

Diagnostic Tests in Medicine

Keywords: combinatorics, probability and statistics, conditional probability, frequency of occurrence

Every person undergoes many diagnostic tests during their lifetime, either at the doctor's office or at home. It doesn't matter whether it is a test for COVID-19, celiac disease, general mammogram screening, or a home pregnancy test. In any case, we want to know whether the test result is reliable. For each such test, two basic pieces of information are provided:

- *Test Sensitivity*: the probability that the test will detect the disease when I have it.
- *Test Specificity*: the probability that the test will be negative when I do not have the disease.

The so-called *prevalence of the disease*, i.e., the ratio of the number of individuals who have the disease to the number of all individuals in the population, is also important for evaluation.

Test Result Interpretation

A number of home tests for allergies and intolerances are sold in pharmacies and drugstores, including COVID-19 or pregnancy tests. For example, the leaflet of a test for celiac disease (gluten intolerance) states that the sensitivity of the test is 96,3% and the specificity of the test is 89,7%. We know that the number of celiacs in the population is roughly 1%, so the prevalence of the disease is 1/100.

We will be primarily interested in the question: If I test positive, what is the probability that I actually have celiac disease? Try to guess the answer first. In the following exercise, we will perform the exact calculation.

Exercise 1. The test sensitivity is 96,3%, the test specificity is 89,7% and the prevalence of the disease is 1/100. What is the probability that a patient who tests positive for celiac disease actually has the disease?

Solution. This is an example of a problem on conditional probability that can be easily solved using Bayes' theorem:

Let A be a random event and let B_1, \dots, B_n be a complete set of events. If $P(A) > 0$, $P(B_1) > 0, \dots, P(B_n) > 0$, then

$$P(B_k|A) = \frac{P(A \cap B_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^n P(A|B_i)P(B_i)}, \quad k = 1, \dots, n.$$

The symbol $P(B|A)$ denotes the probability that event B occurs given that event A has occurred.

Let's define the following probabilities:

- $P(C)$ the probability that I'm positive (I have the disease),
- $P(N)$ the probability that I'm negative (I don't have the disease),
- $P(+|C)$ the probability that the test is positive if I'm positive,
- $P(-|C)$ the probability that the test is negative if I'm positive,

Results matter!

- $P(+|N)$ the probability that the test is positive if I'm negative,
- $P(-|N)$ the probability that the test is negative if I'm negative,
- $P(C|+)$ the probability that I'm positive if the test is positive.

Since the prevalence is 1/100, then

$$P(C) = 0,01 \quad P(N) = 0,99$$

and from the given data

$$P(+|C) = 0,963 \quad P(-|C) = 0,037 \quad P(+|N) = 0,103 \quad P(-|N) = 0,897.$$

According to Bayes' theorem, it holds that

$$\begin{aligned} P(C|+) &= \frac{P(+|C)P(C)}{P(+|C)P(C) + P(+|N)P(N)} = \\ &= \frac{0,963 \cdot 0,01}{0,963 \cdot 0,01 + 0,103 \cdot 0,99} = 0,086. \end{aligned}$$

The probability that a patient who tests positive really has the disease is therefore only 0,086, i.e., 8,6 %

Note. It is very likely that many students will not guess this result correctly before the calculation. This is not surprising, as studies have shown that up to 95% of doctors cannot answer questions like this correctly.

Exercise 2. What is the probability that a patient who tests negative is really negative?

Solution. We will use the notation from the previous exercise and Bayes' theorem. We get

$$\begin{aligned} P(N|-) &= \frac{P(-|N)P(N)}{P(-|N)P(N) + P(-|C)P(C)} = \\ &= \frac{0,897 \cdot 0,99}{0,897 \cdot 0,99 + 0,037 \cdot 0,01} = 0,999583. \end{aligned}$$

Thus, the probability that a patient who tests negative is actually healthy is 0,999583, i.e., 99,96 %.

Interpretation Without Using Probability

Let's try to solve the first exercise again, but we will reframe the assignment to avoid the concept of conditional probability.

Exercise 3. Out of 100 000 people, about 1 000 have celiac disease. Of these 1 000 people, the test detects celiac disease in 960 cases. Of the remaining 99 000 people who don't have celiac disease, 10 200 people will also test positive. Let's imagine the group of those people who test positive. What proportion of them actually have celiac disease?

Solution. In total, we have

$$960 + 10\,200 = 11\,160$$

positive test results, but only 960 are true positive cases, so:

$$\frac{960}{11\,160} = 0,086 \implies 8,6\%.$$

We see again that the proportion of people who have the disease among those who tested positive is 8,6%.

In this calculation, we simplified by working with whole numbers and rounding to the nearest ten. Thus, we have sacrificed a little precision for simplicity. In addition, we assumed that everything would turn out as probability predicts, i.e., that exactly 1 000 people have celiac disease. This is not always true. Although this method is simpler and often easier to understand, it is not typically found in textbooks. However, with this exercise statement, the success rate of doctors in estimating the result is almost 100%.

Literature

1. Singh P., Arora A., Strand T.A., Leffler D.A., Catassi C., Green P.H., Kelly C.P., Ahuja V., Makharia G.K., *Global Prevalence of Celiac Disease: Systematic Review and Meta-analysis*. Clin Gastroenterol Hepatol. 2018 Jun; 16(6):823-836. doi: 10.1016/j.cgh.2017.06.037.
2. Gigerenzer, G., *Calculated risks*, Simon and Schuster, 2003.