Math Minute 4.4 **What Does a Positive Test Result Really Mean?**

What does it mean if your test result for a genetic disease (e.g., cystic fibrosis; CF) is positive? What is the probability you actually have the disease, given that the test was positive? You must know the answer to this question before you can respond intelligently to a positive test result. The probability that the positive test result is correct depends on three other probabilities that are commonly associated with disease testing: prevalence, sensitivity, and specificity. To explain these three probabilities and how they are used to determine the probability that your positive result is correct, we consider the case of testing a single Caucasian individual for CF. You can explore other diseases and populations in the Math Minute Discovery Questions.

Prevalence is the overall probability of an individual having the disease, estimated by the proportion of the population with the disease. The prevalence of CF in Caucasians is $P(CF) = 1/2500 = 0.0004$ (assuming no prior knowledge of CF disease alleles in the individual's pedigree).

Sensitivity is the probability that the genetic test will be positive when the individual really has the disease. The sensitivity of the CF genetic test varies with the number of alleles tested. If the 21 common alleles are tested, sensitivity is $P(+ \mid CF) = 0.85$. The vertical bar between "+" and "CF" can be read as "assuming" or "given"; the probability of a positive test result given the person has cystic fibrosis is 85%.

Specificity is the probability that the test will be negative given that the individual does not have the disease. Poor specificity may be caused by technical errors or difficulties in the testing procedures. The specificity of a test can be estimated by studying those who test negative, to see if they later develop the disease. We will suppose that the specificity of the CF test is $P(- \mid \text{no CF}) = 0.999$. Because a person who does not have CF must test either positive or negative, another way of saying the same thing is $P(+ \mid \text{no CF}) = 0.001$. In other words, only one out of 1,000 individuals who do not have CF would test positive.

Now we put together prevalence, sensitivity, and specificity to compute $P(CF \mid +)$, the probability of having CF given a positive test result. The formula for calculating $P(CF \mid +)$ is called Bayes' Rule, an important topic in probability and statistics:

$$P(CF \mid +) = \frac{P(+ \mid CF)P(CF)}{P(+ \mid CF)P(CF) + P(+ \mid \text{no CF})P(\text{no CF})}$$

$$= \frac{0.85 \times 0.0004}{0.85 \times 0.0004 + 0.001 \times 0.9996} \approx 0.2538.$$

Bayes' Rule tells us that if you test positive for CF, there is only about a 25% chance that you actually have the disease. This result is counterintuitive because the test is fairly sensitive (0.85) and very specific (0.999). However, we have proven mathematically that the probability of a correct positive test is much smaller than the sensitivity and specificity might suggest, so we know our calculation is correct.

An alternative way to compute $P(CF \mid +)$ might help you understand Bayes' Rule and see why $P(CF \mid +)$ is lower than you might have expected. Suppose 10 million people are tested for CF. You would expect about 4,000 people in this population to have CF. Of these 4,000, you would expect only 3,400 ($4{,}000 \times 0.85$) to test positive. In contrast, of the 9,996,000 without CF, you would expect about 9,996 (i.e., $9{,}996{,}000 \times 0.001$) to test positive. Thus, 13,996 people would test positive, but only about 25% (4,000/13,996) of them actually would have CF. Of course, randomness in the population means that none of these numbers are exact, which is why Bayes' Rule uses probabilities instead of specific population counts.

MATH MINUTE DISCOVERY QUESTIONS

1. Who should be more skeptical of a positive CF test, a Caucasian or an African American? Use a calculator to compute the appropriate probabilities

to support your answer. (The prevalence and sensitivity of CF testing in African Americans are given in this section of the text.)
Note: You can use genetictests.xls to help you answer Math Minute Discovery Questions 2, 3, and 5.

2. Explore the effects of sensitivity on disease testing by computing P(CF | +) when sensitivity is between 0.6 and 0.99, keeping the prevalence 0.0004 and the specificity 0.999.

3. Explore the effect of specificity on disease testing by computing P(CF | +) when specificity is between 0.9 and 0.9999, keeping the prevalence 0.0004 and the sensitivity 0.85.

4. Which do you think should be a higher policy priority, finding more CF alleles, or decreasing the false positive rate of CF tests? Explain your answer.

5. Explore the effect of prevalence by computing P(CF | +) when prevalence is between 0.0004 and 0.4, keeping the sensitivity 0.85 and the specificity 0.999. For what value of prevalence would P(CF | +) be maximized?

LINKS
deCODE
MATH MINUTES
genetictests.xls

Math Minute 4.4 What Does a Positive Test Result Really Mean?

1. Who should be more skeptical of a positive CF test, a Caucasian or an African American? Use a calculator to compute the appropriate probabilities to support your answer. (The prevalence and sensitivity of CF testing in African Americans are given in this section of the text.)

Compute P(CF | +) for a Caucasion: 0.2538 (see text of Math Minute 4.4), and P(CF | +) for an African American: 0.039 [using P(+ | CF) = 0.69 and P(CF) = 1/17,000]. Therefore an African American should be more skeptical of a positive CF test.

Note: You can use genetictests.xls to help you answer Math Minute Discovery Questions 2, 3, and 5.

2. Explore the effects of sensitivity on disease testing by computing P(CF | +) when sensitivity is between 0.6 and 0.99, keeping the prevalence 0.0004 and the specificity 0.999.

When sensitivity ranges from 0.6 to 0.99, P(CF | +) ranges between 0.19 and 0.28.

3. Explore the effect of specificity on disease testing by computing P(CF|+) when specificity is between 0.9 and 0.9999, keeping the prevalence 0.0004 and the sensitivity 0.85.

Specificity has a major effect, with P(CF | +) ranging from 0.00339 to 0.773 as specificity ranges from 0.9 to 0.9999.

4. Which do you think should be a higher policy priority, finding more CF alleles, or decreasing the false positive rate of CF tests? Explain your answer.

Specificity has a great effect on P(CF | +). For example, with a prevalence of 1/17,000 (African American rate), increasing the specificity by an order of magnitude (from 0.999 to 0.9999) increases this probability from 0.039 to 0.289. Another jump in order of magnitude (specificity = 0.99999, 1 error in 100,000 tests) would cause the probability to increase to 0.802.

Increasing sensitivity by finding more CF alleles has a smaller effect on P(CF | +). For example, with a prevalence of 1/17,000 (African American rate), increasing the sensitivity from 0.69 to 0.95 only increases the probability from 0.039 to 0.053.

However, P(CF | +) is not the only consideration. This probability only models how skeptical one should be of a positive test, and does not reflect the ability of repeated tests to determine disease status. The inequity in sensitivity of CF tests across racial groups could be addressed by finding more CF alleles, and even though P(CF | +) does not increase dramatically, there is at least a better chance of detecting the presence of CF through repeated testing.

5. Explore the effect of prevalence by computing $P(CF \mid +)$ when prevalence is between 0.0004 and 0.4, keeping the sensitivity 0.85 and the specificity 0.999. For what value of prevalence would $P(CF \mid +)$ be maximized?

Prevalence	P(CF | +)
.0004	.2538
.004	.7734
.04	.9725

P(CF | +) would be maximized at a prevalence of 1. If everyone has CF, then no one should be skeptical of a positive test. The rareness of CF is what makes P(CF | +) so small, even for high specificity and sensitivity.