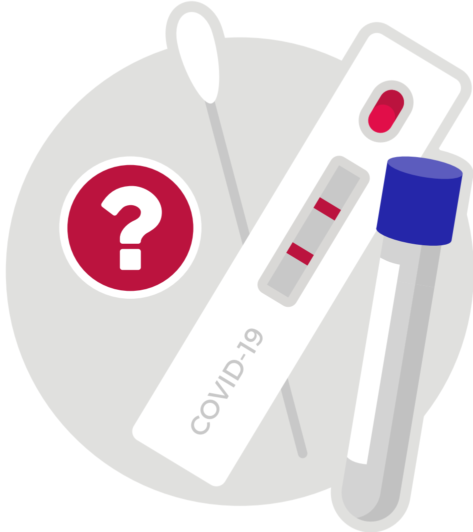
Examples of Covid-19 infection testing for high school students



STATISTICS UNVEILED...

In the time of Covid-19, knowledge of statistics is not only important to epidemiologists, economists, doctors, and politicians, but to everyone. In fact, misunderstanding statistics can pose risks to our lives and the lives of people around us.

Imagine that you take a Covid-19 rapid-test. Such a test is not 100% reliable, and if you can't interpret the result due to your lack of understanding of statistics, this can have drastic consequences. This is because it can easily happen that a rapid-test can show that you are negative for the virus when you are actually infected. If that happens, you could become a super-spreader without having the slightest idea that you are risking the lives of people around you.

COVID-19 RAPID TESTS

As the name suggests, rapid tests are cheap and simple in comparison with PCR tests, and they produce results while one waits.[1] If they become available over-the-counter at some point, one will easily be able to take a test at home. Why is the sale of rapid tests currently restricted, then? The problem is that they are not 100% reliable, and even a small chance of error can result in completely misleading conclusions. Thus, there are concerns that many people won't be able to interpret their test results correctly. Will you know how to interpret yours?

What is test error? When the result of a 90% reliable test is positive, does that mean that I'm infected with 90% probability? Absolutely not! This is similar to the situation when you observe something resembling an UFO in the sky. Even if it seems convincing, it is much more likely that you are experiencing a glitch in your eyesight than an alien visit. Likewise, a positive result from an inexact test is much more likely to be caused by test error than by an actual infection. The real situation is slightly more complicated as there are two types of test errors.[2]

1 More about Covid-19 testing: https://en.wikipedia.org/wiki/COVID-19_testing
2 More about the two types of errors: https://en.wikipedia.org/wiki/Sensitivity_and_specificity

- **The probability of positive test results in case of an infected individual** is called ***test sensitivity*** and is denoted s_p. It tells us what percentage of infected people will be detected by the test. Low test sensitivity is a serious problem. Not detecting a disease delays the necessary response. For example, in case of cancer, an early diagnosis can be crucial for successful treatment. For Covid-19, early detection is vital to prevent further spread of the infection. Getting an erroneously negative test result is worse than not being tested at all, because it can lead to less careful behaviour. If that happens, one can unwittingly infect parents, grandparents, classmates, and others.
- **The probability of negative test results in case of a healthy individual** is called test ***specificity*** and is denoted s_n. It is the share of healthy individuals that the test correctly detects. If the test fails to detect a healthy individual, it causes him/her unnecessary stress and more. In the case of Covid-19, it can cause one to be unnecessarily quarantined for weeks. It also distorts statistics reflecting the spread of the infection. In some cases, it might even prevent detection of other causes of health problems.

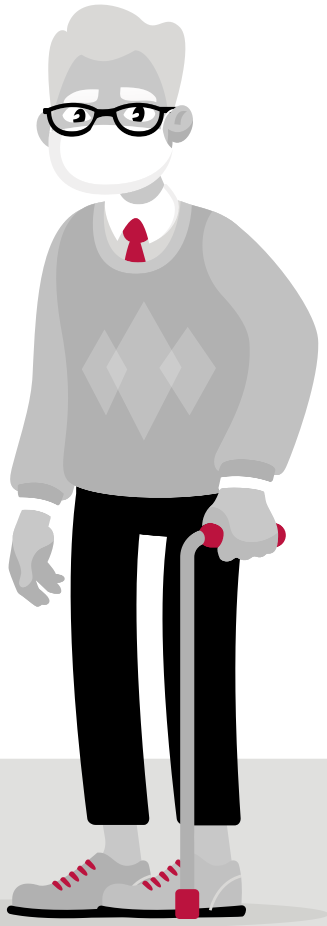
Further, we need to distinguish two types of infection probabilities:

- **The probability of having the infection before being tested** is called the ***prior probability*** and is denoted p. This is the probability at which one is likely to be infected before taking the test. This probability is much smaller for a random individual with no known risk factors than in the case of someone who has symptoms, or who shared space (household, classroom, bus, etc.) with someone who was found to be infected.
- **The probability of having the infection after being tested** is called the ***posterior probability*** and is denoted p'. This reflects not only how many opportunities one had to get infected, but also the actual test results and the reliability of the test used. This probability is the key information for doctors interpreting test results.

This is what you need to know to be able to calculate it all using the online calculator[3]. As an input, you set the parameters p, s_p, and s_n. The first parameter is usually chosen intuitively. The other two can be found in the info leaflet included with the test. The calculator allows you to look at numerous situations. We recommend starting by exploring examples with 90% sensitivity and specificity, that is $s_p = s_n = 90\% = 0.9$.

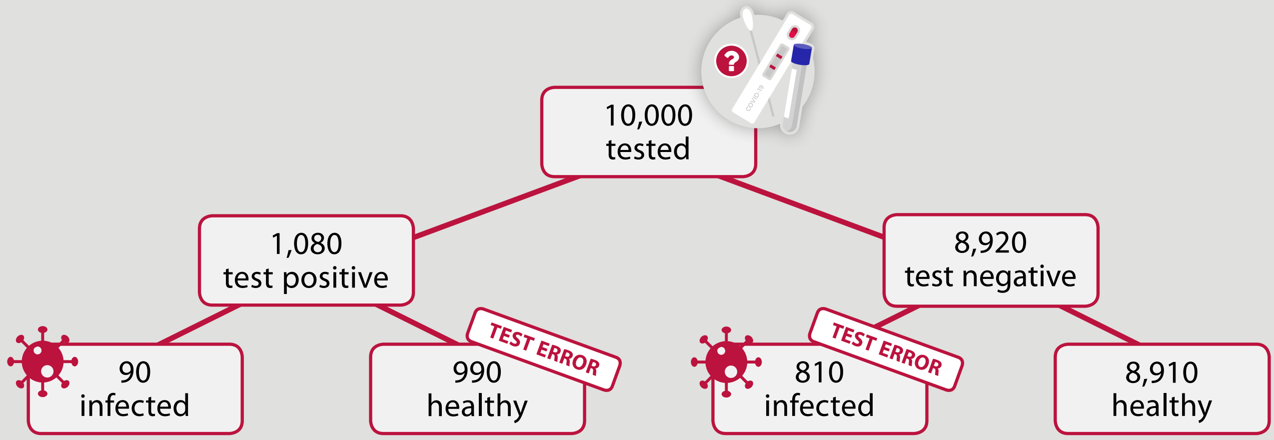
3 Bayessian calculator for Covid-19 https://ideaapps.cerge-ei.cz/anti_covid/?lang=en

THE CASE OF A CAUTIOUS SENIOR

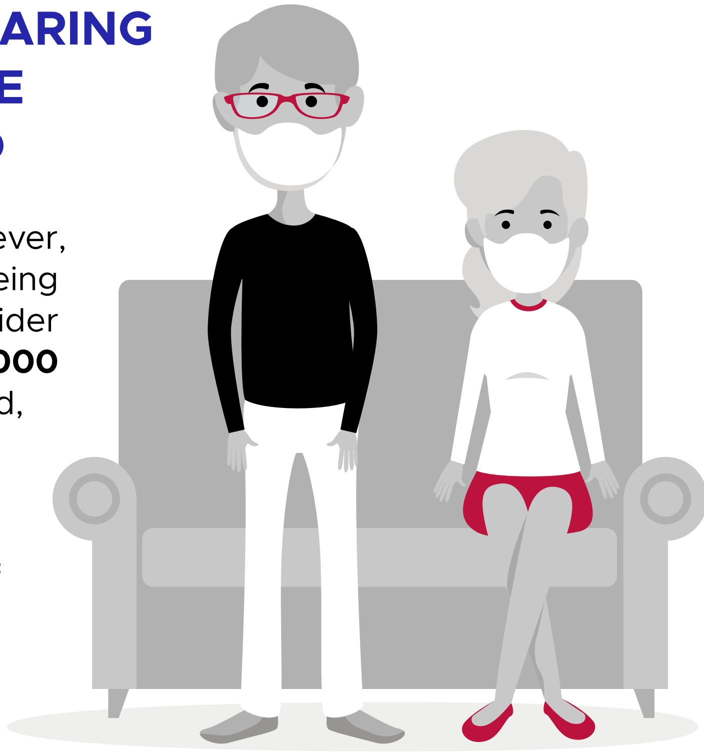


The senior rarely leaves home, but he takes a rapid test and tests positive. In his case, the probability of infection was minimal, say p = 1%. Let us consider 10,000 such seniors. Out of them (on average) 10,000p = **100** are infected, and 10,000(1 – p) = **9,900** are healthy. Out of those who are infected 100s_p = 100 • 0.9 = **90** test positive. At the same time, 9,900(1 – s_n) = 9,900 • 0.1 = **990** of the healthy ones also test positive. In total, 1,080 seniors test positive, but only 90 of them are actually infected. This means that for each senior who tests positive the probability of being infected is only $p' = \frac{90}{1080} \cdot 100\%$ = **8.3%**.

ILLUSTRATIVE APPLICATION OUTPUT FOR THIS EXAMPLE

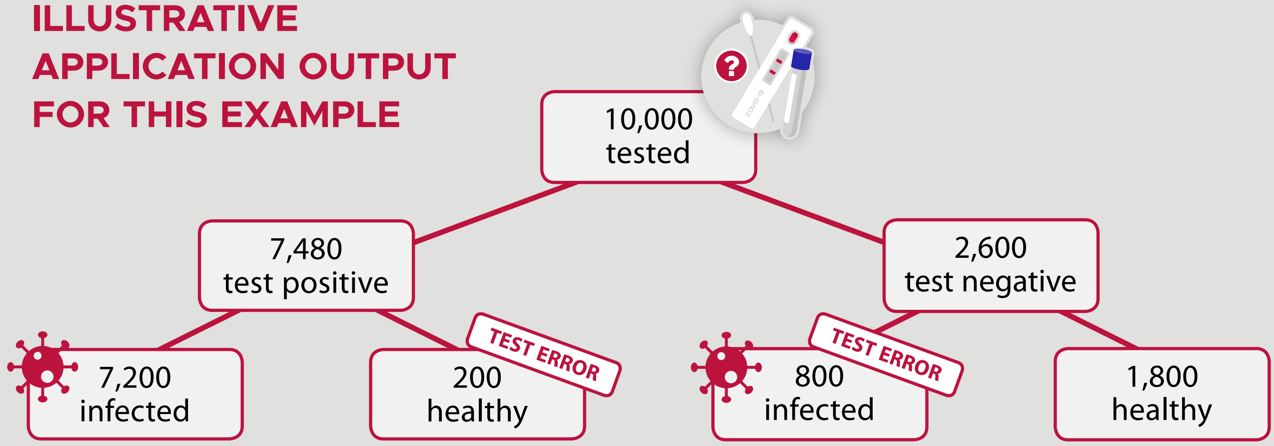


THE CASE OF A HUSBAND SHARING A HOUSEHOLD WITH HIS WIFE WHO IS PROBABLY INFECTED



Imagine that the husband tests negative. However, before getting tested, his probability of being infected is high, say p = 80%. Let us consider 10,000 similar households. In 10,000p = **8,000** of these households the husband is infected, and in the remaining 10,000(1 – p) = 2,000 of the households the husband is healthy. Out of the healthy husbands 2,000 • s_n = **1,800** test negative, but also 8,000(1 – s_p) = 8,000 • 0.1 = **800** of the infected husbands test negative. It follows that in spite of testing negative, the probability that the husband is infected is $p' = \frac{80}{1800 + 800} \cdot 100\%$ = **30.8%**. That is quite high and so he should stay in quarantine.

ILLUSTRATIVE APPLICATION OUTPUT FOR THIS EXAMPLE



The calculation we have used rely on so called **Bayes theorem**.[4] The calculation itself is not complicated, but for your convenience the online calculator does it for you. The application allows you to explore how to correctly interpret test results based on a variety of circumstances. If you understand that well, nobody should be afraid to sell a rapid test to you specifically.

IDEA CERGE EI
CONTENT
Pavel Kocourek
PROGRAMMING
Taras Hrendash, Jiří Münich
PRODUCTION
Blanka Javorová, Markéta Malá

4 More about the Bayes' theorem: https://en.wikipedia.org/wiki/Bayes%27_theorem