Covid-19-false-positive

Maike Heinrich

24 5 2021

Probability of getting a false positive Covid-19 PCR-test

Getting a positive result in a Covid-19 PCR-test doesn't always mean to actually have the disease. While the probability of being actually healthy when getting a negative test-result is almost 100%, the probability of getting a false positive test is around 42%. This means that almost half of all people, who are tested positive on Covid-19 are actually healthy. This can be explained by Bayes' Rule, which says that one probability affects another. In this case, the probability of getting the disease, which is the incidence and the accuracy of the test. A PCR test has a specificity of around 99.9%, which means the rate for correctly detecting negative results, and a sensitivity of around 90% - the detection rate for positive results. According to experts, the error rate is nevertheless quite small as many laboratories repeat positive tests.

Probability of getting a false positive Covid-19 self test

A quite reliable result is not the case for quick- and self tests. Although their accuracy has to be at least 97% in detecting correctly negative results, according to specifications of the World Health Organization, the false positive rate is exorbitant higher. While a negative test is reliable in 97% of the cases, the probability of getting a positive result and actually being healthy is about 96%. This illustrates the necessity of repeating a positive test by a PCR test in a laboratory.

The following codes show, how this probability is calculated and name the range of the variation of probabilities to get a false positive Covid-19 test.

In this example the current incidence is set on 150 and the specificity of the PCR-test is set on 99.9%, the specificity of a self test is set on 97%

```r
#Covid-19-Incidence of 150/100000 citizens
prev<-0.00150
#number of people tested
N<-100000

#probability of getting Covid-19 by Incidence
outcome<-sample(c("Corona","Healthy"),N,replace=TRUE, prob=c(prev,1-prev))
N_C<-sum(outcome==("Corona"))
N_H<-sum(outcome==("Healthy"))

#Accuracy of a Covid-19 Test (99,9%)
specitivity<-0.999
sensitivity<-0.9

#Calculating the test-outcome with Bayes' Rule
test<-vector("character",N)
```

```
test[outcome=="Corona"]<-sample(c("+","-"),N_C,replace=TRUE,prob=c(sensitivity,1-sensitivity))
test[outcome=="Healthy"]<-sample(c("-","+"),N_H,replace=TRUE,prob=c(specitivity,1-specitivity))
```

Creating a tabel for the results where "-" means a negative test, "+" stands for a positive test and "Corona" and "Healthy" for the actual result.

```
#Creating a table for the outcome
result<-table(outcome,test)
result
```

```
##          test
## outcome      -      +
##    Corona    17    135
##    Healthy 99751    97
```

People who will get a false positive result (This is a random number based on probability and will vary every time this code will be run)

```
#Number of people per 100000, getting a false positive result
result[2,2]
```

```
## [1] 97
```

Probability of being healthy although the test was positive

```
#Probability in percent of being healthy with getting a positive test
(result[2,2]/(result[2,2]+result[1,2]))*100
```

```
## [1] 41.81034
```

Because this value is based on probability it is random and varies each time, the code is run. To see the range of this variation a Monte Carlo simulation is made to be able to calculate the 95% confidence intervals - the two probabilities between which 95% of all calculated random probabilities will be found.

```
#Running a Monte Carlo simulation to calculate the 95% confidence interval of this probability
#samplesize for Monte Carlo simulation
B<-10000
MC_prob<-replicate(B,{MC_test<-vector("character",N)
MC_test[outcome=="Corona"]<-sample(c("+","-"),N_C,replace=TRUE,prob=c(sensitivity,1-sensitivity))
MC_test[outcome=="Healthy"]<-sample(c("-","+"),N_H,replace=TRUE,prob=c(specitivity,1-specitivity))
MC_result<-table(outcome,MC_test)
(MC_result[2,2]/(MC_result[2,2]+MC_result[1,2]))*100
})
mean(MC_prob)
```

```
## [1] 42.09736
```

Now the confidence interval can be calculated

```
#Calculation for 95% confidence interval
N_se<-1000 #samplesize for standard error
X_hat<-mean(MC_prob)/100 #expected value
SE_hat<-sqrt(X_hat*(1-X_hat)/N_se) #standard error
ci<-c(lower<-X_hat-qnorm(0.975)*SE_hat,upper<-X_hat+qnorm(0.975)*SE_hat)
ci
```

```
## [1] 0.3903734 0.4515738
```

In between this range, 95% of the calculated probabilities of being healthy although being tested positiv for Covid-19 can be found.

```
ci[1]*100
```

```
## [1] 39.03734
```

```
ci[2]*100
```

```
## [1] 45.15738
```

Now to illustrate the huge difference to the false positive results of a self test the codes below show the calculation for a test with the specificity of 97% to correctly detect negative results and the sensitivity of 80% for correct positives.

```
specificity_st<-0.97
sensitivity_st<-0.8
test_st<-vector("character",N)
test_st[outcome=="Corona"]<-sample(c("+","-"),N_C,replace=TRUE,prob=c(sensitivity_st,1-sensitivity_st))
test_st[outcome=="Healthy"]<-sample(c("-","+"),N_H,replace=TRUE,prob=c(specificity_st,1-specificity_st))
```

Table of the results of a self test

```
result_st<-table(outcome,test_st)
result_st
```

```
##          test_st
## outcome       -     +
##    Corona     35   117
##    Healthy 96911  2937
```

Probability for a false positive self test

```
prob_healthy_st<-(result_st[2,2]/(result_st[2,2]+result_st[1,2]))*100
prob_healthy_st
```

```
## [1] 96.16896
```

Range of probabilities for getting a false positive self test

```r
MC_prob_st<-replicate(B,{MC_test_st<-vector("character",N)
MC_test_st[outcome=="Corona"]<-sample(c("+","-"),N_C,replace=TRUE,prob=c(sensitivity_st,1-sensitivity_st
MC_test_st[outcome=="Healthy"]<-sample(c("-","+"),N_H,replace=TRUE,prob=c(specificity_st,1-specificity_s
MC_result_st<-table(outcome,MC_test_st)
(MC_result_st[2,2]/(MC_result_st[2,2]+MC_result_st[1,2]))*100})
mean(MC_prob_st)
```

```
## [1] 96.09981
```

```r
#confidence interval
X_hat_st<-mean(MC_prob_st)/100 #expected value
SE_hat_st<-sqrt(X_hat_st*(1-X_hat_st)/N_se) #standard error
ci_st<-c(lower_st<-X_hat_st-qnorm(0.975)*SE_hat_st,upper_st<-X_hat_st+qnorm(0.975)*SE_hat_st)
ci_st
```

```
## [1] 0.9489989 0.9729973
```

The 95% range of probabilities of getting a false positive Covid-19 selftest is between these two percentages.
So repeating the test is recommended.

```r
ci_st[1]*100
```

```
## [1] 94.89989
```

```r
ci_st[2]*100
```

```
## [1] 97.29973
```