

Dichotomous tests

Introduction

In this chapter and the next, we move from assessing test reliability (reproducibility) to assessing accuracy. We are no longer comparing repeated administrations of an imperfect test or comparing one imperfect test with another; we are comparing the test result to the patient's true disease state (D+ or D-) in order to quantify how often the test gives the right answer. This requires that we have a "gold standard" (also known as "reference standard") against which to compare our test. In this chapter, we introduce these concepts with dichotomous tests. In Chapter 4, we extend the discussion to multilevel and continuous tests.

Sensitivity, specificity, prevalence, predictive value, and accuracy

We will review the definitions of sensitivity, specificity, prevalence, predictive value and accuracy using as an example the evaluation of a rapid bedside test for influenza virus reported by Poehling et al. (2002). Simplifying somewhat, the study compared results of a rapid bedside test for influenza called QuickVue with the true influenza status in children hospitalized with fever or respiratory symptoms. The authors used as a gold standard for diagnosing influenza either a positive viral culture or two positive polymerase chain reaction tests. We present the data using just the polymerase chain reaction test results as the gold standard.¹ The results were as shown in Table 3.1.

"Sensitivity" is the probability that a patient with the disease will have a positive test. In this case, there were eighteen patients with influenza, of whom fourteen had a positive test, so the sensitivity was $14/18 = 78\%$. A mnemonic for sensitivity is PID, which stands for "Positive In Disease." (This is easy to remember because the other PID, Pelvic Inflammatory Disease, is a problem that requires clinician sensitivity.) A

¹ Although the authors' gold standard may have been more appropriate, using it resulted (by coincidence) in the number of false positives exactly equalling the number of false negatives, which is potentially confusing.

Table 3.1. Results of “QuickVue” influenza test in a 2 × 2 table

	Flu+	Flu–	Total
Test+	14	5	19
Test–	4	210	214
Total	18	215	233

perfectly sensitive test (sensitivity = 100%) will never give a false negative (never be negative in disease), so a “perfectly **S**ensitive test, when **N**egative, rules **O**UT disease” (mnemonic, SnNOUT). An example would be the highly sensitive urine pregnancy test in a young woman with abdominal pain, where the disease in question is ectopic pregnancy. A negative urine pregnancy test rules out ectopic pregnancy.

“Specificity” is the probability that a patient without the disease will have a negative test. In our example above, there were 215 patients without the disease, of whom 210 had a negative test, so the specificity was $210/215 = 98\%$. A mnemonic for specificity is NIH for “**N**egative **I**n **H**ealth.” (Remember this by recalling that the other NIH, the National Institutes of Health, are very specific in their requirements on grant applications.) A perfectly specific test (specificity = 100%) will never give a false positive (never be positive in health), so a “perfectly **S**pecific test, when **P**ositive, rules disease **I**N (SpPIN). An example of this would be “pathognomonic” findings, such as visualization of head lice for that infestation or gram-negative diplococci on gram stain of the cerebrospinal fluid for meningococcal meningitis. These findings are highly specific; they never or almost never occur in patients without the disease, so their presence rules the disease in. Note that, although “NIH” is a helpful way to remember specificity, we want the test not just to be “negative in health,” we want it to be negative in everything that is not the disease being tested for, including other diseases that may mimic it.

“Prevalence” is the proportion of patients in the at-risk population who *have* the disease *at one point in time*. It should not be confused with “incidence,” which is the proportion who *get* the disease *over a period of time*. In Table 3.1, there were 233 children hospitalized for fever or respiratory symptoms, of whom 18 had the flu. In this population, the prevalence of flu was $18/233$ or 7.7%.

“Positive predictive value” is the probability that a patient with a positive test has the disease. In Table 3.1, there are nineteen patients with a positive test, of whom fourteen had the disease, so the positive predictive value was $14/19 = 74\%$. This means that, in a population like this one (hospitalized children with fever or respiratory symptoms), about three out of four patients with a positive bedside test will have the flu.

“Negative predictive value” is the probability that a patient with a negative test does not have the disease. In Table 3.1, there were 214 patients with a negative test, of whom 210 did not have the flu, so the negative predictive value was $210/214 = 98\%$. This means that, in a population such as this one, the probability that a patient

with a negative bedside test does not have the flu is about 98%.² Another way to say this is: the probability that a patient with a negative test *does* have the flu is about $100\% - 98\% = 2\%$.

The negative predictive value of 98% is not as good as it sounds. The probability of flu before obtaining the test result, the pre-test probability, was already only 7.7%, so the probability that a patient in this study did not have the flu was already $100\% - 7.7\% = 92.3\%$ before the test was done. Negative predictive values will always be high when the pre-test probability of disease is low.

“Accuracy” has both general and more precise definitions. We have been using the term “accuracy” in a general way to refer to how closely the test result agrees with the true disease state as determined by the gold standard. The term accuracy also refers to a specific numerical quantity: the percent of all results that are correct. In other words, accuracy is the sum of true positives and true negatives, divided by the total number tested. Table 3.1 shows 14 true positives and 210 true negatives out of 233 tested. The accuracy is therefore $(14 + 210)/233 = 96.1\%$.

Accuracy can be understood as a prevalence-weighted average of sensitivity and specificity:

$$\text{Accuracy} = \text{Prevalence} \times \text{Sensitivity} + (1 - \text{Prevalence}) \times \text{Specificity}.$$

Although completeness requires that we provide this numerical definition of accuracy, it is not a particularly useful quantity. Because of the weighting by prevalence, for all but very common diseases, accuracy is mostly determined by specificity. Thus, a test for a rare disease can have extremely high accuracy just by always coming out negative.

Importance of the sampling scheme

It is not always possible to calculate prevalence and positive and negative predictive values from a 2×2 table as we did above. What if this study had sampled children with and without flu separately (a “case-control” sampling scheme) with one non-flu control for each of the 18 patients with the flu, as in Table 3.2?

Table 3.2. Sample 2×2 table for the flu test when subjects with and without flu are sampled separately, leading to meaningless “prevalence” of 50%

	Flu+	Flu–	Total
Test+	14	1	15
Test–	4	17	21
Total	18	18	36

² It is just a coincidence that the negative predictive value $210/215$ and the specificity $210/214$ both round to 98%. As we shall see, the probability that a patient without the disease will have a negative test (specificity) is *not* the same as the probability that a patient with a negative test does not have the disease (negative predictive value).

Box 3.1: Dichotomous tests: definitions

	Disease+	Disease–	Total
Test +	a True Positives	b False Positives	a + b Total Positives
Test–	c False Negatives	d True Negatives	c + d Total Negatives
Total	a + c Total With Disease	b + d Total Without Disease	a + b + c + d Total N

Sensitivity: the probability that the test will be positive in someone with the disease: $\frac{a}{a + c}$

Mnemonics: PID = Positive In Disease; SnNOUT = Sensitive tests, when Negative, rule OUT the disease

Specificity: the probability that the test will be negative in someone who does not have the disease: $\frac{d}{b + d}$

Mnemonics: NIH = Negative In Health; SpPIN = Specific tests, when Positive, rule IN a disease

The following four parameters can be calculated from a 2 × 2 table only if there was cross-sectional sampling:

Prevalence: the probability of disease in the entire population: $\frac{a + c}{a + b + c + d}$.

Positive Predictive Value: the probability that a person with a positive test has the disease: $\frac{a}{a + b}$.

Negative Predictive Value: the probability that a person with a negative test does NOT have the disease: $\frac{d}{c + d}$.

Accuracy: the proportion of those tested in which the test gives the correct answer: $\frac{a + d}{a + b + c + d}$.

We could still calculate the sensitivity as $14/18 = 78\%$ and would estimate specificity as $17/18 = 94\%$, but calculating the prevalence as $18/36 = 50\%$ is meaningless. The 50% proportion was determined by the investigators when they decided to have one non-flu control for each flu patient; it does not represent the proportion of the at-risk population with the disease. When patients are sampled in this case-control fashion, we cannot estimate prevalence or positive or negative predictive value – both of which depend on prevalence.³ Calculating prevalence, positive predictive value, and negative predictive value from a 2 × 2 table requires sampling the D+ and D– patients together from a whole population, rather than sampling separately by disease status. This is called “cross-sectional” (as opposed to “case-control”) sampling.

³ “Accuracy” also depends on prevalence, but as mentioned above, it is not a useful quantity.

Box 3.2: Brief digression: the “|” symbol

The “|” symbol is used to represent a conditional probability. It is read “given.” The expression $P(A|B)$ is read “the probability of A given B” and means the probability of A being true (or occurring) if B is known to be true (or to occur). Here are some examples:

$P(\text{Headache}|\text{Brain tumor}) = \text{Probability of headache given that the patient has a brain tumor} = 0.7.$

$P(\text{Brain tumor}|\text{Headache}) = \text{Probability of a brain tumor given that the patient has a headache} = 0.001.$

Note, as illustrated above, $P(A|B)$ will generally be quite different from $P(B|A)$.

Using the “|” symbol,

Sensitivity = $P(+|D+)$ = Probability of a positive test given disease.

Specificity = $P(-|D-)$ = Probability of a negative test given no disease.

Positive Predictive Value = $P(D + |+)$ = Probability of disease given a positive test.

Negative Predictive Value = $P(D - |-)$ = Probability of no disease given a negative test.

Combining information from the test with information about the patient

We can express a main idea of this book as:

What you thought before + New information = What you think now

This applies generally, but with regard to diagnostic testing, “what you thought before” is also known as the prior (or pre-test) probability of disease. This is the probability that the patient had the disease before the test result was known. For screening tests, this is often just the prevalence of disease. For diagnostic tests, it will depend on the patient’s signs and symptoms and other test results, and on how much these suggest the disease in question.

“What you think now” is also known as the posterior (or post-test) probability of disease. Now that you know the test result, what is your revised estimate of the probability of disease? In the case of a positive dichotomous test, this is the same as positive predictive value. What about the posterior probability of disease after a negative test? This can be confusing. It is the probability that a patient with a negative test *has* the disease. Hence, it is $1 - \text{Negative Predictive Value}$. (The negative predictive value is the probability that a patient with a negative test *does not have* the disease.) We will spend a fair amount of time in this and the next chapter discussing how to use the result of a diagnostic test to update the prior probability and obtain the posterior probability of disease. The first method that we will discuss is the “2 × 2 Table Method”; the second uses likelihood ratios.

2 × 2 table method for updating prior probability

This method uses the sensitivity and specificity of a test to fill in the 2 × 2 table that would result if the test were applied to an entire population with a given prior

probability of disease. Thus, we assume either that the entire population is studied or that a random sample is taken, so that the proportions in the “Disease” and “No Disease” columns are determined by the prior probability of disease, P(D+). As mentioned above, this is referred to as cross-sectional sampling, because subjects are sampled according to their frequency in the population, not separately based on either disease status or test result.

The formula for posterior probability after a positive test is:

$$\frac{\text{Sensitivity} \times \text{Prior Probability}}{\text{Sensitivity} \times \text{Prior Probability} + (1 - \text{Specificity}) \times (1 - \text{Prior Probability})}$$

To understand what is going on, it helps to fill the numbers into a 2 × 2 table, as shown in a step-by-step “cookbook” fashion in Example 3.1.

Example 3.1 2 × 2 Table Method Instructions for Screening Mammography Example

One of the clinical scenarios in Chapter 1 involved a 45-year-old woman who asks about screening mammography. If this woman gets a mammogram and it is positive, what is the probability that she actually has breast cancer? Based on Kerlikowske et al. (1996a, 1996b), the prevalence of undetected invasive breast cancer in previously unscreened women at age 45 is about 2.8/1000, that is, 0.28%. The sensitivity of mammography is about 75% and the specificity about 93%. Here are the steps to get her posterior probability of breast cancer:

1. Make a 2 × 2 table, with “Disease” and “No Disease” on top and “Test+” and “Test – ” on the left, like the one below.

Blank 2 × 2 Table to Use for Calculating Posterior Probability

	Disease	No Disease	Total
Test +			
Test –			
Total			

2. Put a large, round number below and to the right of the table for your total N. We’ll use 10,000.
3. Multiply that number by the prior probability (prevalence) of disease to get the left column total, the number with disease or (a + c). In this case, it is $2.8/1000 \times 10,000 = 28$.
4. Subtract the left column total from the total N to get the total number without disease (b + d). In this case, it is $10,000 - 28 = 9972$.
5. Multiply the “total with disease” (a + c) by the sensitivity, $a/(a + c)$ to get the number of true positives (a); this goes in the upper left corner. In this case, it is $28 \times 0.75 = 21$.
6. Subtract this number (a) from the “total with disease” (a + c) to get the false negatives (c). In this case, it is $28 - 21 = 7$.
7. Multiply the “number without disease” (b + d) by the specificity, $d/(b + d)$, to get the number of true negatives (d). Here, it is $9972 \times 0.93 = 9274$.

8. Subtract this number from the “total without disease” ($b + d$) to get the false positives (b). In this case, $9972 - 9274 = 698$.
9. Calculate the row totals. For the top row, $21 + 698 = 719$. For the bottom row, $7 + 9274 = 9281$.

The completed table is shown below.

Completed 2×2 Table to Use for Calculating Posterior Probability

	Breast (Cancer)	No Breast Cancer	Total
Mammogram (+)	21	698	719
Mammogram (-)	7	9274	9281
Total	28	9972	10,000

10. Now you can get posterior probability from the table by reading across in the appropriate row and dividing the number with disease by the total number in the row with that result. So the posterior probability if the mammogram is positive (positive predictive value) = $21/719 = 2.9\%$, and our 45-year-old woman with a positive mammogram has only about a 2.9% chance of breast cancer!

If her mammogram is negative, the posterior probability ($1 - \text{Negative Predictive Value}$) is $7/9281 = 0.075\%$, and the negative predictive value is $1 - 0.075\% = 99.925\%$. This negative predictive value is very high. However, this is due more to the very low prior probability than to the sensitivity of the test, which was only 75%.

Likelihood ratios for dichotomous tests

One way to think of the likelihood ratio is as a way of quantifying how much a given test result changes the probability of disease in your patient. More exactly, it is the factor by which the odds of disease either increase or decrease as a result of your test. (Note the distinction between odds and probability below). There are two big advantages to using likelihood ratios to calculate posterior probability. First, as discussed in the next chapter, unlike sensitivity and specificity, likelihood ratios work for nondichotomous tests. Second, they simplify the process of estimating posterior probability.

You have seen that it is possible to get posterior probability from sensitivity, specificity, prior probability, and the test result by filling in a 2×2 table. You have also seen that it is kind of a pain. We would really love to just multiply the prior probability by some constant derived from a test result to get the posterior probability. For instance, wouldn't it be nice to be able to say that a positive mammogram increases the probability of breast cancer about 10-fold, or that a white blood cell count of more than 15,000 triples the probability of bacteremia?

But there is a problem with this: probabilities cannot exceed one. So if the prior probability of breast cancer is greater than 10%, there is no way you can multiply it by ten. If the prior probability of bacteremia is more than one-third, there is no way

Box 3.3: Avoiding a common error: be clear on the denominator of “False Positives” and “False Negatives”!

A common source of confusion arises from the inconsistent use of terms like “False-Positive Rate” and “False-Negative Rate.” The numerators of these terms are clear – in 2 × 2 tables like the ones above, they correspond to cells b and c, respectively. The trouble is that the denominator is not used consistently. For example, the False-Negative Rate is generally defined as (1 – Sensitivity), i.e., the denominator is (a + c). But sometimes the term is used when the denominator is (c + d) or even (a + b + c + d).

Here’s an example of how this error can get us into trouble. We have often heard the following rationale for requiring a urine culture to rule out a urinary tract infection (UTI), even when the urinalysis (UA) is negative:

1. The sensitivity of the UA for a UTI is about 80%.
2. Therefore, the false-negative rate is 20%.
3. Therefore, after a negative UA, there is a 20% chance that it’s a false negative and that a UTI will be missed.
4. The 20% chance of missing a UTI is too high; therefore, always culture, even if the UA is negative.

Do you see what has happened here? The decision to culture should be based on the posterior probability of UTI after the UA (which is the prior probability before the culture). We do want to know the chance that a negative UA represents a false negative, so it seems like the false-negative rate should be relevant. But the false-negative rate we want is (1 – Negative Predictive Value), not (1 – Sensitivity). In the example above, in Statement 2, we began with a false-negative rate that was (1 – Sensitivity), and then in Statement 3, we switched to (1 – Negative Predictive Value). But we can’t know negative predictive value just from the sensitivity; it will depend on the prior probability of UTI (and the specificity of the test) as well.

This is illustrated below for two different prior probabilities of UTI in a 2-month-old boy. In the high-risk scenario, the baby is uncircumcised, has a high (39.3°C) fever, and a UTI risk of about 40%. In the low-risk scenario, the baby is circumcised, has a lower (38.3°C) fever, and a UTI risk of only ~2% (Newman et al. 2002). The sensitivity of the UA is assumed to be 80% and the specificity 85%.

	High-Risk Boy: Prior = 40%				Low Risk-Boy: Prior = 2%		
	UTI	No UTI	Total		UTI	No UTI	Total
UA+	320	90	410	UA+	16	147	163
UA–	80	510	590	UA–	4	833	837
Total	400	600	1000	Total	20	980	1000
Posterior probability after negative UA = 80/590 = 13.5%				Posterior probability after negative UA = 4/837 = 0.4%			

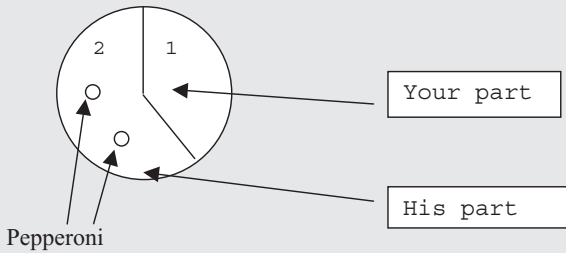
For the high-risk boy, the posterior probability after a negative UA is still 13.5%, perhaps justifying a urine culture. In the low-risk boy, however, the posterior probability is down to 0.4%, meaning that 250 urine cultures would need to be done on such infants for each 1 expected to be positive.

There are many similar examples of this problem, where Test A is not felt to be sufficiently sensitive to rule out the disease, so if it is negative, we are taught that Test B needs to be done. This only makes sense if Test A is never done when the prior probability is low.

Box 3.4: Understanding odds and probability using pizzas

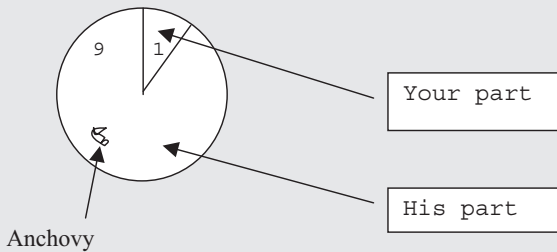
It might help to visualize a delicious but insufficient pizza to be completely divided between you and a hungry friend when you are on-call together. If your portion is half as big as his, it follows that your portion is one-third of the pizza. Expressing the ratio of the size of your portion to the size of his is like odds; expressing your portion as a fraction of the total is like probability. If you get confused about probability and odds, just draw a pizza!

Call night #1: Your portion is half as big as his. What percent of the pizza do you eat?



Answer: 1/3 of the pizza (if odds = 1:2, probability = 1/3).

Call night #2: You eat 10% of the pizza. What is the ratio of the size of your portion to the size of your friend's portion?



Answer: Ratio of the size of your portion to the size of her portion, 1:9 (if probability = 10%, odds = 1:9).

you can triple it. To get around this problem, we switch from probability to odds. Then we will be able to say:

$$\text{Prior Odds} \times \text{Likelihood Ratio} = \text{Posterior odds}$$

Necessary digression: a crash course in odds and probability

This topic trips up a lot of people, but it really is not that hard. “Odds” are just a probability (P) expressed as a ratio to $(1 - P)$; in other words, the probability that something *will* happen (or already exists) divided by the probability that it *won't* happen (or does not already exist). For our current purposes, we are mostly interested in the odds for diagnosing diseases, so we are interested in:

$$\frac{\text{Probability of having the disease}}{\text{Probability of *not* having the disease}}$$

If your only previous experience with odds comes from gambling, do not get confused – in gambling they use betting odds, which are based on the odds of *not*

Box 3.5: Practice with odds and probabilities

Convert the following probabilities to odds:

- a) 0.01
- b) 0.25
- c) $3/8$
- d) $7/11$
- e) 0.99

Convert the following odds to probabilities:

- a) 0.01
- b) 1:4
- c) 0.5
- d) 4:3
- e) 10

winning. That is, if the tote board shows a horse at 2:1, the odds of the horse winning are 1:2 (or a little less to allow a profit for the track).

We find it helpful always to express odds with a colon, like a:b. However, mathematically, odds are ratios, so 4:1 is the same as $4/1$ or 4, and 1:5 is $1/5$ or 0.2.

Here are the formulas for converting from probability to odds and vice versa:

If probability is P, the corresponding odds are $P/(1 - P)$.

- If the probability is 0.5, the odds are $0.5:0.5 = 1:1$.
- If the probability is 0.75, the odds are $0.75:0.25 = 3:1$.

If odds are a:b, the corresponding probability is $a/(a + b)$.

- If the odds are 1:9, the probability is $1/(1 + 9) = 1/10$.
- If the odds are 4:3, the probability is $4/(4 + 3) = 4/7$.

If the odds are already expressed as a single number (e.g., 0.5 or 2), then the formula simplifies to: Probability = Odds/(Odds + 1) because the “b” value of the a:b way of writing odds is implicitly equal to 1.

The only way to learn this is just to do it. Box 3.5 has some problems to practice on your own right now. (The answers are in Appendix 3.4.)

One thing you probably noticed in these examples (and could also infer from the formulas) is that, when probabilities are small, they are almost the same as odds. Another thing you notice is that *odds are always higher than probabilities* (except when both are zero). Knowing this may help you catch errors. Finally, probabilities cannot exceed one, whereas odds can range from zero to infinity.

The last thing you will need to know about odds is that, because they are just ratios, when you want to multiply odds by something, you multiply only the numerator (on the left side of the colon). So if you multiply odds of 3:1 by 2, you get 6:1. If you multiply odds of 1:8 by 0.4, you get odds of $(0.4 \times 1):8 = 0.4/8 = 0.05$ or 5:100.

Table 3.3. 2×2 table for likelihood ratio derivation

	Disease+	Disease–	Total
Test+	a True Positives	b False Positives	a + b Total Positives
Test–	c False Negatives	d True Negatives	c + d Total Negatives
Total	a + c Total With Disease	b + d Total Without Disease	a + b + c + d Total N

Deriving likelihood ratios (“lite” version)

Suppose we want to find something by which we can multiply the prior odds of disease in order to get the posterior odds. What would that something have to be?

Recall the basic 2×2 table and assume we study an entire population or use cross-sectional sampling, so that the prior probability of disease is $(a + c)/N$ (Table 3.3).

What, in terms of a , b , c , and d , are the prior odds of disease? The prior odds are just the probability of having disease divided by the probability of *not* having disease, based on knowledge we have before we do the test. So

$$\begin{aligned} \text{prior odds} &= \frac{P(\text{disease})}{P(\text{no disease})} = \frac{\text{Total with disease/Total N}}{\text{Total without disease/Total N}} \\ &= \frac{(a + c)/N}{(b + d)/N} = \frac{(a + c)}{(b + d)} \end{aligned}$$

Now, if the test is positive, what are the posterior odds of disease? We want to calculate the odds of disease as above, except now use information we have derived from the test. Because the test is positive, we can focus on just the upper (positive test) row of the 2×2 table. The probability of having disease is now the same as the posterior probability: True Positives/All Positives or $a/(a + b)$. The probability of not having disease if the test is positive is: False Positives/All Positives or $b/(a + b)$. So the posterior odds of disease if the test is positive are:

$$\frac{P(\text{Disease}|\text{Test+})}{P(\text{No Disease}|\text{Test+})} = \frac{\text{True Positive/Total Positive}}{\text{False Positive/Total Positive}} = \frac{a/(a + b)}{b/(a + b)} = \frac{a}{b}$$

So now the question is: by what could we multiply the prior odds $(a + c)/(b + d)$ in order to get the posterior odds (a/b) ?

$$\frac{a + c}{b + d} \times ? = \frac{a}{b}$$

The answer is:

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

So,

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

This must be the likelihood ratio (LR) we having been searching for!⁴

But look more closely at the formula for the LR that we just derived – some of it should look familiar. Remember what $a/(a + c)$ is? That’s right, sensitivity! And $b/(b + d)$ is $(1 - \text{Specificity})$. So the LR for a positive dichotomous test is just $\text{Sensitivity}/(1 - \text{Specificity})$.

You do not need to derive this every time you want to know what a LR is, although you could. Instead, just remember this one formula:

$$\text{Likelihood ratio (result)} = \frac{P(\text{result}|\text{disease})}{P(\text{result}|\text{no disease})}.$$

Stated in words, this says that the likelihood ratio for a test result is the probability of obtaining this test result in those *with* the disease divided by the probability of obtaining this result in those *without* the disease. This formula is a good one to memorize, because, as we will see in Chapter 4, it works for all tests, not just dichotomous ones. The numerator refers to patients *with* the disease, and the denominator refers to patients *without* the disease. One way to remember it is WOWO, which is short for “With Over WithOut.”⁵ Each possible test result has a LR. For dichotomous tests, there are two possible results and therefore two LRs: LR(+), the LR of a positive result, and LR(–), the LR of a negative result.

To derive the formula for the LR for a negative result, you might first find it helpful to go back to the 2×2 table and retrace the steps we took to get the LR for a positive result, but instead use the cell values for the negative test, which appear in the lower row of the 2×2 table. If you do this, at the end you should have derived for the “?” factor the formula $\frac{c/(a + c)}{d/(d + b)}$. If you think about what other ways we have to express this, you should come up with

$$\text{Likelihood ratio}(-) = \frac{P(-|\text{disease})}{P(-|\text{no disease})} = \frac{1 - \text{Sensitivity}}{\text{Specificity}}$$

or the likelihood of a negative result in patients with the disease divided by the likelihood of a negative result in patients without the disease.

⁴ In case you are wondering why we call this the “lite” derivation, it is because the formula for the LR works even when sensitivity and specificity come from a study that does not have cross-sectional sampling, but this derivation would not work in such a study.

⁵ Thanks to Dr. Warren Browner for this mnemonic.

Example 3.2 Using LRs to Calculate Posterior Probability

Let us return to Example 3.1, where the prevalence (prior probability) of breast cancer was 0.28%, the sensitivity of the mammogram was 75%, and the specificity was 93%. The LR for a positive mammogram would then be $[\text{Sensitivity}/(1 - \text{Specificity})] = 0.75/0.07 = 10.7$. Since odds and probabilities are almost the same when probabilities are low, let us first try a short cut: simply multiply the prior probability by the LR:

$$0.0028 \times 10.7 = 0.030 = 3\%$$

This is close to the 2.9% we calculated with the 2×2 table method used before. However, if the prior probability and/or the LR are higher, this shortcut will not work. For example, in a 65-year-old woman (prior probability $\approx 1.5\%$) with a mammogram “suspicious for malignancy” (LR ≈ 100), we would get $[0.015 \times 100] = 1.5$, which doesn’t make any sense as a posterior probability, because it is greater than one. In general, if the product of the prior probability and likelihood ratio is more than about 10%, we have to convert to odds and back again. For the example above, the steps are:

1. Convert prior probability (P) to prior odds $[P/(1 - P)] = 0.015/(1 - 0.015) = 0.0152$.
2. Find the LR for the patient’s test result (r): $\text{LR}(r) = \frac{P(r|D+)}{P(r|D-)} = 100$.
3. Multiply prior odds by the LR of the test result: $0.0152 \times 100 = 1.52$.
4. Convert posterior odds back to probability $\left(P = \frac{\text{odds}}{1 + \text{odds}}\right)$:

$$P = 1.52/(1 + 1.52) = 1.52/2.52 = 0.60.$$

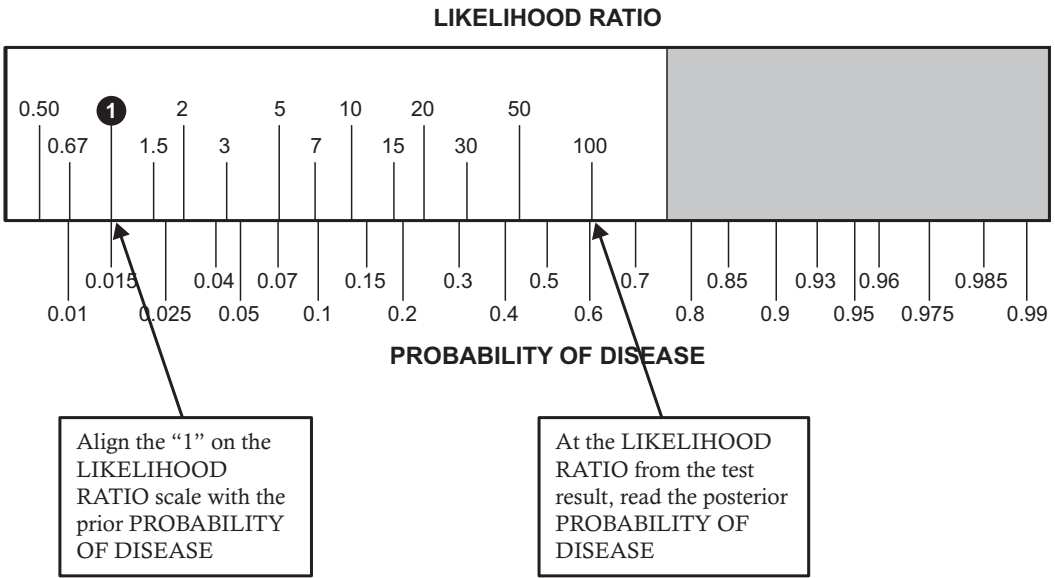
So if the *prior* probability of breast cancer was 1.5%, a mammogram “suspicious for malignancy” would raise the *posterior* probability to about 60%.

Using the LR slide rule

Although LRs make calculation of posterior probability a little easier than the 2×2 table method, it still is rather burdensome, especially if the probabilities are too high to skip the conversion from probability to odds and back. An alternative is to use a LR slide rule, which not only provides an answer, but is also useful for visualizing the process of going from prior to posterior probability. It uses a probability scale that is spread out so that distances on it are proportional to the logarithm of the prior odds. We review logarithms in Appendix 4.1. A cutout version of the LR slide rule is included with this book. To use the slide rule to calculate posterior probability from prior probability and LR:

1. Line up the 1 on the LR portion (sliding insert) with the prior probability on the probability (lower) portion.
2. Find the LR of the test result on the LR (top) half and read off the posterior probability just below.

We will see how the LR slide rule can help us understand testing thresholds.



This example shows the position if the prior probability is 0.015 and the likelihood ratio is 100. The posterior probability is about 0.6.

Figure 3.1 LR slide rule

Treatment and testing thresholds

Recall that in Chapter 1 we said that a good reason to do a diagnostic test is to help you make a decision about administering or withholding treatment. There are two main factors that limit the usefulness of tests:

1. They sometimes give wrong answers.
2. They have a “cost,” which includes the financial cost as well as the risks, discomfort, and complications that arise from testing.

Even a costless test has limited usefulness if it is not very accurate, and even a 100% accurate test has limited usefulness if it is very costly. In the following sections, we will show how test inaccuracy and costs narrow the range of prior probabilities for which the expected benefits justify performing the test. Readers interested in a more in-depth discussion should read about decision analysis (Sox 1986; Sox et al. 1988; Detsky et al. 1997a, 1997b; Krahn et al. 1997; Naglie et al. 1997; Naimark et al. 1997).

As an example, we will consider the question of whether to use a rapid bedside test, such as the QuickVue test discussed earlier in this chapter, to guide antiviral treatment for the flu. An antiviral medication, such as oseltamivir, reduces the duration of flu symptoms by about 1 day.

Quantifying costs and benefits

In order to calculate the range of prior probabilities for which the expected benefits justify testing, we need to quantify three things:

1. *How bad is it to treat someone who does not have the disease?* This quantity is generally denoted “C” (for cost) (Sox et al. 1988; Hilden and Glasziou 1996). C

is the cost⁶ of (unnecessarily) treating someone without the disease. In the flu example, we will take the cost of this unnecessary treatment as just the monetary cost of the antiviral medication, about \$60.

2. *How bad is it to fail to treat someone who has the disease?* This quantity is generally denoted “B” (Sox et al. 1988; Hilden and Glasziou 1996). You can think of B as the cost of failing to achieve the **B**enefit of treatment. For example, if the value we assign to patients with the flu feeling better 1 day sooner is \$160, but the medication costs \$60, the net benefit of treatment is $160 - 60 = 100$, so we can think of that missed opportunity to get the \$100 benefit of treatment as the net cost of not treating someone with the flu.
3. *What is the cost of the test?* This cost includes the cost of the time, reagents, etc. to do the test, as well as the cost of complications or discomfort from doing the test itself (including assigning a dollar value to any pain and suffering involved). We will denote this test cost as “T.”

A note about the term “cost”: Some of our colleagues have objected to using the term “cost,” because readers might construe it to refer only to monetary costs. Our various “costs” include all harm, pain, suffering, time, and money associated with 1) treating someone unnecessarily, 2) failing to treat someone who needs treatment, and 3) performing the diagnostic test.

The treatment threshold

First introduced by Pauker and Kassirer in 1975 (Pauker and Kassirer 1975), the “treatment threshold probability,” P_{TT} , is the (posterior) probability of disease at which the expected costs of the two types of mistakes we can make (treating people without the disease and not treating people with the disease) are balanced. By expected costs, we mean we multiply the cost of these mistakes (C and B) by their probability of occurring. For example, the expected cost of not treating is P (the probability of disease) $\times B$. This is because the probability that not treating is the wrong decision is the probability that the person has the disease, or P , and the cost of that wrong decision is B . This makes sense: if $P = 0$, then not treating will not be a mistake, and the cost will be zero. On the other hand, if $P = 1$, the person has the disease, and the cost of not treating is $1 \times B = B$. If $P = 0.5$, then half the time the cost will be zero, and half the time the cost will be B , so the expected cost is $0.5 \times B$. We can graph this expected cost of not treating as a function of the probability of disease: $P \times B$ is the equation for a straight line with slope B and intercept 0, as shown in Figure 3.2

Similarly, the expected cost of treating is $(1 - P) \times C$. The probability that treating is the wrong decision is the probability that the person does not have the disease ($1 - P$), and the cost of treating someone who does not have the disease is C . Because $(1 - P) \times C = C - C \times P$, the expected cost of treating is a straight line,

⁶ What we refer to as “cost” might more strictly be termed “regret,” the difference in outcome between the action we took and the best action we could, in retrospect, have taken. (See Hilden and Glasziou 1996.) The regret associated with treating a patient who turns out to have the disease is zero, since it was the best action we could have taken. Similarly, the regret associated with not treating an individual who turns out not to have the disease is also zero.

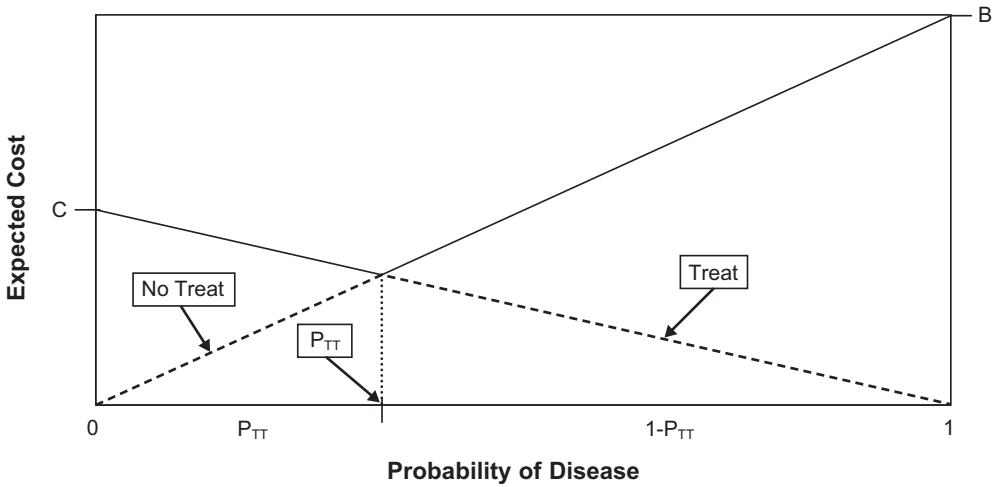


Figure 3.2 Expected costs of not treating and treating by probability of disease. For probabilities from 0 to P_{TT} , “No Treat” has the lowest expected cost. For probabilities from P_{TT} to 1, “Treat” has the lowest expected cost.

with intercept C and slope $-C$. The place where these two lines cross is the treatment threshold, the probability of disease, P_{TT} , at which the expected costs of not treating and treating are equal (Fig. 3.2). Put mathematically, P_{TT} is the probability of disease at which:

$$P_{TT} \times B = (1 - P_{TT}) \times C$$

And therefore, the treatment threshold odds are given by:

$$\frac{P_{TT}}{(1 - P_{TT})} = \frac{C}{B}$$

and the threshold probability is

$$P_{TT} = \frac{C}{(C + B)}$$

Stop here to convince yourself that this formula makes sense. If treating someone who does not have the disease is half as bad as failing to treat someone who does have the disease, we should be willing to treat two people without disease to avoid failing to treat one person who has it, and the threshold probability P_{TT} should be $1/3$. Using the formula above, if $B = 2 \times C$, then we get $P_{TT} = C/(C + 2C) = C/3C = 1/3$. Similarly, if the two types of mistakes are equally bad, $C = B$, and P_{TT} should be 0.5 .

Finally, look at the graph in Figure 3.2 and visualize what happens as C gets closer to zero. Can you see how the treatment threshold, P_{TT} , slides down the “no treat” line, approaching zero? This makes sense: if the cost of treating people without disease is low relative to the benefit of treating someone who has it, you will want to treat even when the probability of disease is low. Similarly, imagine what happens when C goes up in relation to B . The treatment threshold, P_{TT} , will move to the right.

As did Pauker and Kassirer in 1980 (Pauker and Kassirer 1980), we now extend the threshold calculation to the case where a dichotomous diagnostic test is available.

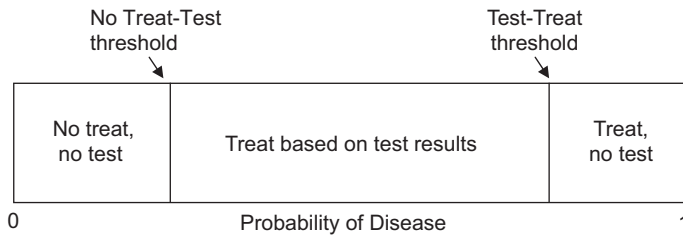


Figure 3.3 The no treat–test and test–treat probability thresholds, between which the test can affect treatment decisions.

There are now two threshold probabilities: the “no treat–test threshold” and the “test–treat threshold.”

Testing thresholds for an imperfect but costless test

We will first assume that the test itself has absolutely no monetary cost or risks to the patient. Even if a test is very inexpensive or free, if it isn’t perfect, there are some situations in which testing is not indicated because it cannot change the treatment decision. If a dichotomous test has less than perfect specificity (i.e., false positives are possible) and the treatment has some risks (i.e., $C > 0$), there will be some low prior probability below which you would not want to treat even if the test were positive. This is because the low prior probability keeps the posterior probability low, so that the false positives would overwhelm the true positives and there would be too many people treated unnecessarily. That defines a lower testing threshold, the No Treat–Test threshold, below which there is no point performing the test. For a dichotomous test, this lower threshold is related to the LR for a positive result.

At the other end of the spectrum, if the test has less than perfect sensitivity (i.e., false negatives are possible) and the treatment has some benefits (i.e., $B > 0$), there will be some high prior probability above which you would want to treat even if the test were negative. This is because the high prior probability keeps the posterior probability high, so that false negatives would overwhelm the true negatives and testing would lead to too many failures to treat patients with the disease. That defines a higher testing threshold, the Test–Treat threshold, above which one should just treat, rather than do the test. This higher threshold is related to the LR of a negative result for a dichotomous test.

Between these two testing thresholds, there is a zone in which the results of the test have the potential to affect your decision to treat (Fig. 3.3).

Example 3.3 In patients with the flu, we quantified the net benefit of antiviral treatment at \$100 and the cost of unnecessary treatment at \$60. Then, our treatment threshold should be $C/(C + B) = 60/160 = 37.5\%$. That is, after we do our rapid bedside antigen test, if the probability of influenza is greater than 37.5%, we will treat the patient. We will assume that the sensitivity of the rapid antigen test is 75% and specificity is 95%. (These are close to, but slightly worse than, the estimates from Table 3.1.) What are our testing thresholds in this case – that is, for what range

of prior probabilities of influenza should the results of the bedside test affect the decision to treat? (For now, we are assuming that the test is free and harmless to the patient.) Here are the steps to follow:

1. Calculate LRs for positive and negative test results:

$$\text{LR}(+) = \text{Sensitivity}/(1 - \text{Specificity}) = 0.75/(1 - 0.95) = 0.75/0.05 = 15$$

$$\text{LR}(-) = (1 - \text{Sensitivity})/\text{Specificity} = (1 - 0.75)/0.95 = 0.25/0.95 = 0.26$$

2. Convert the treatment threshold of 0.375 to odds:

$$\text{odds} = P/(1 - P) = 0.375/(1 - .375) = 0.6$$

3. Divide LR(+) and LR(-) into treatment threshold to get the prior odds for the testing thresholds:

(since posterior odds = prior odds \times LR, then posterior odds/LR = prior odds)

$$\text{Posterior odds}/\text{LR}(+) = (0.6/15) = 0.04(\text{for positive test})$$

$$\text{Posterior odds}/\text{LR}(-) = 2.28(\text{for negative test})$$

4. Convert each of these prior odds (for testing thresholds) back to a prior probability
 $P = \text{odds}/(1 + \text{odds})$:

$$P = 0.04/1.04 = 0.04 \text{ (for positive test)}$$

$$P = 2.28/3.28 = 0.70 \text{ (for negative test)}$$

5. Interpret the result:

- If the prior probability of influenza is $<4\%$ (the no treat–test threshold), then even if the rapid antigen test is positive, the post-test probability will still be below 37.5% (the treatment threshold), and you would not treat the patient.
- If the prior probability is $>70\%$ (the test–treat threshold), then even if the antigen test is negative, the post-test probability will be above 37.5%, and you would treat the patient in spite of the negative test result.
- If the prior probability is between 4% and 70%, the test *may* be indicated, because it at least has the potential to affect management.

So far, we have not considered costs or risks of the test (as opposed to those of the treatment). When these are factored in as well, the testing range will be narrower.

Visualizing testing thresholds

The LR slide rule's log(odds) scale provides a nice way of visualizing testing thresholds when the accuracy of a test (rather than its costs or risks) is the main thing that limits its usefulness. In the flu example (Example 3.3), the positive and negative LRs of the bedside antigen test can be visualized as arrows. If they are placed with their points on the treatment threshold, their origins will define the testing thresholds as in Figure 3.4.

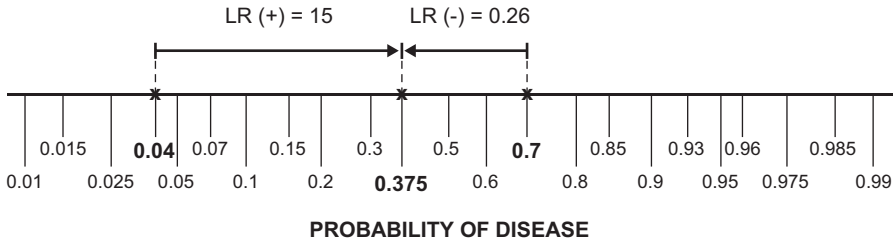


Figure 3.4 LR slide rule arrows demonstrate the concept of test and treatment thresholds.

If the prior probability of influenza is less than about 0.04, even if the test is positive, the posterior probability will remain below 0.375, and we shouldn't treat. Similarly, if the prior probability is more than 0.7, even if the test is negative, the posterior probability will remain high enough to treat. These are the same numbers we got algebraically in Example 3.3.

You can also visualize the testing threshold using an X-shaped graph like Figure 3.2. In this case, we draw a line for the expected cost of testing and treating according to the result of the test. When the probability of disease is zero, the expected cost is $C \times (1 - \text{Specificity})$. This is the cost of unnecessary treatment (C) times the probability that the test will be falsely positive in patients without the disease. Similarly, when the probability of disease is 1, the expected cost is $B \times (1 - \text{Sensitivity})$. This is the cost (B) of failing to treat times the probability that the test will be falsely negative. If we connect these two points with a straight line, we can see that, at very low and very high probabilities of disease, "No Treat" and "Treat" have lower expected costs than "Test," because testing too often leads to wrong answers (Fig. 3.5).

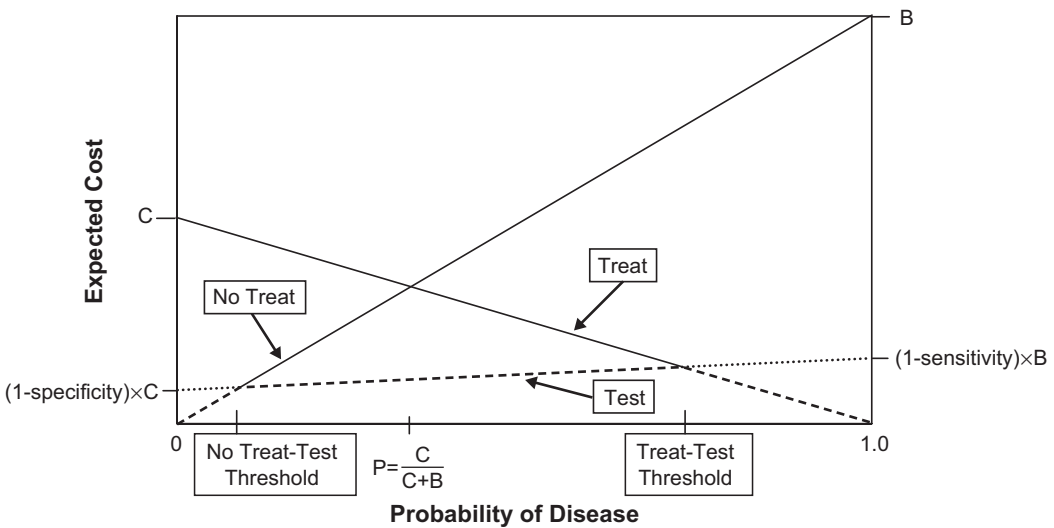


Figure 3.5 The expected cost of the "Test" option is higher than the cost of "No Treat" below the No Treat-Test threshold, and higher then the cost of "Treat" above the Test-Treat threshold.

Testing thresholds for a perfect but risky or expensive test

In the preceding discussion, we showed that, when tests are imperfect, there are some prior probabilities for which the test is not worth doing because the results do not have the potential to affect management. But some tests, with close to 100% sensitivity or specificity, *do* have the potential to change management, even when the prior probability of disease is very close to zero or one. However, because there are risks and costs to tests themselves, even a nearly perfect test may not be worth doing in some patients. Even though it has the potential to change management in some clinical situations, the probability of it doing so is too small to justify the cost of the test. To explore this issue, we now assume that the test is perfect (Sensitivity = Specificity = 100%), but that it has some “cost.” Keep in mind that “cost” could represent monetary cost, which is easy to quantify, or risks to the patient (such as pain and loss of privacy), which are harder to quantify. In this situation, there are still two threshold probabilities: 1) the No Treat–Test threshold, where the expected benefits of identifying and treating D+ individuals first justify the testing costs; and 2) the Test–Treat threshold, where the expected savings from identifying and not treating D– individuals no longer justify the testing costs.

If the bedside test for influenza were perfect and the prior probability of influenza were 5%, we would have to test twenty patients to identify one case of the flu. If the prior probability were 10%, we would have to test ten patients to identify one case. For a perfectly sensitive test, the number needed to test to identify one D+ individual is simply $1/P(D+)$, where $P(D+)$ is the prior probability of disease.

To find the No Treat–Test threshold probability, we need to ask how many individuals we are willing to test to identify one D+ individual.

We have already utilized B, the cost of not treating a D+ individual, which we can also think of as the net benefit of treating someone with the disease, and C, the cost of unnecessarily treating a D– individual; now we utilize T, the cost of the test. For a perfect test, the No Treat–Test threshold probability is T/B . (This means we have to assign a dollar benefit to treating people with disease so that T and B can be measured in the same units.)

Assume that the perfect bedside flu testing kits cost \$10 each ($T = \10). If $B = \$100$ after subtracting the cost of the drug, then $T/B = \$10/\$100 = 10\%$. This makes sense: for every ten patients we test, on average one will have the flu and be treated, which is worth \$100, but the cost of testing those ten people is also $10 \times \$10 = \100 . The costs of testing and benefits of treating are equal, and we break even. If the prior probability of flu is less than 10%, on average we will have to test more than ten people (and hence spend more than \$100) for each one who tests positive and gets the treatment; hence, on the average costs of testing would exceed the benefits.

To understand the test–treat threshold probability, we reverse the logic. We again assume that C, the cost of treating someone without the flu, is just the \$60 cost of the medication. Start by assuming that the probability of influenza is 100%. There is no point in testing to identify D– individuals because there aren’t any, so we would just treat without testing. As the probability of flu decreases from 100%, it eventually reaches a point where the \$60 treatment cost we save by identifying a D– individual

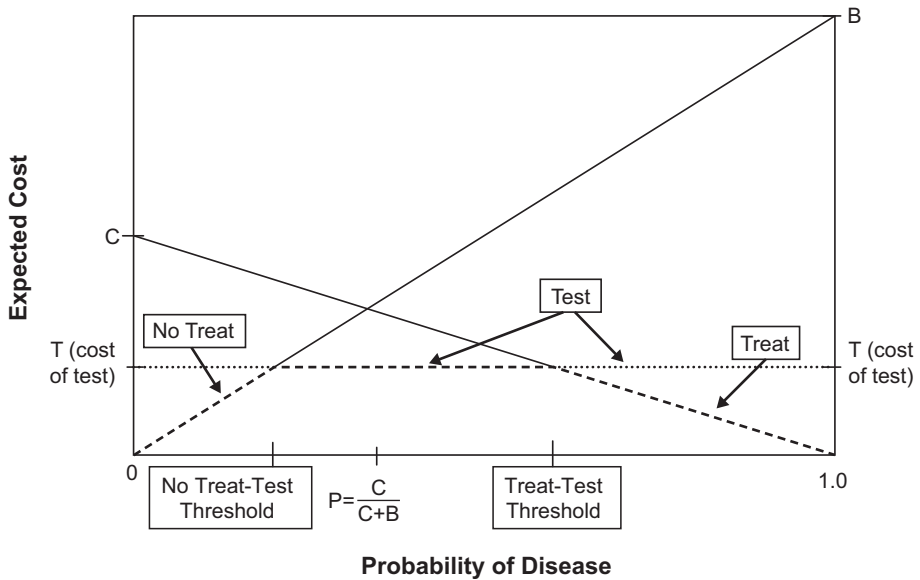


Figure 3.6 “No treat–test” and “test–treat” thresholds for a perfect but costly test.

justifies the cost of testing to identify that individual. This occurs when the probability of not having the disease is T/C , corresponding to a probability of having the disease of $(1 - T/C)$, the Test–Treat threshold. This makes sense, too. When the probability of nondisease is $1/6$, the number needed to test to identify one patient without the disease is six. We test six patients at a testing cost of \$10 each in order to save \$60 on the one without disease, and hence we come out even. We have to convert this $1/6$ probability of nondisease to a probability of disease by subtracting from 100%, so the test–treat threshold probability of disease is $1 - 1/6 = 5/6 = 83.3\%$.

You can easily visualize testing thresholds for a perfect but costly test by drawing a horizontal line at expected cost = T for the testing option (Fig. 3.6).

Testing thresholds for an imperfect and costly test

Using the same parameters, $C = \$60$, $B = \$100$, $T = \$10$ (or \$0), Sensitivity = 0.75 (or 1.0), and Specificity = 0.95 (or 1.0), Table 3.4 gives the testing thresholds assuming the test is 1) imperfect and costless, 2) perfect and costly, and 3) imperfect and costly. For interested readers, the formulas for the testing thresholds of an imperfect and costly test are given in Appendix 3.2. The graph showing expected costs would be the same as Figure 3.5, except that the testing line would be displaced upward by an amount equal to the testing cost (T).

In order to do these calculations, we have to express misclassification costs (B and C) and testing costs in common units. It is usually difficult to reduce the costs and risks of testing, as well as failing to treat someone with disease or treating someone without the disease, to units such as dollars. We present the algebra and formulas here, not because we want you to use them clinically, but because we want you to understand them and want to show that the theory here is actually quite simple.

Table 3.4. Thresholds for a flu test, taking into account accuracy, cost, and both

Test characteristics	No treat–test threshold	Test–treat threshold
Imperfect ^a but costless	0.04	0.70
Perfect but costly ^b	0.10	0.67
Imperfect ^a and costly	0.17	0.57

^a Sensitivity = 0.75; Specificity = 0.95.

^b T = \$10.

Testing thresholds exist both because the test is imperfect (and might lead to too many misclassifications) and because the test has costs and risks that might outweigh the benefits of the additional information. Sometimes, especially when the test is expensive and risky but accurate, the testing costs so outweigh the misclassification risks that you can ignore the misclassification risks. Would you do the test if it were perfect? If the answer is “no,” then the risks and costs of the test, not the misclassification risks, are driving your decision. We don’t do lumbar punctures on well-looking febrile infants. This is not so much because we are worried about false positives, but because the low probability of a positive does not justify the discomfort, risk, and expense of the test. Would you do the test if it were free of discomfort, risks, and costs? If the answer is “no,” then the misclassification risks, not the costs and risks of the test itself, are driving your decision. This is one reason we don’t perform screening mammography on 30-year-old women. The false positives would vastly overwhelm the true positives and cause an enormous burden of stress and ultimately unnecessary follow-up testing.

Summary of key points

1. The accuracy of dichotomous tests can be summarized by the proportion in whom the test gives the right answer in five groups of patients:
 - those with disease (sensitivity),
 - those without the disease (specificity),
 - those who test positive (positive predictive value),
 - those who test negative (negative predictive value), and
 - the entire population tested (accuracy).
2. Although sensitivity and specificity are more useful for evaluating tests, clinicians evaluating patients will more often want to know the posterior probability of disease given a particular test result.
3. Posterior probability can be calculated by using the sensitivity and specificity of the test and the prior probability of disease. This can be done by using the 2×2 table method or by converting probabilities to odds and using the LR of the test result (defined as $P(\text{Result}|\text{Disease})/P(\text{Result}|\text{No disease})$).
4. The treatment threshold (P_{TT}) is the posterior probability of disease at which the expected cost of treating those without disease equals the expected cost of not treating those with the disease. The formula for this is: $P_{TT} = C/(C + B)$.

5. If a test is less than perfectly accurate or has costs or risks, it does not make sense to use it on patients with very low probabilities of disease – probabilities below the “No Treat–Test” threshold.
6. Similarly, sometimes the probability of disease is so high that it makes sense to skip the test and proceed directly to treatment. This occurs when the probability is greater than the “Test–Treat” threshold.
7. Both the “No Treat–Test” and “Test–Treat” thresholds can be visualized graphically or calculated algebraically if the cost of treating someone without the disease (C), the cost of failing to treat someone with the disease (B), and the cost of the test (T) can all be estimated on the same scale.

Appendix 3.1: General summary of definitions and formulas for dichotomous tests

	Disease	No Disease	Totals
Test+	a	b	a + b
Test–	c	d	c + d
Totals	a + c	b + d	N = (a + b + c + d)

$$\text{Sensitivity} = a/(a + c)$$

$$= P(+|D+)$$

$$\text{Specificity} = d/(b + d)$$

$$= P(-|D-)$$

$$1 - \text{Sensitivity} = P(-|D+)$$

$$1 - \text{Specificity} = P(+|D-)$$

If sampling is cross-sectional (i.e., diseased and nondiseased are not sampled separately), then:

$$\text{Prevalence} = \text{prior probability} = (a + c)/N$$

$$\text{Positive Predictive Value (PPV)} = \text{Posterior probability if test +} = a/(a + b)$$

For tests with dichotomous results:

$$\text{LR}(+) = P(+|D+)/P(+|D-) = \text{sensitivity}/(1 - \text{specificity})$$

$$\text{LR}(-) = P(-|D+)/P(-|D-) = (1 - \text{sensitivity})/\text{specificity}$$

$$\text{Probability} = P = \text{odds}/(1 + \text{odds});$$

$$\text{Odds} = P/(1 - P) \text{ or}$$

$$\text{If odds} = a/b, \text{ probability} = a/(a + b)$$

$$\text{Prior odds} \times \text{LR} = \text{posterior odds} \quad (\text{ALWAYS TRUE!})$$

Appendix 3.2: Rigorous derivation of likelihood ratios

Here is a real derivation – it is not that hard!

First, you need to accept some basic theorems of probability:

1. $P(A \text{ and } B) = P(B \text{ and } A)$

2. $P(A \text{ and } B) = P(A|B)P(B)$. This just says the probability of both A and B is the probability of B times the probability of A *given* B.

From 1 and 2 (which both seem self-evident), it is easy to prove Bayes's theorem:

3. $P(A|B)P(B) = P(A \text{ and } B) = P(B \text{ and } A) = P(B|A)P(A)$. Therefore, $P(A|B) = P(B|A)P(A)/P(B)$, which is how Bayes's theorem is generally written.

Now by Bayes's theorem (where $r =$ a specific test result):

- 4. Posterior probability = $P(D+|r) = P(r|D+)P(D+)/P(r)$
- 5. $1 -$ Posterior probability = $P(D-|r) = P(r|D-)P(D-)/P(r)$

Dividing 4 by 5 gives:

$$6. \frac{P(D+|r)}{P(D-|r)} = \frac{P(r|D+)}{P(r|D-)} \times \frac{P(D+)}{P(D-)}$$

Posterior odds = $LR(r) \times$ Prior odds

Note that this derivation applies regardless of the form the result takes (dichotomous, continuous, etc.) and requires no assumptions other than the probability theorems we started with.

Appendix 3.3: Formulas for testing thresholds for dichotomous tests

B = Net Benefit of Treating a D+ individual

C = Cost of Unnecessarily Treating a D- individual

C/B = Treatment Threshold Odds

T = Cost of Test

3.3a: For an imperfect but costless test:

$$\begin{aligned} \text{No Treat-Test Threshold Odds} &= \frac{C/B}{LR(+)} \\ &= \frac{(C)P(+|D-)}{(B)P(+|D+)} \end{aligned}$$

$$\text{No Treat-Test Threshold Prob} = \frac{(C)P(+|D-)}{(B)P(+|D+) + (C)P(+|D-)}$$

$$\begin{aligned} \text{Test-Treat Threshold Odds} &= \frac{C/B}{LR(-)} \\ &= \frac{(C)P(-|D-)}{(B)P(+|D-)} \end{aligned}$$

$$\text{Test-Treat Threshold Prob} = \frac{(C)P(-|D-)}{(B)P(-|D+) + (C)P(-|D-)}$$

Example: Imperfect but costless test for influenza

B = Net Benefit of Antiviral Treatment = \$100

C = Net Cost of Antiviral Treatment = \$60

Sensitivity = $P(+|D+) = 0.75$; $1 -$ Sensitivity = $P(-|D+) = 0.25$

Specificity = $P(-|D-) = 0.95$; $1 -$ Specificity = $P(+|D-) = 0.05$

$$\begin{aligned} \text{No Treat-Test Threshold Prob} &= \frac{(C)P(+|D-)}{(B)P(+|D+) + (C)P(+|D-)} \\ &= \frac{(60)0.05}{(100)0.75 + (60)0.05} \\ &= 0.04 \end{aligned}$$

$$\begin{aligned} \text{Test-Treat Threshold Prob} &= \frac{(C)P(-|D-)}{(B)P(-|D+) + (C)P(-|D-)} \\ &= \frac{(60)0.95}{(100)0.25 + (60)0.95} \\ &= 0.70 \end{aligned}$$

3.3b: For a perfect but costly test:

No Treat-Test Threshold Probability = T/B

Test-Treat Threshold Probability = 1 - T/C

Example: Perfect but costly test for influenza

B = Net Benefit of Antiviral Treatment = \$100

C = Antiviral Treatment Cost = \$60

T = Cost of the Perfect Bedside Test = \$10

No Treat-Test Threshold Probability = T/B = \$10/\$100 = 0.10

Test-Treat Threshold Probability = 1 - T/C = 100% - \$10/\$60 = 0.833

3.3c: For an imperfect and costly test:

$$\text{No Treat-Test Threshold Odds} = \frac{(C)P(+|D-) + T}{(B)P(+|D+) - T}$$

$$\text{No Treat-Test Threshold Prob} = \frac{(C)P(+|D-) + T}{(B)P(+|D+) + (C)P(+|D-)}$$

$$\text{Test-Treat Threshold Odds} = \frac{(C)P(-|D-) - T}{(B)P(-|D+) + T}$$

$$\text{Test-Treat Threshold Prob} = \frac{(C)P(-|D-) - T}{(B)P(-|D+) + (C)P(-|D-)}$$

Example: Imperfect and costly test for influenza

B = Net Benefit of Antiviral Treatment = \$100

C = Antiviral Treatment Cost = \$60

T = Cost of Test = \$10

Sensitivity = P(+|D+) = 0.75; 1 - Sensitivity = P(-|D+) = 0.25

Specificity = P(-|D-) = 0.95; 1 - Specificity = P(+|D-) = 0.05

$$\begin{aligned} \text{No Treat-Test Threshold Prob} &= \frac{(C)P(+|D-) + T}{(B)P(+|D+) + (C)P(+|D-)} \\ &= \frac{(60)0.05 + 10}{(100)0.75 + (60)0.05} \\ &= 0.167 \end{aligned}$$

$$\begin{aligned} \text{Test-Treat Threshold Prob} &= \frac{(C)P(-|D-) - T}{(B)P(-|D+) + (C)P(-|D-)} \\ &= \frac{(60)0.95 - 10}{(100)0.25 + (60)0.95} \\ &= 0.573 \end{aligned}$$

Appendix 3.4: Answers to odds/probability conversions in Box 3.5

If probability is P, Odds are $P/(1 - P)$

	Probability	Odds
a.	.01	1/99
b.	.25	1/3
c.	3/8	3/5
d.	7/11	7/4
e.	.99	99

If odds are a/b, probability is $a/(a + b)$.

	Odds	Probability
a.	.01	1/101
b.	1:4	1/5
c.	.5	$.5/1.5 = 1/3$
d.	4:3	4/7
e.	10	10/11

References

- Detsky, A. S., G. Naglie, et al. (1997a). "Primer on medical decision analysis: Part 1. Getting started." *Med Decis Making* 17(2): 123–5.
- Detsky, A. S., G. Naglie, et al. (1997b). "Primer on medical decision analysis: Part 2. Building a tree." *Med Decis Making* 17(2): 126–35.
- Hilden, J., and P. Glasziou (1996). "Regret graphs, diagnostic uncertainty and Youden's Index." *Stat Med* 15(10): 969–86.
- Kerlikowske, K., D. Grady, et al. (1996a). "Effect of age, breast density, and family history on the sensitivity of first screening mammography." *JAMA* 276(1): 33–8.
- Kerlikowske, K., D. Grady, et al. (1996b). "Likelihood ratios for modern screening mammography. Risk of breast cancer based on age and mammographic interpretation." *JAMA* 276(1): 39–43.
- Krahn, M. D., G. Naglie, et al. (1997). "Primer on medical decision analysis: Part 4. Analyzing the model and interpreting the results." *Med Decis Making* 17(2): 142–51.
- Nagle, G., M. D. Krahn, et al. (1997). "Primer on medical decision analysis: Part 3. Estimating probabilities and utilities." *Med Decis Making* 17(2): 136–41.
- Naimark, D., M. D. Krahn, et al. (1997). "Primer on medical decision analysis: Part 5. Working with Markov processes." *Med Decis Making* 17(2): 152–9.
- Newman, T. B., J. A. Bernzweig, et al. (2002). "Urine testing and urinary tract infections in febrile infants seen in office settings: the Pediatric Research in Office Settings' Febrile Infant Study." *Arch Pediatr Adolesc Med* 156(1): 44–54.
- Pauker, S. G., and J. P. Kassirer (1975). "Therapeutic decision making: A cost-benefit analysis." *N Engl J Med* 293(5): 229–34.

- Pauker, S. G., and J. P. Kassirer (1980). "The threshold approach to clinical decision making." *N Engl J Med* **302**(20): 1109–17.
- Poehling, K. A., M. R. Griffin, et al. (2002). "Bedside diagnosis of influenzavirus infections in hospitalized children." *Pediatrics* **110**(1 Pt 1): 83–8.
- Sox, H. C., Jr. (1986). "Probability theory in the use of diagnostic tests. An introduction to critical study of the literature." *Ann Intern Med* **104**(1): 60–6.
- Sox, H. C., M. A. Blatt, et al. (1988). *Medical Decision Making*. Boston, Butterworths.

Chapter 3 Problems: dichotomous tests

1. You are informed by your doctor that you have tested positive for Grunder-schnauzer disease. You may ask one question to help you figure out whether you really have it. What do you want to know (choices are sensitivity, specificity, prevalence, predictive value, etc.)?
2. Consider the following excerpt from an abstract about a test for Cat Scratch Disease (Zangwill et al. 1993):

ABSTRACT

METHODS. We conducted a physician survey to identify cases of cat scratch disease occurring over a 13-month period in cat owners in Connecticut. We interviewed both the patients (or their parents) and controls matched for age who owned cats. Serum from the patients was tested for antibodies to *Rochalimaea henselae* with a new, indirect fluorescent-antibody test.

RESULTS. Of 45 patients [cases], 38 had serum samples with titers of 1:64 or higher for antibody to *R. henselae*, as compared with 4 of 112 samples from controls ($P < 0.001$). *The positive predictive value of the serologic test was 91 percent [italics added] . . .*

- a) Make a 2×2 table that summarizes the results.
 - b) Is the authors' calculation of predictive value (91%) correct?
3. A "rapid strep" test for Group A streptococcal throat infection that used an optical immunoassay (OIA) was reported to have about 91% sensitivity and 95% specificity compared with culture (Roddey et al. 1995). The authors concluded that because "approximately 9% of cultures positive for group A Strep . . . would have been missed by the OIA, we believe that a throat culture should be processed in the case of a negative OIA result." Do you agree? How could you improve on this conclusion?
 4. Haydel et al. (2000) studied the usefulness of various clinical findings to predict a positive head CT scan in patients with minor head trauma and a possible loss of consciousness. The head CT scan was not the diagnostic test being evaluated; a positive head scan was the "disease," and the clinical findings were the "tests." They reported their results in the table below.

Association between selected clinical findings and CT results in 520 patients with minor head injury

Finding ^a	Total (N = 520)	Positive	Negative	P value ^b	Likelihood ratio ^c
		CT scan no. (%)	CT scan (N = 484)		
Short-term memory deficits	9 (2)	5 (14)	4 (1)	<0.001	15.0
Drug or alcohol intoxication	180 (35)	22 (61)	158 (33)	0.001	11.0
Headache	123 (24)	12 (33)	111 (23)	0.16	2.0

^a Some patients had more than one finding.

^b P values were determined by χ^2 analysis.

^c The LR indicates the likelihood of a positive CT scan in patients with the finding in question as compared with the likelihood in patients without the finding.

- a) Consider the second row of the table, which includes data relating drug or alcohol intoxication to the CT scan results. Calculate the LR for this finding.
 - b) The authors definition of LR in footnote **c** is actually a prevalence ratio (like a risk ratio): the probability of disease in those with the finding divided by the probability in those without the finding. For that same row of the table, calculate the prevalence ratio.
 - c) You have a patient with minor head trauma similar to those in the Haydel et al. study, whose prior probability of a positive CT scan you estimate at about 10% before you find out about drug or alcohol intoxication. If your history reveals drug or alcohol intoxication, what is your estimate of the probability that his CT scan will be positive?
5. A study (Gaitan-Cepeda et al. 2005) of people with HIV/AIDS undergoing highly active antiretroviral therapy (HAART) found that the probability of immune failure in the presence of oral candidiasis was 91%.
- a) View the exam for oral candidiasis as a test for immune failure. Which of the test characteristics defined in this chapter does the 91% figure represent?
 - b) Here are the results of the study's cross-sectional sample of patients with HIV/AIDS on HAART:

		Immune Failure		Total
		Yes	No	
Oral Candidiasis	Yes	31	3	34
	No	75	37	112
Total		106	40	146

- i. What was the prevalence of immune failure in this sample?
- ii. What was the probability of immune failure in the absence of oral candidiasis?
- iii. What would you call this number?
- iv. Calculate the sensitivity of oral candidiasis for immune failure.
- v. Calculate the specificity.

- c) Do you agree with the authors that oral candidiasis is a good marker for immune failure in patients on HAART?
6. You have a patient with pharyngitis (a sore throat) who you consider treating with penicillin.

Assume:

- i. The drug cost of a course of penicillin to treat acute Group A streptococcal throat infection (“strep throat”) is \$23 (www.drugstore.com, Penicillin VK 500 mg #30, 9/23/08), and the expected cost in patient inconvenience, risk of adverse or allergic reactions, and contribution to antibiotic resistance is another \$17. So, the total expected treatment cost is \$40.
- ii. Treating someone who really has strep throat (and not some other pharyngitis) decreases symptom severity, length of illness, transmission to others and the (already minute) risk of rheumatic fever. The value of this averages about \$100, but since the cost of treatment is \$40, the net benefit of treating someone with strep throat is \$60. This can also be viewed as the net cost of failing to treat someone with strep throat. Penicillin will not help the patient if the sore throat is caused by something other than Group A strep.
 - a) Draw a graph like figure 3.2, labeling the axes, lines and intercepts.
 - b) At what probability of strep throat should you treat with penicillin? Show the point on the graph.
 - c) If a rapid strep test were 90% sensitive and 91% specific, for what range of prior probabilities would it have the potential to affect management? (Ignore the cost of the test.) Do this calculation using likelihood ratios, then draw a line for “testing” on the graph.
 - d) Now assume that the perfect “rapid strep” test for Group A streptococcal throat infection has been developed. The test causes negligible discomfort and results are available nearly instantaneously, but the test costs \$30. When does it make sense to use this test? Draw a line for testing on the graph and explain.
 - e) Extra credit: Imagine you have a patient with a particularly severe sore throat. For her, you estimate the net benefit of treatment to be \$160 rather than \$60. Redraw the graph from part d and explain how your answer would change, if at all. (The algebra is optional – focus on the concept.)

References for problem set

- Gaitan-Cepeda, L. A., M. Martinez-Gonzalez, et al. (2005). “Oral candidosis as a clinical marker of immune failure in patients with HIV/AIDS on HAART.” *AIDS Patient Care STDS* **19**(2): 70–7.
- Haydel, M. J., C. A. Preston, et al. (2000). “Indications for computed tomography in patients with minor head injury.” *N Engl J Med* **343**(2): 100–5.
- Roddey, O. F., Jr., H. W. Clegg, et al. (1995). “Comparison of an optical immunoassay technique with two culture methods for the detection of group A streptococci in a pediatric office.” *J Pediatr* **126**(6): 931–3.
- Zangwill, K. M., D. H. Hamilton, et al. (1993). “Cat scratch disease in Connecticut. Epidemiology, risk factors, and evaluation of a new diagnostic test.” *N Engl J Med* **329**(1): 8–13.