**The Young Epidemiology Scholars Program (YES) is supported by
The Robert Wood Johnson Foundation and administered by the College Board.**

Screening and Diagnostic Tests

Manuel Bayona
Department of Epidemiology
School of Public Health
University of North Texas
Health Science Center
Fort Worth, Texas

and

Chris Olsen
Mathematics Department
George Washington High School
Cedar Rapids, Iowa

Contents

Lesson Plan

TITLE: Screening and Diagnostic Tests

SUBJECT AREA: Science, mathematics

TOPIC: Biology; making connections in mathematics

OBJECTIVES: The student should understand the necessity, purpose, procedures and limitations of diagnostic tests and their role in screening for a disease.

TIME FRAME: Two to 3 days.

PREREQUISITE KNOWLEDGE: Students should be comfortable with the ideas of conditional probability and be acquainted with Bayes's Theorem

MATERIALS NEEDED: A calculator would greatly facilitate calculations and cut down on the time needed for the students to finish the assignments.

PROCEDURE: It is well known that probability is a difficult topic for students. This module should be used only after some prior discussion of probability; the module itself focuses on the application of the probability. A reasonable procedure might be to lecture on day 1 on the topic and terminology, guide the students through the calculations or have them work in small groups on day 2, and give them the quiz as a take-home assignment on day 2 for discussion on day 3.

ASSESSMENT: As mentioned, the quiz could be given as a take-home assignment on day 2. There is no reason why these calculations should be done alone or without formulas—the worksheet assignment and discussion of the quiz will occur while students are still learning and becoming more comfortable with these ideas.

LINKS TO STANDARDS:

Mathematics

- Conditional probability, Bayes's Theorem
- Communication and connections

Science

- Evidence, models and explanation
- Nature of scientific knowledge

Bibliography

Gordis L. *Epidemiology* 2nd ed. Philadelphia: WB Saunders; 2000.

Lilienfeld DE, Stolley PD. *Foundations of Epidemiology*. New York: Oxford University Press; 1994.

Rothman, KJ, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1998.

Notes for Teachers

Goals and Objectives of This Module

The goal of this module is to introduce students to the concepts and practice of diagnostic tests as are commonly given by physicians. Many people do not appreciate that these diagnostic tests are based on probability and that there is a random or chance component to the outcomes of such tests. A greater understanding of this should help patients appreciate the nature of medical tests. Screening is also useful in a wide range of nonmedical areas (e.g., airport security screening, selection of tax forms for audit).

From a mathematical point of view, diagnostic testing provides a very important real life example of the use of probability and increases appreciation of the role of Bayes's Theorem in the logic behind the evaluation of diagnostic tests.

Epidemiologic Concepts to Be Covered

The following epidemiologic concepts related to screening and diagnostic exams are covered in this module: sensitivity, specificity and predictive value.

Prerequisites

The mathematical prerequisite generally would be placement in second-year algebra or wherever a strong unit of probability is taught. Bayes's Theorem should be presented, and the students should have practiced the calculations involved in the theorem.

Guide for Teachers

This module is designed to be an extension of the unit wherein conditional probability and Bayes's Theorem are presented. The lesson should be presented in 1–2 days, depending on the students' prior experience and success with conditional probability.

Probability is a topic that is not among the easiest to teach, and conditional probability is not the easiest of topics. Although teachers will differ in their judgment about providing formulas, it is suggested that the formulas be provided for the students' reference during their homework as well as the quiz used for assessment. There is great potential for “symbol shock” with the mathematics of this topic. If students are aware they need not memorize the formulas, they will find it easier to focus on the application of the mathematics.

Textbooks will differ in the probability topics covered. Because of this, background information about the probability needed for this module is provided for a quick review by instructors or possibly for handing out to students as a supplement to their text. Notation is also not consistent from text to text, and it is hoped that the notation used in this module is close enough to the students' text so that they are not unduly concerned.

Web Links

There are many Web links for the topic of diagnostic screening. Perhaps the best strategy for finding these is to use the conjunction of specificity and sensitivity in your Web browser. Each of the terms used singly may produce much chaff and only a little wheat.

Some of the Web sites that are particularly good as of this writing are:

The Medical University of South Carolina Doctoring Curriculum Web site: http://www.musc.edu/dc/icrebm/sensitivity.html

A set of Web page organized by the Rapid Diagnostics Web site and supported with funding from the United States Agency for International Development: http://www.rapid-diagnostics.org/accuracy.htm

Relation to Standards of the National Council of Teachers of Mathematics

This module reinforces the NCTM Data Analysis and Probability standard for grades 9–12 but will primarily address the strands of connections and communication. Using the mathematics of probability in the context of epidemiology will help students develop a view of mathematics as an integrated whole, rather than simply abstract manipulations of symbols.

References

An excellent and very readable account of screening and diagnostic is

Stolley P, Lasky T. *Investigating Disease Patterns: The Science of Epidemiology*. New York: WH Freeman; 1998.

A more mathematical but still readable treatment of these ideas is

Gordis L. *Epidemiology*. 2nd ed. Philadelphia: WB Saunders; 2000.

An excellent source for a review of the probability concepts involved in this module is

Ross S. *A First Course in Probability*. 6th ed. Upper Saddle River, NJ: Prentice Hall; 2002.

Data for the problems mentioned in this module were taken from the following sources:

Bellomo G, Narducci PL, Rondoni F, et al. Prognostic value of 24-hour blood pressure in pregnancy. *JAMA*. 1999;282: 1447–1452.

Dominguez-Bello M.G, Michelangeli F, Romero R, et al. Modification of Christensen urease test as an inexpensive tool for detection of *Helicobacter pylori*. *Diagn*. *Microbiol*. *Infect*. *Dis*. 1997;28:149–152.

Golightly ME. Laboratory considerations in the diagnosis and management of Lyme borreliosis. *Am*. *J*. *Clin*. *Pathol*. 1993;99(2):168–174.

Stafstrom CE, Rostasy K, Minster A. The usefulness of children's drawings in the diagnosis of headache. *Pediatrics*. 2002;109(3):460–472.

Background Information for Teachers

Part I: Basic Probability

In classical probability, the probability of an event, A, denoted by $P(A)$, is defined as the ratio of the number of outcomes favorable to A to the total number of outcomes. This classical definition has proven to be inadequate for empirical work because the usual requirement of equal likelihood is rarely met in practice. Today the so-called frequentist or empirical definition of probability is used, as follows: The **probability** of an event, A, denoted by $P(A)$, is defined to be the value approached by the relative frequency of occurrence of A in a very long series of trials of a chance experiment. Thus if the number of trials is quite large,

$$P(A) = \frac{\text{number of times } A \text{ occurs}}{\text{number of trials}}$$

If a chance experiment is repeated n times under essentially identical conditions and if the event A occurs m times, then as n grows large the ratio $^m/_n$ approaches a fixed limit that is by definition the probability of event A. At any given place in the sequence of trials of the chance experiment, the ratio is an approximation of the ideal probability:

$$P(A) \approx \frac{m}{n}$$

By virtue of this definition, $0 \leq P(A) \leq 1$. Impossible events are customarily given a probability of 0, and certain events are given a probability of 1. Impossible and certain events are generally not of great practical interest.

There exists a fairly small set of basic algebraic rules for probability:

1. For any event A, $0 \leq P(A) \leq 1$.
2. If two events, A and B, cannot both occur, $P(A \text{ or } B) = P(A) + P(B)$.
3. For any event, A, $P(A) + P(A^c) = 1$.

Events that cannot both occur are referred to as **disjoint** or **mutually exclusive events.** The event "not A," called the **complement** of A, is denoted by the symbol A^c.

Two fundamental algebraic operations on probabilities correspond to the usual understanding of the English words "or" and "and." The union of two events corresponds to the English "or" and is defined as follows:

The union of two events, A and B, denoted $A \cup B$,
is defined as the event either A or B or both occur.

The intersection of two events corresponds to the English "and" and is defined as follows:

The intersection of two events, A and B, denoted $A \cap B$,
is defined as the event both A and B occur.

Part II: Conditional Probability

A concept of crucial importance in understanding diagnostic and screening procedures in general and epidemiology in particular is that of **conditional probability.** This concept has applications in other fields such as business and economics, where is included in the so-called statistical decision theory. Such theory uses conditional probability to assess decision making under uncertainty and includes decision trees that are useful tools for formal decision making. The idea behind conditional probability is that knowledge about the occurrence or nonoccurrence of an event may provide information that allows a better estimate of the probability of a different event. For example, estimating the probability that a randomly selected human will be over six feet tall is very much helped by knowing whether the individual is a man or a woman. This concept is formalized as another algebraic operation:

Let A and B denote two events, with $P(B) > 0$. Then the conditional probability of the event A given that the event B has occurred, denoted by $P(A \mid B)$, is:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

It may also be true that knowledge of one event gives absolutely no information that will lead to a better calculation of the probability of a different event. This concept is formalized in probability as **independence.** The idea behind independence is that the calculation of the probability of event A is unaffected by the probability of event B, and the corresponding formal definition of independence is:

Two events, A and B, are independent if $P(A \mid B) = P(A)$. If A and B are not independent, they are said to be dependent.

The importance of conditional probability and independence for diagnostic screening is fundamental: Diagnostic and screening procedures depend on dependence. If, say, the appearance of an X-ray were independent of a patient's having tuberculosis, an X-ray would be useless for diagnosis.

The above algebraic operations of probability can be used to calculate the probabilities of unions and intersections of events. For any two events, A and B,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P(A \cap B) = P(A) \times P(B \mid A)$$

Less complex calculation formulas may be used if certain conditions are met:

If A and B are disjoint, $P(A \cup B) = P(A) + P(B)$ because $P(A \cap B) = 0$.
If A and B are independent, $P(A \cap B) = P(A) \times P(B)$ because $P(B \mid A) = P(B)$.

It is worth noting here that the concepts of disjoint (mutually exclusive) events and independent events are very commonly confused but are very different. If two events are mutually exclusive, knowledge that one event has occurred does provide information about the occurrence of the other event and thus the two events cannot be independent.

The formulas for the calculation of unions and intersections will, on occasion, allow simplification of some probability expressions:

$$P(A \cup A^C) = P(A \text{ or } A^C) = P(A) + P(A^C) = 1$$
$$P(A \cap A^C) = P(A \text{ and } A^C) = 0$$

Part III: Bayes's Theorem

The probabilities related to diagnosis and screening are based on the Bayes's Theorem, a theorem not without controversy in the history of mathematics and statistics. Bayes's Theorem is usually presented in all its generality, and comprehending it is a heroic exercise in algebraic interpretation. Fortunately it turns out that applying Bayes's Theorem to diagnostic and screening procedures will not require a general understanding because we are fundamentally interested in two events, which either will or will not occur, and we will be able to present Bayes's Theorem as simply as possible. The need for Bayes's Theorem arises because of the peculiarities of our knowledge about diseases and their effects. In diagnostic and screening procedures, we would like to make probability statements about the likelihood of a malady, given some symptom, e.g., P(Malady A | Symptom B). This probability cannot be estimated directly, owing to the state of our ignorance. What we can estimate, based on the observations of many doctors over many years, is a different probability, P(Symptom B | Malady A). Bayes's Theorem will allow the calculation of the probability we actually want, using the probability that we are able to estimate in the real world.

To set up the logic, we first observe two facts about maladies and diagnostic procedures. First, it must be true that a person has or does not have a particular malady. Second, it must be true that a person does or does not test positive using a diagnostic test. Two terms associated with this set of events are mutually exclusive, which we have already seen, and exhaustive. Events are said to be **exhaustive** if their probabilities sum to 1. Thus the events getting 1, 2, 3, 4, 5 or 6 with the roll of a die are mutually exclusive and also exhaustive: No two of these events can occur on a single roll, and no events other than these can occur on a roll.

Our first step toward Bayes's Theorem involves a formula that takes advantage of mutually exclusive and exhaustive events—call them B_1 and B_2. This formula is known as the law of total probability and relates an event A to the two events B_1 and B_2:

If the events B_1 and B_2 are mutually exclusive and exhaustive, then for any event A,

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) \\ &= P(A \mid B_1) \times P(B_1) + P(A \mid B_2) \times P(B_2) \end{aligned}$$

Suppose now that we wish to find the probability, $P(B_1 \mid A)$. Using some of our formulas above, we see that:

$$\begin{aligned} P(B_1 \mid A) &= \frac{P(B_1 \cap A)}{P(A)} \\ &= \frac{P(B_1 \cap A)}{P(A \cap B_1) + P(A \cap B_2)} \\ &= \frac{P(B_1 \cap A)}{P(A \mid B_1) \times P(B_1) + P(A \mid B_2) \times P(B_2)} \\ &= \frac{P(A \cap B_1)}{P(A \mid B_1) \times P(B_1) + P(A \mid B_2) \times P(B_2)} \\ &= \frac{P(A \mid B_1) \times P(B_1)}{P(A \mid B_1) \times P(B_1) + P(A \mid B_2) \times P(B_2)} \end{aligned}$$

This very important result is known as Bayes's Theorem. The importance of Bayes's Theorem, which can be easily lost even in this simplified algebraic form, is that we can calculate the conditional probability of B_1 given A—$P(B_1 \mid A)$—using the reverse conditional probability of A given B_1—$P(A \mid B_1)$.

Translating this into medical practice, suppose that the symbols $D+$, $D-$, $T+$ and $T-$ are defined as having a disease, not having a disease, testing positive on a diagnostic test and testing negative on a diagnostic test, respectively. Then what is desired from a diagnostic procedure is the probability of having the disease, given that the test is positive. (It is also of importance to assess the probability of not having the disease, given that the test is negative.) Using Bayes's Theorem,

$$\begin{aligned} P(D+ \mid T+) &= \frac{P(D+ \cap T+)}{P(T+)} \\ &= \frac{P(D+ \cap T+)}{P(T+ \cap D+) + P(T+ \cap D-)} \\ &= \frac{P(D+ \cap T+)}{P(T+ \mid D+) \times P(D+) + P(T+ \mid D-) \times P(D-)} \\ &= \frac{P(T+ \cap D+)}{P(T+ \mid D+) \times P(D+) + P(T+ \mid D-) \times P(D-)} \\ &= \frac{P(T+ \mid D+) \times P(D+)}{P(T+ \mid D+) \times P(D+) + P(T+ \mid D-) \times P(D-)} \end{aligned}$$

Given this result, it is clear that the desired probability can be calculated in terms of probabilities that can be estimated using data collected by health professionals.

Screening and Diagnostic Tests

Conditional probability is a very helpful tool used in many ways in everyday life. It is used to assess how probable the occurrence of an event is given that another related event has happened. This tool is of particular importance in business and economics, for example. Business statistical decision theory uses conditional probability to make decisions on what to do based on previous knowledge of the chances or probability of success.

Conditional probability in general and Bayes's Theorem in particular are also very important in medicine. At the level of individuals, an understanding of these topics is crucial in understanding diagnostic tests performed by our doctors. At the level of public health, epidemiologists depend on the accumulation of these diagnostic tests to judge how pervasive a particular disease is and how fast it is spreading—both essential to combating a disease and minimizing its effects over a population.

Science as we know it is fundamentally the result of a philosophy that says observation is important, and therefore measurement—the translation of observations into numeric form—is a primary concern. As all of us who have been in a science lab understand, measurement is not perfect. Suppose, for example, we are trying to measure the weight of a 100-kg man. If we take this person to 10 different scales, we might actually get 10 different measures due to slight inaccuracies in the scales. In addition, different people observing the same scale will see slightly different numbers.

These sorts of problems are collectively known as measurement problems, and it should not be surprising that these problems occur in medicine when physicians diagnose disease in individual patients. To see how this happens, consider the tick-borne disease known as Lyme disease, named in 1977 when arthritis was observed in a cluster of children in and around Lyme, Connecticut. Investigation revealed that Lyme disease is caused by the bacterium *Borrelia burgdorferi*, transmitted to humans by infected deer ticks. Lyme disease is diagnosed by a test that detects particular antibodies in the blood. From a measurement perspective, testing an individual for Lyme disease can have four possible results:

1. A person has Lyme disease and gets a positive diagnostic test result.
2. A person has Lyme disease and gets a negative diagnostic test result.
3. A person does not have Lyme disease and gets a positive diagnostic test result.
4. A person does not have Lyme disease and gets a negative diagnostic test result.

Using a 2 × 2 table, let's represent these possible outcomes of a diagnostic test.

Diagnosis: Lyme Disease

	Has Lyme	**Does Not Have Lyme**
Positive Test	Correct diagnosis	Incorrect diagnosis
Negative Test	Incorrect diagnosis	Correct diagnosis

Before the more litigious among you run off to make appointments with attorneys, it must be pointed out that these errors of diagnosis are not indictments of physicians—they are concessions to the limitations of chemistry. It may be that a positive test will occur if one has a similar disease that produces similar antibodies that confuse the diagnostic test. It may be that a negative test will result because of some rare chemical interaction due to an individual's biochemistry. Or it is possible that a test may give an ambiguous result and be difficult for the physician to interpret.

Doctors, of course, are aware that tests can be less than definitive and will frequently schedule more tests or have the lab reanalyze a blood sample with the same diagnostic test. Patients are also to some extent aware that diagnoses are not etched in stone and will sometimes go for a second opinion. We would like to explain just how a particular diagnostic test is evaluated, using the concepts of conditional probability. We begin with the commonsense notion that a patient either does or does not have a particular malady and that a diagnostic test will return either a positive or negative result. Our generic 2 by 2 table would look something like this:

Diagnosis: Disease X

	Has X	**Does Not Have X**
Positive Test	Correct diagnosis	Incorrect diagnosis
Negative Test	Incorrect diagnosis	Correct diagnosis

When a diagnostic test is developed, its capability of delivering a correct diagnosis—a positive test for those who have X and a negative test for those who do not have X—is analyzed during the development process. For purposes of discussion, let's suppose that the diagnostic test for our disease X is a blood test, such as the common test for Lyme disease. As a part of the quality control by the manufacturer, samples of blood from individuals known to have X and samples of blood from individuals known to be X free are acquired.

Now the astute student might be wondering why, if there is a method that gives known results, anyone would want a new diagnostic test and moreover one that would be fallible. There are two reasons for this. In some cases the best method for ascertaining the known disease state is an autopsy after the death of the patient. Because the information comes too late to do any good, it can't be a helpful diagnostic test for any patients. In other instances, a company may wish to provide a diagnostic test that is just as good as the existing diagnostic test, and is less expensive or can provide test results in a more timely manner. In either case the method of ascertaining the known disease state is called the gold standard, and a new diagnostic procedure would be tested against the gold standard. What is desired is a diagnostic test that agrees with the gold standard:

Diagnosis: Disease X

	Gold Standard Has X	**Gold Standard No X**
Positive Test	Agreement	Disagreement
Negative Test	Disagreement	Agreement

Suppose we used our new diagnostic technique on 10,000 blood samples, with the following results:

Mythical Diagnostic Test Evaluation Results: Round 1

	Gold Standard Has X	**Gold Standard No X**
Positive Test	1,000	0
Negative Test	0	9,000

As can be easily seen, the test results are in perfect agreement with the gold standard. Needless to say, results like this don't happen in the real world. We are more likely to see results that look like this:

Mythical Diagnostic Test Evaluation Results: Round 2

	Gold Standard Has X	**Gold Standard No X**	**Total**
Positive Test	950	900	1,850
Negative Test	50	8,100	8,150
Total	1,000	9,000	10,000

In this table we can see that the diagnostic test is correct sometimes and incorrect other times. The laboratory developing the diagnostic test procedure is responsible for estimating the probabilities of error, and these probabilities are estimated from the experimental results. Based on the data in the table above, these estimates are:

$$P(\text{Positive test} \mid \text{Patient has disease}) = \frac{950}{1{,}000} = 0.950$$

$$P(\text{Positive test} \mid \text{Patient does not have the disease}) = \frac{900}{9{,}000} = 0.100$$

$$P(\text{Negative test} \mid \text{Patient has disease}) = \frac{50}{1{,}000} = 0.050$$

$$P(\text{Negative test} \mid \text{Patient does not have the disease}) = \frac{8{,}100}{9{,}000} = 0.900$$

To simplify matters, we will create some notation. Let D and d represent having and not having the disease, respectively; also, let + and − represent getting a positive and negative test result, respectively. Then we can rewrite the above probabilities as follows:

$$P(+ \mid D) = \frac{950}{1{,}000} = 0.950 \qquad P(+ \mid D) = \frac{900}{1{,}000} = 0.100$$

$$P(- \mid D) = \frac{50}{1{,}000} = 0.050 \qquad P(+ \mid D) = \frac{8{,}100}{9{,}000} = 0.900$$

In the world of medicine and epidemiology, these probabilities have special definitions:

Probabilities and Terminology

Probability	**Terminology**
$P(+ \mid D)$	Sensitivity
$P(- \mid d)$	Specificity
$P(+ \mid d)$	False-positive rate
$P(- \mid D)$	False-negative rate

We will return now to our Lyme disease diagnostic test and substitute some real values, consistent with actual values for the common Lyme disease diagnostic test. The prevalence of a characteristic or condition is defined as the proportion of people who have that particular characteristic or condition. The actual prevalence rate for Lyme disease is approximately 0.008 in the United States, and we will assume this value for our calculations.

Lyme Disease Diagnostic Test

	Gold Standard Lyme	Gold Standard No Lyme	Total
Positive Test	15	60	75
Negative Test	1	1,924	1,925
Total	16	1,984	2,000

On the basis of these numbers, we would make the following probability calculations, using L to indicate having Lyme disease and *not L* to indicate not having Lyme disease:

$$\text{Sensitivity} = P(+ \mid L) = \frac{15}{16} = 0.9375$$

$$\text{False-positive rate} = P(+ \mid \textit{not } L) = \frac{60}{1{,}984} = 0.0302$$

$$\text{False-negative rate} = P(- \mid L) = \frac{1}{16} = 0.0625$$

$$\text{Specificity} = P(- \mid \textit{not } L) = \frac{1{,}924}{1{,}984} = 0.9698$$

Now that we have defined these important quantities and shown how to calculate them, we must sheepishly admit that none of these quantities, by themselves, are helpful in diagnosing Lyme disease! This shocking turn of events can be understood by considering the nature of conditional probability. In each case, these probabilities are probabilities of test results, given patients' disease status.

That is, we have probabilities like this:

$$\text{Sensitivity} = P(+ \mid L)$$
$$\text{False-positive rate} = P(+ \mid \textit{not } L)$$
$$\text{False-negative rate} = P(- \mid L)$$
$$\text{Specificity} = P(- \mid \textit{not } L)$$

From the standpoint of performing a diagnosis, we can immediately detect a problem when we consider the given of each of the conditional probabilities: For each of these the disease status is assumed known. What physicians and patients want from a diagnostic test is the probabilities of disease status, given the test results. In other words:

$P(L \mid +)$ = The probability of having the disease, if a test is positive.
$P(\textit{not } L \mid -)$ = The probability of not having the disease, if a test is negative.

The probability of having the disease, given a positive test, is known as the **positive predictive value.** The probability of not having the disease, given a negative test, is known as the **negative predictive value.** These quantities are, as it were, just what the doctor ordered. Given that a random person has a positive test, what's the probability that she or he actually has the disease? Given a random person's negative test, what is the probability of not having the disease? These are the informative diagnostic probabilities, and it is clear that we have not yet actually calculated them. How can we take the probabilities we have and compute the probabilities we want? It is precisely here that Bayes's Theorem comes to our aid. In terms of our symbols for test results, Lyme disease status and the probabilities estimated during the laboratory's development of the diagnostic test, we can calculate the positive predictive value using Bayes's Theorem:

$$P(L \mid +) = \frac{P(+ \mid L)P(L)}{P(+ \mid L)P(L) + P(+ \mid \textit{not } L)P(\textit{not } L)}$$

Substituting the values from our table above,

$$\begin{aligned} P(L \mid +) &= \frac{(0.9375)(0.008)}{(0.9375)(0.008) + (0.0302)(0.992)} \\ &= 0.200 \end{aligned}$$

Similarly, we can calculate the negative predictive value:

$$P(\textit{not } L \mid -) = \frac{P(- \mid \textit{not } L)P(\textit{not } L)}{P(- \mid L)P(L) + P(- \mid \textit{not } L)P(\textit{not } L)}$$

$$\begin{aligned} P(\textit{not } L \mid -) &= \frac{(0.9698)(0.992)}{(0.0625)(0.008) + (0.9698)(0.992)} \\ &= 0.9995 \end{aligned}$$

Notice that the diagnostic test's ability to detect Lyme disease is not particularly good, but its ability to absolve patients of worry about Lyme disease is excellent. It is not uncommon for a diagnostic test to be better at one of these tasks and is one reason why physicians might schedule more than one diagnostic test for the same disease.

At this point we have traced the development of a diagnostic test from inception. In the development stage, the test is compared with an existing gold standard for detecting a disease state. Then, by using Bayes's Theorem and the estimates of the probabilities from the development stage, probabilities that are more closely relevant to doctor and patient are found. When the predictive values and the probabilities of false-positive and -negative are evaluated,

doctor and patient are able to interpret the results of such tests as a part of the diagnostic strategy by the physician.

We will now give you a chance to practice these calculations on some actual diagnostic tests. Remember, these calculations can get involved, so you should be careful to show your work.

Worksheet: Practice Problems in Screening (Student's Version)

Name _____

Directions: Using the definition of prevalence given above, answer the following questions. Be sure to show your work.

1. Describe in a few words the difference between specificity and sensitivity as the terms apply to diagnostic testing.

2. Describe in a few words the difference between the predictive value of a positive test and the predictive value of a negative test.

3. Bacteria? In my stomach? *Helicobacter pylori* is a spiral bacterium that lives in the stomach and duodenum (section of intestine just below stomach). The inside of the stomach is exposed to gastric juice and is protected by a thick layer of mucus that covers the stomach lining. *H. pylori* lives in this mucous lining. It is responsible for peptic ulcer disease and chronic gastritis and is thought to be a risk factor for stomach cancer. About half the world's population is infected with *H. pylori,* most living in developing countries. A new low-cost diagnostic test, the local urease test (LUT), has been developed to test for the presence of the bacterium. The performance of LUT in actual trials is presented in the table

below. When you perform the calculations, use the U.S. prevalence, about 0.0001. (That is, the probability of being infected in the United States is about 0.0001, or 1 in 10,000.)

	Cases (Infected)	**Controls (Noninfected)**
+ LUT	126	1
- LUT	0	84

a. Calculate the specificity and sensitivity of this diagnostic test and briefly interpret your results regarding the probability of a positive test when the person has the infection and the probability of a negative test when the person does not have the infection

b. Calculate the positive and negative predictive values of this diagnostic test and briefly interpret your results regarding the diagnostic abilities of this test—that is, the probability of having the disease given that the test was positive and the probability of not having the disease given that the test was negative.

4. White coat hypertension. Preeclampsia is a disorder that occurs only during pregnancy and affects both the mother and the unborn baby. Affecting at least 5% of all pregnancies, it is a rapidly progressive condition often characterized by high blood pressure. Sudden weight gain, headaches and changes in vision are important symptoms of preeclampsia. A slight difficulty in diagnosing hypertension in the doctor's office is that patients may be nervous just at being in the place and have a temporarily elevated blood pressure because of their nervousness. A false diagnosis of elevated blood pressure would subject pregnant women to unneeded and expensive diagnostic procedures or drug intervention for a condition that does not exist.

In a hospital study conducted from 1994 to 1997, two methods of diagnostic screening were compared. The first diagnostic method was a 24-hour blood pressure monitoring device. The second was a blood pressure measurement performed at an office visit. Women who had been previously clinically evaluated for hypertension were subjects for this study. The results for each diagnostic procedure are presented below. A positive result is a blood pressure indication that is outside the normal range. Each subject was measured using both techniques.

24-hr BP Monitoring Diagnostic Data

	Hypertensive	Not Hypertensive	Total
Positive 24-hr BP Measure	60	6	66
Negative 24-hr BP Measure	12	169	181
Total	72	175	247

Office BP Measurement Diagnostic Data

	Hypertensive	Not Hypertensive	Total
Positive Office BP Measure	63	39	102
Negative Office BP Measure	9	136	145
Total	72	175	247

a. On page 15 of the screening module, the terminology associated with various probabilities was defined. Which of these probabilities would be calculated if the investigators are concerned about the rate at which nonhypertensive patients are classified as having high blood pressure?

b. Calculate the probabilities from part a for the 24-hour monitoring and the office procedure. Of the two diagnostic procedures, which would be better in response to the investigators' concern? Justify your response by appealing to the data.

c. One difficulty with the 24-hour monitoring is the time it takes—i.e., 24 hours—whereas a doctor's visit is much shorter. Write a short paragraph for a women's health brochure that explains the benefits and costs of each diagnostic method. You should mention the risks of a diagnostic error and balance that against the time factor. This brochure would be intended to help women make more informed choices about which diagnostic method they might like to use. Assume the dollar cost of the two procedures is about the same.

Worksheet: Practice Problems in Screening (Teacher's Version)

Name _____

Directions: Using the definition of prevalence given above, answer the following questions. Be sure to show your work.

1. Describe in a few words the difference between specificity and sensitivity as the terms apply to diagnostic testing.

   Specificity refers to the prevalence of negative test results in people without a disease, while sensitivity refers to the prevalence of positive test results in people with the disease.

2. Describe in a few words the difference between the predictive value of a positive test and the predictive value of a negative test.

   Positive predictive value refers to the probability that a person with positive test has the disease, whereas negative predictive value refers to the probability that a person with a negative test does not have the disease.

3. Bacteria? In my stomach? *Helicobacter pylori* is a spiral bacterium that lives in the stomach and duodenum (section of intestine just below stomach). The inside of the stomach is exposed to gastric juice and is protected by a thick layer of mucus that covers the stomach lining. *H. pylori* lives in this mucous lining. It is responsible for peptic ulcer disease and chronic gastritis and is thought to be a risk factor for stomach cancer. About half the world's population is infected with *H. pylori,* most living in developing countries. A new low-cost diagnostic test, the local urease test (LUT), has been developed to test for the presence of the bacterium. The performance of LUT in actual trials is presented in the table

below. When you perform the calculations, use the U.S. prevalence, about 0.0001. (That is, the probability of being infected in the United States is about 0.0001, or 1 in 10,000.)

	Cases (Infected)	Controls (Noninfected)
+ LUT	126	1
- LUT	0	84

a. Calculate the specificity and sensitivity of this diagnostic test and briefly interpret your results regarding the probability of a positive test when the person has the infection and the probability of a negative test when the person does not have the infection

The sensitivity appears to be 100% (126/126) and the specificity, 99% (83/84). Thus, a person with the disease appears to have a 100% chance to have a positive test, whereas a person without the disease appears to have a 99% chance to have a negative test.

b. Calculate the positive and negative predictive values of this diagnostic test and briefly interpret your results regarding the diagnostic abilities of this test—that is, the probability of having the disease given that the test was positive and the probability of not having the disease given that the test was negative.

The positive predictive value is 0.8% ([126/126][0.0001]/([126/126][0.0001]+[1/85][0.9999]) whereas the negative predictive value is 100%. Thus, in the U.S. population the test would be excellent at ruling out *H. pylori* infection (because of its high negative predictive value) but not good at confirming H. pylori infection (because of its low positive predictive value).

4. White coat hypertension. Preeclampsia is a disorder that occurs only during pregnancy and affects both the mother and the unborn baby. Affecting at least 5% of all pregnancies, it is a rapidly progressive condition often characterized by high blood pressure. Sudden weight gain, headaches and changes in vision are important symptoms of preeclampsia. A slight difficulty in diagnosing hypertension in the doctor's office is that patients may be nervous just at being in the place and have a temporarily elevated blood pressure because of their nervousness. A false diagnosis of elevated blood pressure would subject pregnant women to unneeded and expensive diagnostic procedures or drug intervention for a condition that does not exist.

In a hospital study conducted from 1994 to 1997, two methods of diagnostic screening were compared. The first diagnostic method was a 24-hour blood pressure monitoring device. The second was a blood pressure measurement performed at an office visit. Women who had been previously clinically evaluated for hypertension were subjects for this study. The results for each diagnostic procedure are presented below. A positive result is a blood pressure indication that is outside the normal range. Each subject was measured using both techniques.

24-hr BP Monitoring Diagnostic Data

	Hypertensive	Not Hypertensive	Total
Positive 24-hr BP Measure	60	6	66
Negative 24-hr BP Measure	12	169	181
Total	72	175	247

Office BP Measurement Diagnostic Data

	Hypertensive	Not Hypertensive	Total
Positive Office BP Measure	63	39	102
Negative Office BP Measure	9	136	145
Total	72	175	247

a. On page 15 of the screening module, the terminology associated with various probabilities was defined. Which of these probabilities would be calculated if the investigators are concerned about the rate at which nonhypertensive patients are classified as having high blood pressure?

False-positive rate.

b. Calculate the probabilities from part a for the 24-hour monitoring and the office procedure. Of the two diagnostic procedures, which would be better in response to the investigators' concern? Justify your response by appealing to the data.

The false-positive rates are 3% (6/175) for 24-hour monitoring and 22% (29/175) for office measurement. Therefore 24-hour monitoring would yield fewer false-positive results and so better address the concern not to miss cases of hypertension.

c. One difficulty with the 24-hour monitoring is the time it takes—i.e., 24 hours—whereas a doctor's visit is much shorter. Write a short paragraph for a women's health brochure that explains the benefits and costs of each diagnostic method. You should mention the risks of a diagnostic error and balance that against the time factor. This brochure would be intended to help women make more informed choices about which diagnostic method they might like to use. Assume the dollar cost of the two procedures is about the same.

"If you have high blood pressure that goes untreated, this could harm you or your unborn child. We can test you for high blood pressure either by measuring it in the doctor's office or giving you a blood pressure monitor to wear for 24 hours. Neither test is perfect. Both are safe and they cost about the same. The monitor is more of a bother, but has the advantage of fewer false-positive results, which could save you the additional inconvenience of further diagnostic tests. On the other hand, office measurement is a bit more likely to detect high blood pressure if you actually have it."

Quiz: Screening and Diagnostic Tests

Name _____

1. In a few sentences, distinguish between specificity and sensitivity as related to the probability of finding that a patient has a positive test when the test is applied to a group of patients known to have the disease, and the probability of finding a patient with a negative test when the test is applied to a group of individuals known to be free of the disease.

2. In what way do sensitivity and specificity differ from the false-positive and false-negative rates?

3. Following the formulas below, could you briefly explain in your own words why we can or cannot calculate the predictive values from a 2×2 table from a cohort study based on the Bayes's Theorem? (Hint: Predictive value of a positive test = true positives/all positives; predictive value of a negative test = true negatives/all negatives).

$$
\begin{aligned}
P(D+ \mid T+) &= \frac{P(D+ \cap T+)}{P(T+)} \\
&= \frac{P(D+ \cap T+)}{P(T+ \cap D+) + P(T+ \cap D-)} \\
&= \frac{P(D+ \cap T+)}{P(T+ \mid D+) \times P(D+) + P(T+ \mid D-) \times P(D-)}
\end{aligned}
$$

4. A very common children's complaint is headache. Doctors categorize headaches as being of two kinds: migraine and nonmigraine. Correct diagnosis is important because the treatments will differ for the two types of headache. Diagnosis of headaches is difficult when the patient is a child unable to explain symptoms verbally. Children's drawings have been used with success to analyze subjective feeling. Might this work for headaches? In a recent study children at a pediatric clinic who complained of headaches were given a piece of paper and pencil, and asked:

   Please draw a picture of yourself having a headache. Where is your pain? What does your pain feel like? Are there any other changes or symptoms that come before or during your headache that you can show me in a picture?

One group of physicians categorized the children's drawings as indicating symptoms of migraine, and a different group of physicians provided the usual clinical diagnosis of the children's headaches as a gold standard. Their hope is that they can use these drawings as a diagnostic tool when working with children. The results of their observations are in the table below:

Pediatric Migraine Headache: Pictures for Diagnosis

	Migraine Diagnosed Clinically	**Nonmigraine Diagnosed Clinically**	**Total**
Positive Test: Migraine Features in Drawing	121	18	139
Negative Test: No Migraine Features in Drawing	9	87	96
Total	130	105	235

a. Calculate the positive predictive value of this diagnostic test.

b. Calculate the negative predictive value of this diagnostic test.

c. Because the treatments for migraine are different from those for nonmigraine headaches, the doctors need a good way to diagnose which type of headache the child has. If the physician were told the positive predictive value of this drawing test, would that be enough information to evaluate the diagnostic test? Why or why not?

Quiz: Screening and Diagnostic Tests (Answer Key)

Name _____

1. In a few sentences, distinguish between specificity and sensitivity as related to the probability of finding that a patient has a positive test when the test is applied to a group of patients known to have the disease, and the probability of finding a patient with a negative test when the test is applied to a group of individuals known to be free of the disease.

   Specificity refers to the prevalence of negative test results in people without a disease, while sensitivity refers to the prevalence of positive test results in people with the disease. A test with high specificity has few false positives. A test with high sensitivity has few false negatives.

2. In what way do sensitivity and specificity differ from the false-positive and false-negative rates?

   Specificity = 1 − false-positive rate. Sensitivity = 1 − false-negative rate.

3. Following the formulas below, could you briefly explain in your own words why we can or cannot calculate the predictive values from a 2 × 2 table from a cohort study based on the Bayes's Theorem? (Hint: Predictive value of a positive test = true positives/all positives; predictive value of a negative test = true negatives/all negatives).

$$
\begin{aligned}
P(D+ \mid T+) &= \frac{P(D+ \cap T+)}{P(T+)} \\
&= \frac{P(D+ \cap T+)}{P(T+ \cap D+) + P(T+ \cap D-)} \\
&= \frac{P(D+ \cap T+)}{P(T+ \mid D+) \times P(D+) + P(T+ \mid D-) \times P(D-)}
\end{aligned}
$$

In the typical 2 × 2 table of a diagnostic test applied to a cohort of size N, "a" is an estimate of the product of prob(D+ and T+) and N and "a + c" is an estimate of the product of [prob (D+ and T+) + prob(D− and T+)] and N, where N = a + b + c + d. Correspondingly, "d" is an estimate of the product of prob(D− and T−) and N and "b + d" is an estimate of the product of [prob (D− and T−) + prob(D+and T−)] and N. Therefore, from a 2 × 2 table one can estimate the positive predictive value by a/(a + c) and the negative predictive value by d/(b + d).

4. A very common children's complaint is headache. Doctors categorize headaches as being of two kinds: migraine and nonmigraine. Correct diagnosis is important because the treatments will differ for the two types of headache. Diagnosis of headaches is difficult when the patient is a child unable to explain symptoms verbally. Children's drawings have been used with success to analyze subjective feeling. Might this work for headaches? In a recent study children at a pediatric clinic who complained of headaches were given a piece of paper and pencil, and asked:

   Please draw a picture of yourself having a headache. Where is your pain? What does your pain feel like? Are there any other changes or symptoms that come before or during your headache that you can show me in a picture?

One group of physicians categorized the children's drawings as indicating symptoms of migraine, and a different group of physicians provided the usual clinical diagnosis of the children's headaches as a gold standard. Their hope is that they can use these drawings as a diagnostic tool when working with children. The results of their observations are in the table below:

Pediatric Migraine Headache: Pictures for Diagnosis

	Migraine Diagnosed Clinically	Nonmigraine Diagnosed Clinically	Total
Positive Test: Migraine Features in Drawing	121	18	139
Negative Test: No Migraine Features in Drawing	9	87	96
Total	130	105	235

a. Calculate the positive predictive value of this diagnostic test.

Positive predictive value is 87% (121/139).

b. Calculate the negative predictive value of this diagnostic test.

Negative predictive value is 91% (87/96).

c. Because the treatments for migraine are different from those for nonmigraine headaches, the doctors need a good way to diagnose which type of headache the child has. If the physician were told the positive predictive value of this drawing test, would that be enough information to evaluate the diagnostic test? Why or why not?

No. The positive and negative predictive values depend not only on the characteristics of the test but also the prevalence of the condition (headache in this instance) in the population tested. With a different population the same test could give very different predictive values. Therefore, one would also want to know (at the least) the sensitivity and specificity of the test.