

Evaluating diagnostic tests

Johan F. Coetzee

Department of Anaesthesiology and Critical Care, School of Medicine, Faculty of Health Sciences, University of Stellenbosch, Stellenbosch, South Africa

Abstract

Anaesthesiologists are increasingly more involved in perioperative patient care wherein interpretation of special investigations is crucial to making therapeutic and prognostic decisions. Furthermore, anaesthetic journal publications increasingly rely on diagnostic tests, without paying sufficient attention to the methodology for evaluation of the predictive ability of these tests, particularly when conducting research. The purpose of this article is to provide an introduction to the principles underlying the objective appraisal of diagnostic tests and to provide basic tools for that purpose. Basic concepts about probability and conditional probability are introduced. An explanation is given how Bayes' theorem can be used to apply new information to improve uncertainty about a diagnosis. Various indices of diagnostic accuracy are clarified using simple probabilistic notation applied to a 2x2 table from which they can be calculated. Two clinical examples are employed to illustrate these concepts. The calculations are made easier using a spreadsheet that may be downloaded from the Internet and by using Fagan's nomogram. The use of receiver operating characteristic (ROC) curves for choosing a suitable cut-off point between positive and negative tests is explained.

Keywords: Diagnostic tests; Bayes Theorem; ROC Curve; Probability, Likelihood functions; Odds ratio

Diagnostic tests are used to help ascertain the presence or absence of a condition (disease). There is often uncertainty about how much reliance clinicians can place on a particular test for the purpose of establishing a diagnosis. For example blood concentrations of enzymes can be used to detect organ disease. This often means that test results are considered positive or negative beyond a predefined "cut-off" value. However patients who do not have the disease may have positive test results (false positive) and conversely, diseased persons may have negative test results (false negative). A good test should reliably indicate whether a disease is probable or unlikely. The methodology that is used to evaluate the predictive ability of diagnostic tests is well described in texts on medical statistics¹⁻⁶, but has received little attention in the anaesthetic literature. Anaesthesiologists are increasingly involved in perioperative patient care and often the interpretation of special investigations is crucial to making therapeutic and prognostic decisions. Furthermore, an increasing number of anaesthesiology-related journal articles relate to the development and evaluation of new diagnostic procedures. For example, a search through four leading journals* reveals that between 1991 and 2002 there were 67 articles that discussed special investigations by means of receiver operating characteristic (ROC) curves.

Appraising the predictive abilities of diagnostic tests can be surprisingly difficult. The purpose of this article is to provide an introduction to the principles underlying the objective evaluation of diagnostic tests. Starting from basic concepts about probability, Bayes' theorem will be introduced and an explanation given as to how it can be used in applying new information to improve the uncertainty about a diagnosis. An approach to the application and evaluation of diagnostic research is given. These concepts will be explained using simple probabilistic notation applied to a 2x2 table from which the necessary calculations can be made.

Basic concepts about probability

Probability expresses our degree of certainty about future events. The probability (denoted P) of an event is defined as the number of times we believe that it is likely to occur divided by the number of times that it could possibly occur.

$$P = \frac{\text{No. of times an event is likely to occur}}{\text{No. of times it could possibly occur}}$$

Conventionally, probability is expressed as a number between 0 and 1 (or as a percentage between 0 and 100%). An event with a probability of 0 will never happen and an event with probability 1 is certain to happen. For example when throwing a dice many times, a 4 is likely to occur on average 1 in 6 times: The probability is $1/6 = 0.167$ (Approximately 17% of the time). The probability of an event not occurring is $1 - P$. In the example of the dice the probability of not throwing a 4 is $1 - 0.167 = 0.833$.

* *Anesthesiology, Anesthesia and Analgesia, British Journal of Anaesthesia, Canadian Journal of Anaesthesiology*

Correspondence:

Prof Johan F. Coetzee
email: jfc@sun.ac.za

If two events are truly independent (i.e. the occurrence of one does not influence the occurrence of the other), the probability of either one occurring is the sum of the individual probabilities.

Thus

$$P(A) \text{ or } P(B) = P(A) + P(B) \quad \text{Equation-1}$$

Where P(A) and P(B) are the probabilities of events A and B respectively. The dice example reveals that the probability of throwing either a four or a six with a single throw is $1/6 + 1/6 = 1/3$ or 0.33. On the other hand, the probability of two events occurring simultaneously is the product of the individual probabilities. Thus

$$P(A) \text{ and } P(B) = P(A) \times P(B) \quad \text{Equation-2}$$

For example, the probability of throwing a pair of dice so that a four and a six occur simultaneously, is given by: $1/6 \times 1/6 = 1/36 = 0.028$.

Another method of expressing our degree of certainty about future events is the odds. The odds of an event is defined as the ratio of the number of times that the event is likely to occur to the number of times that it is not likely to occur.

$$\text{Odds} = \frac{P}{1 - P} \quad \text{odds}$$

$$\text{Odds can be converted to probability: } P = \frac{\text{odds}}{(1 + \text{odds})}$$

In the example of the dice, the odds of throwing a 4 are: $0.167 / 0.833 = 0.2$ (i.e. 1/5 or 1:5)

This can be interpreted as indicating that it is 5 times less likely to throw a 4 with a single throw than not to throw a 4.

Conditional probabilities and the definition of Bayes’ theorem

There are three important aspects that affect the ability of a test to predict the presence of a condition (e.g. a disease). These are the pretest probability, the sensitivity and the specificity. The pretest probability denotes the probability of the disease being present prior to testing. This could be the prevalence of the disease in the population, or it could be the index of suspicion that the clinician has estimated, based on his/her clinical history and examination.

The sensitivity of a test is defined as the proportion of persons who have the condition who test positively. This is also known as the true positive rate (TPR). The specificity of a test is defined as the proportion of persons without the disease who have a negative test result. This is also termed the true negative rate (TNR). Coupled to these two basic definitions are the concepts of the false positive rate and the false negative rate. The false positive rate (FPR) is the proportion of healthy persons (without the disease) who test positively. The false negative rate (FNR) is the proportion of diseased persons who test negatively. From these definitions, it can be deduced that $FNR = 1 - \text{sensitivity}$ and $FPR = 1 - \text{specificity}$.

These relationships are best understood (and remembered)

by constructing a 2 x 2 table (Table I). Referring to the 2 x 2 table, the columns denote the presence or absence of disease (D+ or D-) and the rows represent the positive and negative tests (T+ and T-). The prevalence (or pretest probability) of the disease is given by $(a+c) / (a+b+c+d)$, i.e. the proportion of the population who have the disease. In probability notation this is represented by P(D+). The number of true positives (TP) is denoted by a. The sensitivity, or true positive rate (TPR) is given by $a/(a+c)$. Likewise, specificity, the true negative rate (TNR), is calculated as $d/(d+b)$, and so forth for FPR and FNR (see Table I).

In probability notation, sensitivity (or TPR) = $P(T+|D+)$ where the "I" is read as "given". Likewise: False positive rate = $P(T+|D-)$ and specificity = $P(T-|D-)$. These are known as conditional probabilities. A conditional probability is defined as the probability that an event is true, on condition that another event is also true. Note that this is different from the two dice example where the two events (e.g. throwing a “4” and a “6” simultaneously) are completely independent. On the contrary, $P(T+|D+)$ expresses the probability of obtaining a positive test on condition that only patients who have the disease are tested. This is different from the probability of obtaining a positive test from an individual chosen from the population at random, which in this case is given by P(T+).

The probability of event A occurring conditional upon event B, i.e. $P(A|B)$, is formally defined using the product rule as described above in equation-2:

$$P(A \text{ and } B) = P(A|B).P(B) \quad \text{Equation-3}$$

Furthermore the ordering of the events is unimportant so that

$$P(A|B).P(B) = P(B|A).P(A) \quad \text{Equation-4}$$

Solving equation-3 for the conditional probability gives:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \quad \text{Equation-5}$$

Equation-5 is known as Bayes’ theorem.*

Application of Bayes’ theorem to diagnostic testing

Unfortunately, sensitivity and specificity alone do not estimate the likelihood of disease in a particular patient. For this it is necessary to combine these indices with the previous knowledge we have about the patient, in order to determine whether a test result indicates whether the disease in question is present or absent. What we want to know clinically, is what is the probability of disease given a positive test? i.e. $P(D+|T+)$. For example, given prior knowledge about the incidence of coronary heart disease in men over 50 years, we may wish to know what the probability is of coronary heart

* Thomas Bayes (1702-1761) was a non-conformist minister from Tunbridge Wells whose work on probability theory was discovered by his relatives after his death. With Bayes’ theorem, one can relate the knowledge one has about prior probability of an event to the probability of that event after the addition of new knowledge.

Table 1: Calculation of diagnostic accuracy parameters from a 2 x 2 table:

Tables 1(a) and 1(b) illustrate the arrangement of data into a 2 x 2 table for calculation of the indices of diagnostic accuracy.

Table 1(c) lists the various indices, together with their definitions using probability notation and the expressions for their calculation.

D+ = presence of disease; D- = absence of disease; T+ = positive test; T- = negative test.

a = true positives (TP); b = false positives (FP); c = false negatives (FN); d = true negatives (TN).

The total number of individuals who have the disease is (a + c), i.e. (TP + FN); the total number without the disease is (b + d), i.e. (TN + FP).

The total number in the sample tested are (a + b + c + d).

TP = true positives; FP = false positives; FN = false negatives; T- = true negatives.

	D+	D-	Row total
T+	a	b	a+b
T-	c	d	c+d
Column total	a+c	b+d	

	D+	D-	Row Total
T+	TP	FP	TP+FP
T-	FN	TN	FN+TN
Column total	TP+FN	FP+TN	

Name	Probability notation	Calculation	Synonyms / equivalents
Pretest probability	P(D+)	(a+c) / (a+b+c+d)	Prior probability, prevalence
Sensitivity	P(T+ D+)	a / (a + c)	True positive rate (TPR)
Specificity	P(T- D-)	d / (d + b)	True negative rate (TNR)
False Positive Rate (FPR)	P(T+ D-)	b / (b + d)	1 – specificity
False Negative Rate (FNR)	P(T- D+)	c / (a + c)	1 – sensitivity
Positive predictive value	P(D+ T+)	a / (a + b)	Posttest probability of a positive test
Negative predictive value	P(D- T-)	d / (c + d)	
Posttest probability of a negative test	P(D+ T-)	c / (c + d)	
Positive likelihood ratio (LR+)	P(T+ D+)	a / (a + c)	sensitivity
	P(T+ D-)	b / (b + d)	1 - specificity
Negative likelihood ratio (LR-)	P(T- D+)	c / (a + c)	1 - sensitivity
	P(T- D-)	d / (b + d)	specificity
Pretest odds	$\frac{P(D+)}{1 - P(D+)}$		
Posttest odds of a positive or negative test		Pretest odds x Likelihood ratio (Pos. or Neg.)	
Odds ratio	$\frac{P(T+ D+) / P(T- D+)}{P(T+ D-) / P(T- D-)}$	$\frac{a.d}{b.c}$ (Cross-product)	

disease after a positive exercise tolerance test? The prior knowledge is expressed by the prior probability of the disease, P(D+), which is also known as the pretest probability. As mentioned above, this may merely be the prevalence of the disease in the population under consideration, or it may be based upon the clinician's index of suspicion. P(D+|T+) is known as the posterior probability of the test, or the post-test probability. These questions of course are about conditional probability and therefore are appropriate for the application of Bayes' theorem.

By applying Bayes' theorem to equation-5, the following expression to calculate the posttest probability of the disease can be derived (see Appendix).

$$P(D+|T+) = \frac{P(D+).TPR}{P(D+).TPR + \{(1 - P(D+)).FPR\}} \quad \text{Equation A4}$$

Thus using Bayes' theorem, it is possible to calculate the post-test probability of disease, provided the following is known with regard to a diagnostic test: the pretest probability, the true positive rate (sensitivity), and the false positive rate (or the specificity which is 1 – FPR).

Likewise, it is possible to calculate the probability of disease if the test result is negative:

$$P(D+|T-) = \frac{P(D+).(1 - TPR)}{P(D+).(1 - TPR) + \{(1 - P(D+)).(1 - FPR)\}} \quad \text{Equation A5}$$

Example: In a 55 year-old man with haemoptysis and a history of cigarette smoking it is estimated that the pretest probability of lung cancer is 0.4 (i.e. P(D+)). If a mass lesion is found on chest X-ray examination, we can calculate the probability of the patient having lung cancer using equations A4 and A5, if we know the sensitivity and the specificity for that finding. It has been established that these values are: sensitivity (TPR) = 0.6 and specificity (1 - FPR) = 0.96.

$$P(D+|T+) = \frac{0.4 \times 0.6}{(0.4 \times 0.6) + \{(1 - 0.4) \times 0.04\}} = 0.91$$

$$P(D+|T-) = \frac{0.4 \times 0.4}{(0.4 \times 0.4) + \{(1 - 0.4)(1 - 0.04)\}} = 0.22$$

Therefore, in this patient, finding a mass lesion in a chest X-ray photograph practically rules in the presence of lung cancer, while the absence thereof reduces its probability by more than half, but does not rule it out and therefore the clinician should continue to suspect that it may indeed be present.

Bayes' theorem: The dependence of post-test probability upon pretest probability

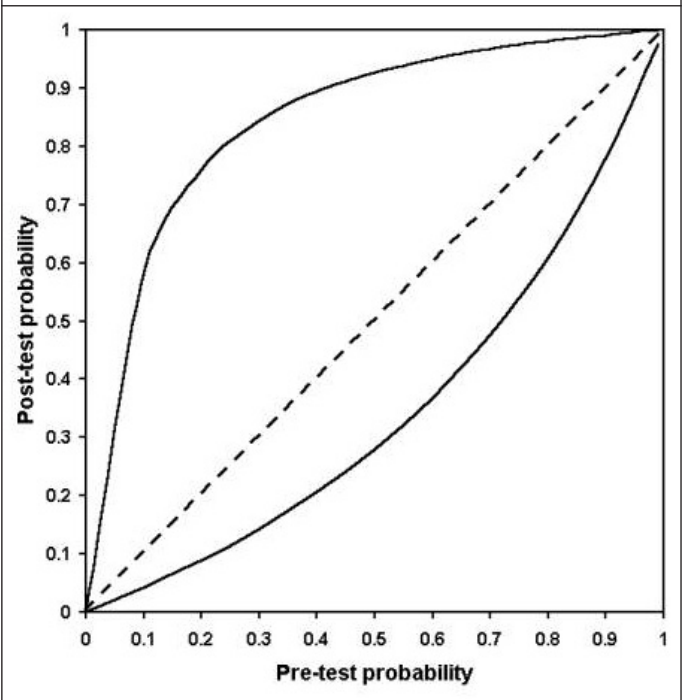
Continuing with the example of the chest X-ray, the probability of lung cancer in 55-year old non-smokers is much less than that of smokers. Let us assume that their pretest probability is 0.001. Calculation of the post-test probability using equations A4 and A5 now reveals that if a mass lesion is found on chest X-ray examination, the post-test probability of lung cancer is only 0.015 and that of a negative test is 0.004. This example illustrates an important point, namely that the post-test probability depends upon the pretest probability, the former increasing as the latter increases. Examining equation-A4 reveals that the posttest probability, P(D+|T+), is related to the pretest probability, P(D+) nonlinearly: Equation A4 has the form:

$$y = \frac{ax}{bx + c}$$

where y is the post-test probability, P(D+|T+) and x is the pretest probability, P(D+). The graph of this function is a rectangular hyperbola. Figure-1 depicts the post-test probability for all possible values of the pretest probability, using the example of an X-ray examination for lung cancer (TPR = 0.6, FPR = 0.04). The diagonal line in Figure-1, the line of identity, defines a post-test probability that is equal to the pretest probability for all values of the pretest probability; i.e. denoting a useless test. An ideal test would be for a positive test to have post-test values of 1 for all values of the pretest probability (i.e. lie along the upper abscissa) and for a negative test the post-test probability would always be zero (i.e. lie along the

lower abscissa). In real life where tests are less than perfect, the post-test probabilities for a positive result lie somewhere above the line of identity, i.e. a meaningful positive test increases the post-test probability (upper curve in Figure 1). For a negative result, the post-test probabilities lie below the line of identity, i.e. a negative test reduces the post-test probability (lower curve in Figure 1). Note that the curve for a negative test is not a mirror image of that for a positive test.

Figure 1: Post-test probabilities corresponding to positive and negative test results calculated for all possible pretest probabilities. The upper curve depicts posttest probabilities for positive tests; the lower curve for negative tests. The dotted diagonal line the line of identity, defines a posttest probability that is equal to the pretest probability for all values of the pretest probability; i.e. denoting a useless test. This graph corresponds to the example of the chest X-ray (see text) where the true positive rate = 0.6 and the false positive rate = 0.04



Bayes' theorem: The influence of the TPR and the FPR on the interpretation of diagnostic tests

It is apparent from equations A4 and A5 that besides being influenced by the pretest probability, the post-test probability is also affected by the TPR and the FPR. This is illustrated by Figure-2 which depicts curves for the chest X-ray example for FPR of 0.02, 0.04 and 0.1. Note that the interpretation of a positive test is heavily influenced by the FPR of the test, whereas a negative test is only slightly affected.* Figure-3 illustrates the influence of the TPR (sensitivity) on post-test probabilities for TPR of 0.9, 0.6 and 0.4. Here the interpretation of a positive test is little influenced by variation in TPR whereas a negative test is heavily influenced.

Using the 2x2 table to calculate the accuracy of diagnostic tests

Equations A4 and A5 (derived from Bayes' theorem) are useful to demonstrate how the post-test probabilities of diagnostic tests are influenced by the pretest probability the sensitiv-

* Note that one may also state that the posttest probability is influenced by the specificity of the test (specificity = 1 - FPR).

Figure 2: Graph depicting the influence of various false positive rates (FPR) on the relationship between posttest probabilities and pretest probabilities. (True positive rate = 0.6)

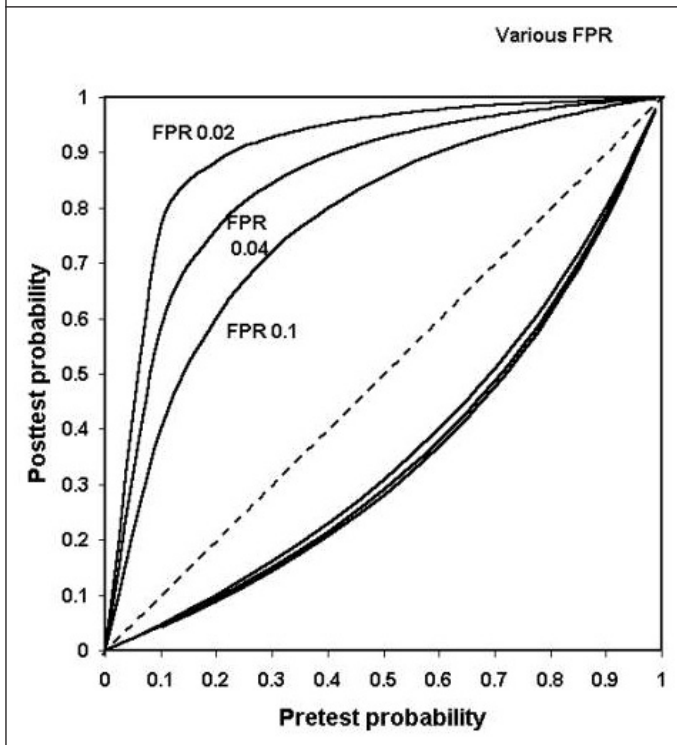
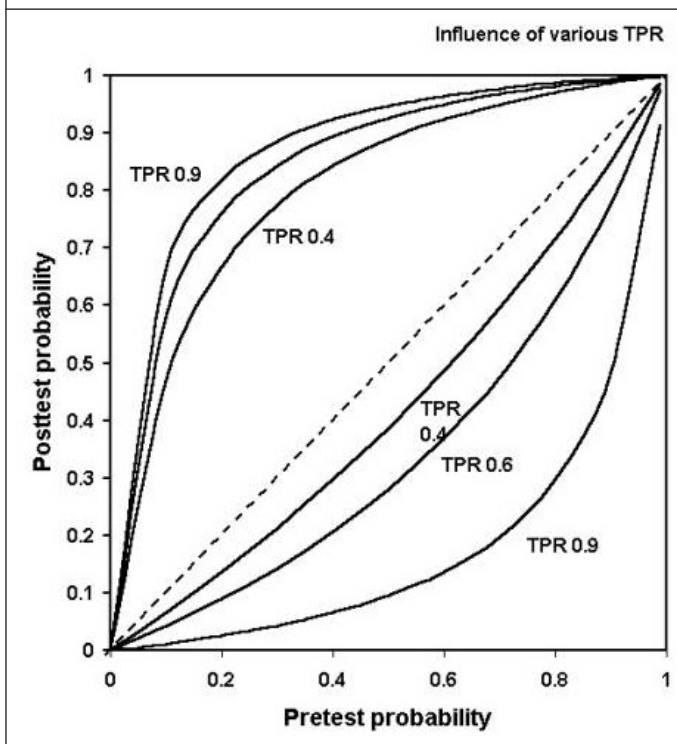


Figure 3: Graph depicting the influence of various true positive rates (TPR) on the relationship between posttest probabilities and pretest probabilities. (False positive rate = 0.04)



ity and the specificity. However when calculating the conditional probabilities of diagnostic tests, many clinicians find it simpler to use a 2x2 table. Referring to Table I, the post-test probability of a positive test is calculated by the expression $a/(a+b)$, i.e. $P(D+|T+)$. The post-test probability of a nega-

tive test is calculated by $c/(c+d)$, i.e. $P(D+|T-)$. If the 2x2 table comprises results from a study group, the above calculation is termed the Positive Predictive Value (PPV). The Negative Predictive Value (NPV) is the probability of disease being *absent* in the study group, given a negative test result, i.e. $P(D-|T-)$.

Discriminating between sick and healthy patients: Likelihood ratio

The usefulness of predicting posterior probabilities is limited, because of their susceptibility to influences by the prior probability of the disease. The likelihood ratio (LR) compares the probability for a positive test result in persons who have the disease with the probability for a positive test result in healthy persons. This ratio simplifies the calculation of predictive values and furthermore it is independent of the disease prevalence. There are two kinds of likelihood ratios, positive and negative. The positive likelihood ratio (LR+) is defined as the probability of a diseased person having a positive test result compared to the probability of a non-diseased person having a positive test result. The negative likelihood (LR-) ratio is defined as the probability of a diseased person having a negative test result compared to the probability of a non-diseased person having a negative test result.

In probability notation:

$$LR+ = \frac{P(T+|D+)}{P(T+|D-)} \quad LR- = \frac{P(T-|D+)}{P(T-|D-)}$$

Using the 2x2 table (Table I) the likelihood ratios can be calculated as follows:

$$LR+ = \frac{a / (a + c)}{b / (b + d)} \quad LR- = \frac{c / (a + c)}{d / (b + d)}$$

From the above definitions, it can be seen that likelihood ratios can also be expressed in terms of TPR, FPR, TNR and FNR, as well as sensitivity and specificity:

$$LR+ = \frac{TPR}{FPR} \quad LR- = \frac{FNR}{TNR}$$

$$LR+ = \frac{\text{sensitivity}}{(1 - \text{specificity})} \quad LR- = \frac{(1 - \text{sensitivity})}{\text{specificity}}$$

Likelihood ratios are a powerful tool for discriminating between sick and healthy patients and as mentioned, have the advantage that they are independent of the pretest probabilities. A LR+ of 10 means that a diseased patient is 10 times more likely to have a positive test than a healthy patient. Unfortunately they do not say much about post-test probabilities, so that the ratios need to be quite large in order to draw conclusions about post-test probabilities (Table II). Estimating the relationship between pretest probability, likelihood ratio and post-test probability is easily

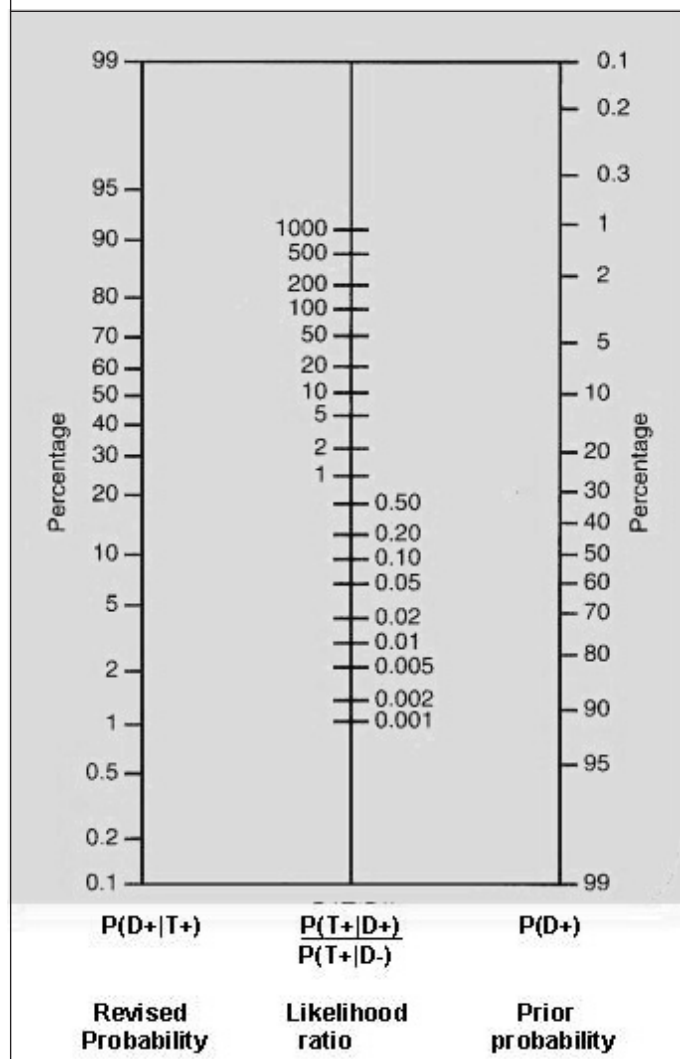
done using the nomogram of Fagan⁷ (Figure 4)*. Using this nomogram, it is possible to estimate post-test probabilities of disease for various pretest probabilities, after you have calculated the likelihood ratio. This powerful tool is equivalent to referring to the graphs of conditional probabilities (Figures 1-3). Both positive and negative likelihood ratios may be employed to obtain post-test probabilities for positive, as well as for negative test results.

In the previously discussed example of the chest X-ray ex-

Table II: Likelihood ratio as an index of the probability of disease:

Likelihood Ratio	Change in disease probability
>10 or <0.1	large change
5 – 10 or 0.1 – 0.2	moderate change
2 – 5 or 0.2 – 0.5	small change
2 and 0.5	little or no change
1	no change
LR = 0 excludes disease, infinity excludes normality	
Likelihood ratios (LR) need to be quite large in order to change the pretest and posttest probabilities of disease.	

Figure 4: Nomogram for converting pre-test probabilities to post-test probabilities for a diagnostic test result with a given likelihood ratio. (http://www.cebm.net/likelihood_ratios.asp)



amination of a 55 year old male smoker (where prevalence = 0.4, TPR = 0.6, FPR = 0.04): $LR+ = TPR / FPR = 0.6/0.04 = 15$. This implies that there is a 15x greater probability that a patient who has lung cancer will have a mass lesion on X-ray. On the other hand, $LR- = FNR / TNR = (1 - TPR) / (1 - FPR) = 0.4/0.96 = 0.42$. This low value implies that a patient without lung cancer has a (1 / 0.42) 2.4 x greater probability of having a negative result than a patient with lung cancer. Referring to Table II, this implies only a small change in the pretest estimation of the probability of lung cancer, as we already have seen in the calculation of the posttest probability according to Bayes' theorem.

A popular method for calculating the effect of a test result on the likelihood of disease being present is to calculate the post-test odds of the presence of disease⁸, which are defined as follows:

$$\text{Post-test odds} = \text{Pretest odds} \times \text{Likelihood ratio}$$

In the example of the chest X-ray examination the pretest odds are: $0.4 / (1 - 0.4) = 0.4 / 0.6 = 1:1.5$. After detecting a mass lesion, the post-test odds are $0.4 / 0.6 \times LR+ = 0.4 / 0.6 \times 15 = 10:1$. This can be interpreted as implying that in the presence of a mass lesion there will be 10 times as many persons with cancer than those without. If a mass lesion is not detected, then the post-test odds are $0.4 / 0.6 \times LR- = 0.4 / 0.6 \times .42 = 0.28 = 1 / 3.6$ or $1 : 3.6$. This can be interpreted as implying that in the absence of a mass lesion, there will be one person who has cancer for every 3.6 persons who does not.

An example: Estimating the predictive value of a test following a study that developed a test for difficult intubation.

Let us assume that a new test for predicting difficult intubation is reported from a study involving 1000 patients and that 100 of these patients (10%) presented difficult intubations. The sensitivity and specificity of the test was reported to be 95% and 90% respectively. We have sufficient information to construct a 2x2 table (Table III).

- Step 1: Total number of sick persons = (prevalence x overall total) = 100
- Step 2: Total number of healthy persons = (1000 – total of sick persons) = (1000 – 100) = 900
- Step 3: True positives (TP) = (sensitivity x total of sick persons) = (0.95 x 100) = 95
- Step 4: False negatives (FN) = (total of sick persons – true positives) = (100 – 95) = 5
- Step 5: True negatives (TN) = (specificity x total of healthy persons) = (0.9 x 900) = 810
- Step 6: False positives (FP) = (total of healthy persons – true negatives) = (900 – 810) = 90

Table I summarizes the predictive values that can be calculated from a 2x2 table. Sensitivity and specificity are characteristics of a test (not the study group to which the test is applied) and these judgments are made prior to performing the test. The ability of the test to predict the presence of a condition is obtained by calculating the predictive values. The positive predictive value

* This nomogram can be downloaded from the Internet website of the Oxford-Centre for Evidence-Based Medicine: http://www.cebm.net/likelihood_ratios.asp

Table III: A 2 x 2 table depicting the example of a test to predict difficult intubation (see text).

	Difficult intubation (D+)	Easy intubation (D-)	Total
Positive test (T+)	95 ³ (sensitivity = 95% of 100)	90 ⁶ (900 - 810)	185
Negative test (T-)	5 ⁴ (100-95)	810 ⁵ (Specificity = 90% of 900)	815
Total	100 ¹ (0.1 x 1000)	900 ² (1000 - 100)	1000

In the example, 1000 patients were evaluated before intubation and the results are depicted in the 2 x 2 table. The superscript numbers in bold script refer to the steps described in the text for calculating the 2 x 2 table, given the sensitivity and the specificity of the test.
 TP = True Positives; FP = False Positives; FN = False Negatives; TN = True Negatives

(PPV) is defined as the ratio between those subjects who actually have the disease and all the subjects who test positively. In Table I the PPV is given by $a/(a+b)$, i.e. $PPV = TP / (TP + FP)$ and it is numerically equal to the post-test probability of Bayes' theorem. However, although numerically the same, there are subtle differences as discussed below. The negative predictive value (NPV) defines the ability of a negative result to define those individuals who do not have the condition. It is defined as the ratio between those subjects who do not have the disease to all of those who tested negatively. In Table I it is given by $d/(c+d)$, i.e. $NPV = TN / (TN + FN)$.

The results of the calculated parameters for estimating the diagnostic accuracy of the test to predict difficult intubation* are presented in Table IV. In our example, the positive predictive value is $95 / (95 + 90) = 0.514$. Therefore, if we get a

positive result, we are only 51.4% sure that there will be a difficult intubation (not much better than flipping a coin). The negative predictive value is $810 / (5 + 810) = 0.994$. Therefore, in this example, a negative result virtually rules out the possibility that a difficult intubation will occur. This case in point illustrates that a 95% sensitivity does not necessarily mean that we have a 95% reliability for diagnosing sick persons because the predictability depends on the prevalence of the disease. As the prevalence decreases, the PPV decreases and the NPV increases, as described above in the discussion of Bayes' theorem (Figure 1). The converse is true when the prevalence increases. A low PPV implies a high false-positive rate and it is important that persons are not unnecessarily alarmed by the occurrence of a positive test, when in fact they do not have the disease!

The LR+ of 9.5 indicates that a person with a difficult intubation was 9.5 times more likely to have exhibited a positive test than a person who did not have a difficult intubation. On the other hand, the LR- of 0.056 indicates that a person with a difficult intubation is 18 times less likely to test negatively than a person who does not experience a difficult intubation. The post-test odds of 1.056 provide essentially the same information as the PPV, namely that after obtaining a positive test, there is approximately a 50:50 chance that intubation will be difficult.

This is one of the problems that arise using screening tests in a low prevalence environment. Although a negative test can virtually rule out the disease, a positive test with a low PPV does not mean very much and can lead to unnecessary anxiety, as well as expensive further testing. Screening tests should preferably have a high PPV. As a general rule, screening tests should give the assurance that a negative result is really negative i.e. there should be very few false negative results. Therefore a highly sensitive test is preferred. A useful acronym is SNOUT i.e. to use a SeNsitive test to rule OUT disease. On the other hand, if there is evidence of disease, a test that results in very few false positive results is preferred. Therefore a highly specific test is favoured. Here the acronym is SPIN

Table IV: Parameters of diagnostic accuracy for the prediction of difficult intubation example.

Parameter	Result
Sample size	1000
Prevalence	0.1
Sensitivity	0.95
Specificity	0.90
Positive predictive value (PPV)	0.514
Negative predictive value (NPV)	0.994
Positive likelihood ratio (LR+)	9.5
Negative likelihood ratio (LR-)	0.056
Pretest odds	0.11
Posttest odds of a positive test	1.06
Posttest odds of a negative test	0.006

95% Confidence Interval (CI) for Sensitivity^ψ: 0.91 - 0.99
 95% Confidence Interval (CI) for Specificity^φ: 0.88 - 0.92

$${}^{\psi}95\% \text{ CI} = \text{Sens} \pm \left[1.96 \times \sqrt{\frac{\text{Sens} \times (1 - \text{Sens})}{\text{TP} - \text{FN}}} \right]$$

$${}^{\phi}95\% \text{ CI} = \text{Spec} \pm \left[1.96 \times \sqrt{\frac{\text{Spec} \times (1 - \text{Spec})}{\text{FP} - \text{TN}}} \right]$$

* A useful Microsoft Excel worksheet ("diagnostic.xls") created by PM MacEaney and DE Malone¹¹ can be accessed and/or downloaded from the Internet (<http://1216.71.117.81/spreadsheets/diagnostic.xls>) (or obtained from the author). This worksheet calculates predictive values from a 2x2 table and performs Bayesian analysis.

because a Specific test rules IN disease.

Outwardly it appears that predictive values derived from a 2 x 2 table after a study are identical to the posterior probabilities as described above in the discussion on Bayes' theorem: indeed they are numerically equal. They are, however, not quite the same. Sox et al.⁵ state: "Both answer the question 'Given this test result, what is the likelihood that my patient has the disease?' However there are important differences between posterior probability and predictive value. Posterior probability is a far more useful means of expressing uncertainty.

Predictive value is:

- Defined as the proportion of **study patients** with a test result who have disease (or no disease).
- Calculated from a 2 x 2 table of results in a defined population of patients
- Dependent on the prevalence of disease in the defined (study) population.

Posterior probability is:

- Defined as the probability of disease after new information is taken into account.
- Calculated from Bayes' theorem
- Dependent on the prior probability of disease, which is defined as an opinion about the likelihood of an event prior to the receipt of new information.

Thus the predictive value is an observable number obtained from a defined population and does not necessarily apply to another population. If the true positive rate and the false positive rate of a clinical finding are known, posterior probability can be used to interpret the finding in any population." The authors discuss the pitfalls of using predictive values as surrogates for posterior probabilities. These mainly concern the design of the study and the application of the findings from one population to another population. Rheeder and Ker⁴ have summarized guidelines on evaluating an article on diagnostic research from the literature^{9,10} and these are presented in Table V.

As with Bayes' theorem, calculation of likelihood ratios

Table V: Evaluating an article on diagnostic research:
<p><i>Is the study valid?</i></p> <ol style="list-style-type: none"> 1. Was there an independent, blind comparison with a reference (gold) standard of diagnosis? 2. Was the diagnostic test evaluated in an appropriate spectrum of patients (such as those in whom it would be used in practice)? 3. Was the reference standard applied regardless of the diagnostic test result (avoiding verification bias) 4. Were the test's methods described clearly enough to permit replication? <p><i>What are the results and their level of precision?</i></p> <ol style="list-style-type: none"> 1. In what form are the results given and how useful are they? (sensitivity / specificity, predictive values, likelihood ratios, odds ratios) 2. Are confidence intervals provided around these mean estimates? <p><i>Can I use these results in clinical practice?</i></p> <ol style="list-style-type: none"> 1. Will the test be reproducible and well interpreted in my practice setting? 2. Are the results applicable to my patients? 3. Will the test results change my management? 4. Will my patients be better off because of the test?

Figure 5: If the distributions of the healthy and the diseased populations are completely separate, then a cut-off point may be selected whereby there would be no false positive or false negative results.

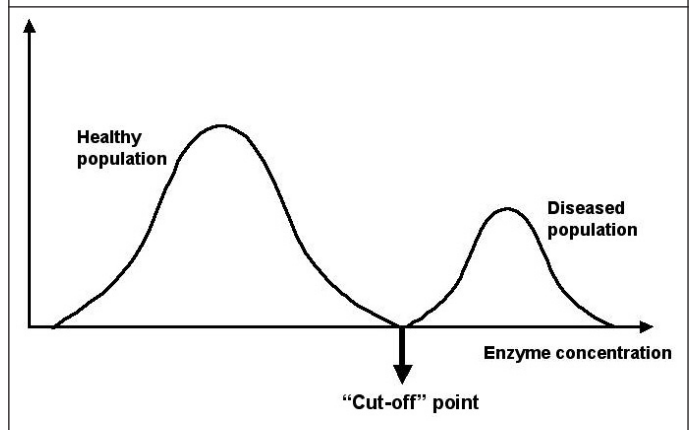
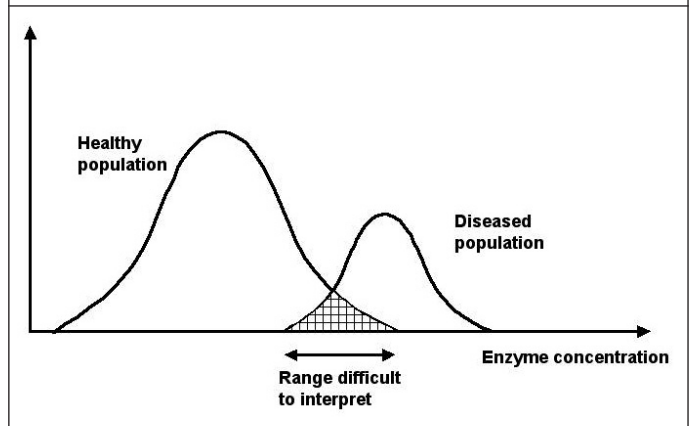


Figure 6: In real life the distributions of the healthy and diseased populations overlap, resulting in false positive and false negative results.



from the 2 x 2 table of a study group is helpful in distinguishing between sick and healthy persons and are independent of the prevalence of the disease in question.

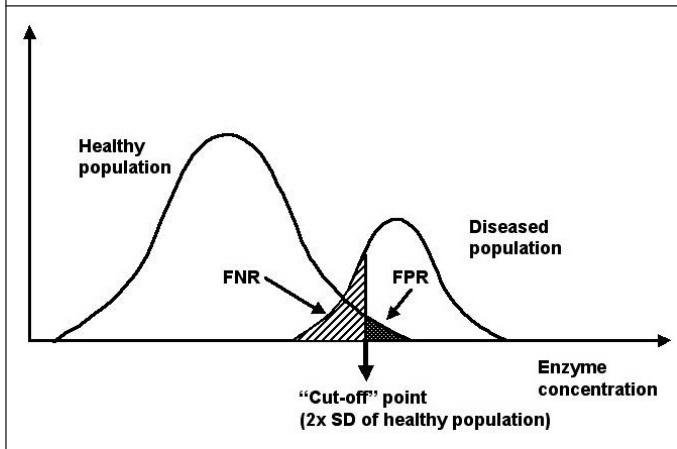
Evaluating diagnostic research: The dilemma of deciding when a test is abnormal

In the preceding discussion, methods were presented whereby it is possible to draw conclusions about the probability of disease after performing a diagnostic test. The assumption was that an optimum cut-off value had been selected to distinguish between positive and negative tests. This section will discuss the problem of finding an optimum value to define the difference between positive and negative test results.

If the test is a continuous variable with a normal distribution, a perfect test can be represented by Figure 5 which illustrates that for an ideal test the distributions of the healthy and the sick populations do not overlap. A "cut-off" point can therefore be chosen that always distinguishes perfectly between sick and healthy individuals. Such tests are rare and often expensive or dangerous. In reality, most distribution curves overlap as depicted by Figure 6. Low values unmistakably denote absence of disease and high values unmistakably indicate disease. However, the area in which the two distribution curves overlap, causes the greatest difficulties in interpretation.

A good test has high sensitivity (TPR) and specificity (TNR). A test with high sensitivity is "sensitive to disease"

Figure 7: In selecting a suitable cut-off point, there is a trade-off between sensitivity and specificity. Shifting the cut-off point to the left to obtain greater sensitivity, results in a greater number of false positive tests (i.e. reduced specificity). On the other hand, shifting the cut-off point to the right to obtain greater specificity, results in a greater number of false negative tests (i.e. reduced sensitivity).



and has a low false negative rate. A test with high specificity is "specific to health" and has a low false positive rate. There is, however, a trade-off between sensitivity and specificity as is illustrated by Figure 7. Laboratories usually report normal values as a mean and a standard deviation (SD). A reasonable value for a cut-off point to distinguish between a normal and an abnormal result could perhaps be 2x standard deviations from the mean of the curve for the healthy patients. In this example, the dark shaded area indicates the FPR and the striped area, the FNR. It is evident that the diagnosis would be missed in a large proportion of sick patients because of a large FNR. If one shifts the cut-off point leftwards to decrease the FNR, the FPR increases and vice versa.

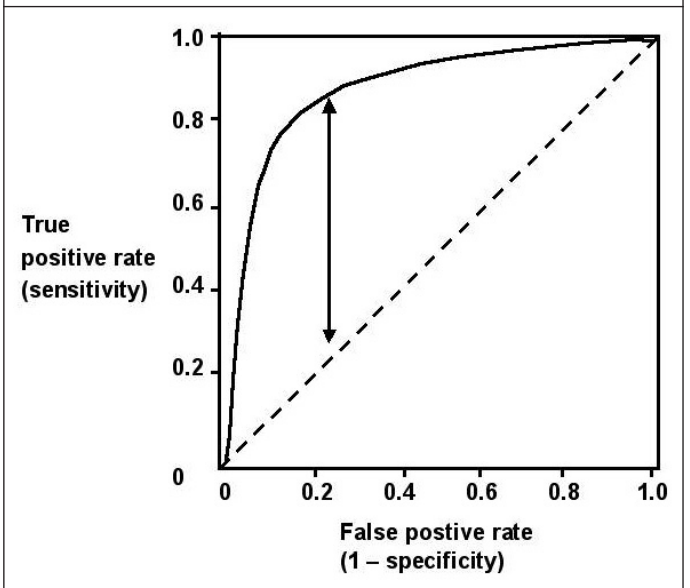
Receiver-operating characteristic (ROC) curves

The relationship between sensitivity and specificity is well illustrated by the use of ROC curves. The curve is obtained by plotting the true positive rate (sensitivity) of the test for all possible cut-off points on the ordinate versus the false positive rate (1 – specificity) on the abscissa (Figure 8). A test that performs well in discriminating between those with and without the disease will have a curve that deviates sharply upwards and leftwards: i.e. the greater the area under the curve, the better the ability of the test to distinguish between diseased and non-diseased states. The dotted diagonal line ("the line of unity") depicts a test that cannot distinguish between individuals with and without disease (i.e. a test result that occurs purely by chance). The diagonal line yields an area below it of 0.5 (50%). In choosing a suitable cut-off value from an ROC, one usually chooses a point on the ROC curve for which the vertical distance from the point to the line of unity is the greatest (double-headed arrow in Figure 8). It can be seen that an ROC curve is a graphical depiction of the positive likelihood ratio of a test {LR+ = sensitivity / (1 – specificity)}:

The Odds Ratio

A statistic often used in case-control studies by epidemiologists is the Odds Ratio of a test or of a risk factor. The odds of obtaining a positive test result in persons who have the disease are:

Figure 8: Receiver operating characteristic curve (ROC) curve. The dotted diagonal line indicates where a test has a 50% post-test probability, i.e. a test that is of no use. The double-sided arrow indicates the maximum distance from the diagonal line, an optimum cut-off point for a test.



Probability of obtaining a positive test result in diseased persons

Probability of obtaining an negative result in diseased persons

$$= \frac{a / (a + c)}{c / (a + c)} = a/c$$

(Reminder: the odds of an event occurring is the ratio of the probability of it occurring to the probability of it not occurring). Likewise the odds of obtaining a positive result in healthy persons is

$$= \frac{b / (b + d)}{d / (b + d)} = b/d$$

The odds ratio for obtaining a positive test result is defined as the following ratio:

Odds of obtaining a positive result in a diseased person


Odds of obtaining a positive result in a non-diseased person

$$\text{Referring to the 2 x 2 table (Table II), the odds ratio} = \frac{a/c}{b/d} = ad/bc$$

The odds ratio is also known as the cross-product ratio because it can be obtained by calculating the ratio of the product of the diagonals in a 2 x 2 table.

	D+	D-
T+	a	b
T-	c	d

References

1. Campbell MJ, Machin D. *Medical Statistics A Commonsense Approach, 3rd Edition*. Chichester, John Wiley & Sons Ltd., 1999, pp 36-48.
2. Cruickshank S. *Mathematics and Statistics in Anaesthesia*. Oxford, Oxford University Press, 1998, pp 177-182.
3. Dawson D. *Basic & Clinical Biostatistics, 4th Edition*. New York, Lange Medical Books/McGraw-Hill, 2004, pp 302-31.
4. Rheeder P, Ker JA. What is wrong with my patient? How to read an article concerning diagnosis. *S.Afr.Med J.* 2002; 92:144-148.
5. Sox HC, Blatt A, Higgins MC, Marton KI. *Medical Decision Making*. Boston, Butterworths, 1988.
6. Oxford Centre for Evidence-Based Medicine. *Centre for Evidence-Based Medicine. Internet Communication*. 2004
7. Fagan TJ. Letter. Nomogram for Bayes theorem. *N.Engl.J.Med* 1975; 293:257
8. Gallagher EJ. Clinical utility of likelihood ratios. *Ann.Emerg.Med* 1998; 31:391-397
9. Greenhalgh T. How to read a paper. Papers that report diagnostic or screening tests. *BMJ* 1997; 315: 540-543
10. Jaeschke R, Guyatt G, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. A. Are the results of the study valid? *Evidence-Based Medicine Working Group. JAMA* 1994; 271: 389-391
11. MacEneaney PM, Malone DE. The meaning of diagnostic test results: a spreadsheet for swift data analysis. *Clin Radiol.* 2000; 55: 227-35. 

Appendix: Derivation of an expression to calculate posttest probabilities of diagnostic tests using Bayes' theorem.

The probability of an event A occurring conditional upon event B i.e. P(A|B) is given by Bayes' theorem which states:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \quad \text{Equation-5}$$

The posttest probability of a test is given by P(D+|T+) and applying Bayes' theorem we obtain:

$$P(D+|T+) = \frac{P(D+ \text{ and } T+)}{P(T+)} \quad \text{Equation A1}$$

According to the multiplication rule (equation-2):

$$P(D+ \text{ and } T+) = P(D+|T+).P(T+) = P(T+|D+).P(D+)$$

Equation A1 now becomes:

$$P(D+|T+) = \frac{P(T+|D+).P(D+)}{P(T+)} \quad \text{Equation A2}$$

P(T+) is the probability of a positive test occurring in any individual chosen at random and can be either a true positive or a false positive test: i.e. P(T+) = {P(T+ and D+)} + {P(T+ and D-)} (summation rule, equation-1, similar to the dice example above)

According to the product rule:

$$P(T+ \text{ and } D+) + P(T+ \text{ and } D-) = P(T+|D+).P(D+) + P(T+|D-).P(D-)$$

Substituting into equation A2:

$$P(D+|T+) = \frac{P(T+|D+).P(D+)}{P(T+|D+).P(D+) + P(T+|D-).P(D-)}$$

We can express the absence of disease, P(D-), as 1 - P(D+) and substituting we obtain:

$$P(D+|T+) = \frac{P(T+|D+).P(D+)}{P(T+|D+).P(D+) + P(T+|D-).{1 - P(D+)}} \quad \text{Equation A3.}$$

Substituting the definitions for TPR and FPR into equation A3 (Table I) we obtain:

$$P(D+|T+) = \frac{P(D+).TPR}{P(D+).TPR + \{(1 - P(D+)).FPR\}} \quad \text{Equation A4.}$$

Likewise it is possible to calculate the probability of disease if the test result is negative:

$$P(D+|T-) = \frac{P(D+).TPR}{P(D+).(1 - TPR) + \{(1 - P(D+)).(1 - FPR)\}} \quad \text{Equation A5.}$$