F+} = \frac{P(+/D^c)(1-P(D))}{P(+)}$$

Since $P(D^c)=1-P(D)$

The false negative rate, say P_{F-}, is the proportion of people, among those responding negative on the test, who nevertheless have the disease, or P(D/—). Again by Bayes' theorem,

$$P_{F-} = P(D/-) = \frac{P(-/D(P(D)}{P(-)} = \frac{P(-/D)P(D)}{P(-)}$$

$$P_{F-} = \frac{1-P(+/D)P(D)}{1-P(+)} \quad \ldots\ldots.19$$

Since P(—/D) = 1—P(+/D) and P(-) = 1—P(+). We still need the overall rates P(+) and P(D) in order to evaluate these two error rates. Actually, we only need P(D), for the following reason. Note that,

Table 7. Results of a trial of a screening test.

Screening	Test Result		
Disease status	+	--	Total
Present (D)	950	50	1000
Absent (D^c)	10	990	1000

Source (8).

$$P(+)=\frac{N_+}{N}=\frac{N_{+D}+N_{+D^c}}{N}=\frac{N_{+D}}{N}+\frac{N_{+D^c}}{N} \quad (20)$$

In equ.20, N_{+D} denotes the number of people who have the disease and respond positive and N_{D^c} denotes the number of people who are free of the disease and respond positive. Multiplying and dividing the first of the two terms on the right-hand side of equ.20 by N_D, the number of people with the disease, we find that

$$\frac{N_{+D}}{N}=\frac{N_{+D}}{N_D}\frac{N_D}{N}=P(+/D)P(D).. \quad (21)$$

Similarly, by multiplying and dividing the second term by $N_D{}^c$, the number of people without the disease, we find that

$$\frac{N_{+D^c}}{N}=\frac{N_{+D^c}}{N_{D^c}}\frac{N_{D^c}}{N}=P(+/D^c)P(D^c).. \quad (22)$$

Substituting the expressions from eqns 21 and 22 in eqn 20, we find that

$$P(+)=P(+/D)P(D)+P(+/D^c)P(D^c). \quad (23)$$

This equation is a special case of the familiar result that an overall rate P(+), is a weighted average of specific rates P(+/D) and P(+/D^c) with the weights being the proportions of people in the specific categories P(D) and P(D^c). Since $P(D^c)=1-P(D)$, eqn 23 becomes

$$\begin{aligned}P(+)&=P(+/D)P(D)+P(+/D^c)1-P(D))\\&=P(+/D^c)+P(D)P(+/D)-P(+/D^c).....\end{aligned} \quad (24)$$

Substituting of eqns 24 and 18 yields, as the expression for the false positive rate.

$$P_{F+}=\frac{P(+/D^c)(1-P(D))}{P(+/D^c)+P(D)(P(+/D)-P(+/D^c))} \quad (25)$$

Substitution of eqn 24 in eqn 19 yields, as the expression for the false negative rate,

$$P_{F-}=\frac{(1-P(+/D)P(D)}{1-P(+/D^c)-P(D)(P(+/D)-P(+/D^c)} \quad (26)$$

Analysis of the false positive and false negative rates associated with screening tests have been performed in psychology (Meehl and Rosen, 1958; Daves,1962), in Opthalmology (Vastola and Kokubu, 1962), and for a variety of Medical disorders (Cochrane and Holland, 1971). We see from eqns 25 and 26 that, in general, the two error rates are functions of the proportions P(+/D) and P(+/D^c), which may be estimated from the results of a trial of the screening test, and of the overall case rate P(D) , for which an accurate estimate is rarely available. Nevertheless, a range of likely values for the error rates may be determined as in the following example.

Example 9: Suppose that the test is applied to a sample of 1000 people known to have the disease and to a sample of 1000 people known not to have the disease. Suppose that this trial resulted in the frequencies shown in the Table 7, we would then have the estimates P(+/D) = 950/1000 =.95 and P(+/D^c) = 10/1000 = 0.01
a pair of probabilities indicating a test that is sensitive [P(+/D) is close to unity]and specific [P(+/D^c) is close to zero] to the disease being studied. Substitution of these two probabilities in eqn 22 gives, as the values for the false positive rate

$$P_{F+}=\frac{0.01(1-P(D))}{0.01+P(D)(.95-0.01)}=\frac{0.01(1-P(D))}{0.01+.94P(D)} \quad (27)$$

the final expression resulting by multiplying both the numerator and the denominator of the preceding expression by 100. Substitution in eqn 23 gives, as the value for the false negative rate,

$$P_{F-}=\frac{(1-.95)P(D)}{1-0.01-P(D)(.95-0.01)}=\frac{0.05P(D)}{.99-.94P(D)}=\frac{5P(D)}{99-94P(D)} \quad (28)$$

Table 8 gives the error rates associated with various values of P(D), the overall case rate. Rarely will the case rate exceed 1% of the population.

If the disease is not too prevalent- if it affects, say, less than 1% of the population, the false negative rate will be

Table 8. Error rates associated with screening test.

P(D)	False positive (P_{F+})	False negative (P_{F-})
1/million	.9999	0
1/100,000	.9991	0
1/10,000	.9906	.00001
1/1000	.913	.00005
1/500	.840	.00010
1/200	.677	.00025
1/100	.510	.00051

Source: Fleiss, 1987.

quite small, but the false positive rate will be rather large. From one point of view, the test is a successful one: Since P_{F-} is less than 5/10,000, therefore, of every 10,000 people who respond negative and are thus presuming given a clean bill of health, no more than five should actually have been informed they were ill. From another point of view, the test is a failure: Since P_{F+} is greater than 50/100, therefore, of every 100 people who respond positive and thus presumably are told they have the disease or are at least advised to undergo further tests, more than 50 will actually be free of the disease. The final decision whether or not the test will depend on the seriousness of the disease and on the cost of further tests or treatment. Because the false positive rate is so great, however, it would be hard to justify using this screening test for any but the most serious diseases.

On method of reducing the false positive or negative rate associated with a diagnostic screening procedure (but thereby increasing its cost) is to repeat the test a number of times, and to declare the final result positive if the subject responds positive to each administration of the test or if he responds positive to a majority of the administrations. For some disorders, a better rule is to administer the test three times and to declare the final result positive if the subject responds positive to at least two of the three administrations. Only those subjects who respond positive to one of the first two administrations and negative to the other will have to be tested a third time. Those who respond positive to both of the first two administrations would be declared positive, and those who respond negative to both would be declared negative (Sandifer et al., 1968).

A more accurate but more complex assessment of the performance of a screening procedure than the above is possible when disease severity is assumed to vary and not as here merely to be present or absent. The appropriate analysis was originally proposed (Neyman, 1968) and later extended (Greenhouse and Mantel, 1950), (Nissey-Meyer, 1964).

CONCLUSION

From the results, we conclude that a positive test result increases the disease risk relative to the prevalence rate often times. While negative test result reduces the risk of the disease relative to its prevalence rate majority of the time also. We conclude also that reducing error rates requires that one should repeat the screening test a number of times and to declare the final result positive if the subject responds positive to each or majority of the administration(s) of the test. More accurate but complex assessment of the performance of a screening procedure is to assume that disease severity varies and not to say that it is present or absent. This article has presented a brief intuitive development of probability and the statement about Bayes' theorem with emphasis on obtaining diagnostic values from which positive or negative predictive values are achieved with the view to making decisions about the well-being of the patient. This is why this theorem is an indispensable research tool in biomedical sciences.

REFERENCES

Bayes T (1763). "An Essay towards solving a problem in the Doctrine of Chances. By the late Rev. Mr. Bayes FRS. Communicated by Mr. price, in a letter to John Canton AMFRS" ;Philosophical Transactions, Giving Some Account of the present Undertaking, studies and Labours of the Ingenious in Many Considerable parts of the World 53: 370-418.

Cochrane AL, Holland WW (1971). Validation of Screening procedures. Br. Med. Bull. 27: 3-8.

Davson-Saunder B, Azen S, Greenberg RS, Reed AH (1987). The instruction of biostatistics in Medical schools. Am. Stat. 41: 263-266.

Dawes RM (1962). A note on base rates and Psychometric efficiency. J. Consult. Psychol. 26: 422-424.

Elstein AA, Davson-Saunder B, Belzer LJ (1985). Instruction in Medical decision Making. Med. Decision Making. S: 229-233.

Fleiss JL (1973). Staistical Methods for Rates and proportions. John Wiley and Sons,Inc. 3: 39-45.

Goodman LA, Kruskal WH (1954). Measures of association for cross classifications. J. Am. Stat. Assoc. 49: 732-764.

Goodman LA, Kruskal WH (1959). Measures of association for cross classifications. II:Further discussion and references. J. Am. Stat. Assoc. 54: 123-163.

Greenhouse SW, Mantel N (1950). The evaluation of diagnostic tests. Biometrics, 6: 399-412.

Hedrick WR, Hykes DL, Starchman DE (1995). Ultrasound physics and Instruction 3rd ed Mosby-year book, Inc: USA. 343-344.

Khamis HJ (1987). Statistics refresher: test of hypothesis and diagnostic test evaluation JDMS. 3: 123-129.

Khamis HJ (1990). An application of Bayes' Rule to diagnostic Test

Evaluation, JDMS, Citing Chambers et al. 4: 212-218.
Meehl PE, Rosen A (1958). Antecedent probability and the efficiency of Psychometric Signs, patterns, or cutting scores. Psychol. Bull. 52: 194-216.
Neyman J (1947). Outline of Statistical treatment of the problem of diagnosis. Pub. Health Rep. 62: 1449-1456.
Nissey-Meyer S (1964). Evaluation of Screening tests in medical diagnosis. Biometrics. 20: 730-755.
Pierre-Simon L (1774). Memoir on the probability of the causes of events, Stat. Sci. 1(3): 364-378.
Rosner B (1990). Fundamental of biostatistics. Ed 3. Boston Mass, PWS-Kent.
Sandifer MG, Fleiss JL, Green LM (1968). Sample selection by diagnosis in clinical drug evaluations. Psychopharm. 13: 118-128.
Stigler SM (1983). Who Discovered Bayes' Theorem? Am. Stat. 37(4): 290-296.
Townsend RR, Simpson GF, Filly RA (1988). Membrane thickness in ultrasound prediction of Chorionicity of twin gestation. J. Ultrasound Med. 7: 327-332.
Vastola EF, Kokubu T (1962). Use of Bayes' theorem in Ophthalmodynamometry. Circulation. 26: 1312-1315.
Yang Q, Khoury MJ, Coughlin SS, Sun F, Flanders WD (2000). The use of population-based Registries in the clinical Validation of Genetic tests for Disease Susceptibility. National Office of Public Health Genomics. pp. 1-11.
Yerushalmy J (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. Pub. Health Rep. 62: 1432-1449a.