Bayes' rule applied on Covid 19 Immunity testing

Author: Micha Bouts, Oehoe Data Science BV, 2020©

Date: 2020-06-04

Intention

Having experienced unconfirmed Covid-19 symptoms of fever, dry cough, pain in throat and obstructed sinus in early March 2020 and not having known whether it truly was Covid-19 or a classic flu and with an interest if meanwhile immunity was gathered, I went on June 2nd 2020 for a serological antibody test. The serological test result turned out negative for SARS-CoV-2 antibodies. This made me think about the probability for truly not having had Covid-19 or not being immune, while having tested negative. While running the underlying exercise I came to realize there is a difference between on the one hand testing for SARS-CoV-2, the virus causing Covid-19, and on the other hand testing for immunity. The first one is done by a Polymerization Chain Reaction (PCR) searching for viral RNA, while the second type of test searches for antibodies. Since I haven't received a PCR test at time of illness, I can only build on a serological test after recovery from past symptoms. The exercise at hand is to use conditional probability to evaluate the chance of (not) being immune. And to learn if anything can be concluded on whether or not Covid-19 was contracted three months ago.

Conditional probability

(source: Conditional Probability, Statistical Inference, Brian Caffo, Jeff Leek, Roger Peng, Johns Hopkins Bloomberg School of Public Health).

Conditional probability is used to calculate the probability of an event based on new information which arrived to the scene.

The conditional probability of an event A, given that B happened before is

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

With Bayes' rule we can reverse the order of conditional probability as follows:

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}$$

Event A and B can both have a positive or negative outcome, arbitrarily chosen what is positive or negative.

Negative Predictive Value (NPV)

The negative predictive value is the probability for not having Immunity for Covid-19 with a negative serological test result.

Following Bayes' rule, this is calculated by

- Event A = negative serological test result = *neg test*
- Event B = truly not having immunity = *Immun neg*

Hence,

$$P(Immun\ neg\ |\ neg\ test)$$
$$= \frac{P(neg\ test\ |\ Immun\ neg)P(Immun\ neg)}{P(neg\ test\ |\ Immun\ neg)P(Immun\ neg) + P(neg\ test\ |\ Immun\ posit)P(Immun\ posit)}$$

The serological prevalence *P*(*Immun posit*) is estimated for Belgium as an intermediate value available for blood donors and healthcare professionals. On May 13th 2020 the available data pointed to 4.7% serological prevalence for blood donors and 8.4% for healthcare professionals respectively (source: https://covid-19.sciensano.be/sites/default/files/Covid19/COVID-19_Weekly%20report_20200529%20-%20NL_0.pdf).

This was averaged to 6.0% (source: https://www.vrt.be/vrtnws/en/2020/05/08/just-6-of-belgians-have-antibodies-against-covid-19/).

Important to notice is that serological prevalence is not the same as disease prevalence. The latter is calculated by the number of confirmed positive cased divided by the population of a country.

For Belgium with an estimated population of 11,492,641 (source: https://en.wikipedia.org/wiki/Belgium, status January 1st 2020 census). The number of confirmed cases is 58,615 as reported on June 2nd 2020 (source: https://epistat.wiv-isp.be/covid/covid-19.html).

Consequently, the serological prevalence is *P*(*Immun posit*) $\approx 0.06 \approx 6.0\%$

While the disease prevalence is

$$P(Covid19\ posit) = \frac{58,615}{11,492,641} \approx 0.0051 \approx 0.5\%$$

There is an order of magnitude difference!

Moving forward with the serological prevalence we get the negative serological prevalence by

$$P(Immun\ neg) = 1 - P(Immun\ posit) \approx 0.94 \approx 94\%$$

Next, we define specificity as the probability that a healthy person will test negative on immunity. In this case, the specificity is determined as a best guess average from a variety of serologic test methods (source: https://www.centerforhealthsecurity.org/resources/COVID-19/serology/Serology-based-tests-for-COVID-19.html)

$$Specificity = P(neg\ test\ |\ Immun\ neg) \approx 0.98$$

Furthermore, the sensitivity is defined as the probability that the immunity test is positive, given that the person actually has immunity for Covid-19. Also for assessing this conditional probability, we made a best guess average (source: https://www.centerforhealthsecurity.org/resources/COVID-19/serology/Serology-based-tests-for-COVID-19.html)

$$Sensitivity = P(posit\ test\ |\ Immun\ posit) \approx 0.95$$

With the sensitivity at hand, we can calculate

$$P(neg\ test\ |\ Immun\ posit) = 1 - Sensitivity$$

Plugging all this into Bayes' rule for the negative predictive value gives us

$$P(Immun\ neg\ |\ neg\ test)$$
$$= \frac{Specificity * P(Immun\ neg)}{Specificity * P(Immun\ neg) + (1 - Sensitivity) * P(Immun\ posit)}$$

Finally,

$$P(Immun\ neg\ |\ neg\ test) = \frac{0.98 * 0.94}{0.98 * 0.94 + (1 - 0.95) * 0.06} \approx 0.997 \approx 99.7\%$$

Likewise, one can calculate the probability for having immunity for Covid-19, under the precondition of a negative serologic test.

$$P(Immun\ posit\ |\ neg\ test)$$
$$= \frac{(1 - Sensitivity) * P(Immun\ posit)}{(1 - Sensitivity) * P(Immun\ posit) + Specificity * P(Immun\ neg)}$$

Which gives us

$$P(immun\ posit\ |\ neg\ test) = \frac{(1 - 0.95) * 0.06}{(1 - 0.95) * 0.06 + 0.98 * 0.94} \approx 0.0003 \approx 0.03\%$$

The latter could also more straightforward be calculated as

$$1 - P(Immun\ neg|\ neg\ test)$$

As a conclusion one can say that having had a negative serological test, the probability is extremely high (99.7%) that I'm truly not-immune. As my GP rightfully warned me for, I'm still susceptible for the virus and should stay vigilant and not live with a false impression of security.

On the other hand, what this test doesn't say anything about is whether or not I truly contracted the SARS-CoV-V2 early March. This serological test and Bayesian probability doesn't lead to any insight as such. It might be that an infection truly happened with only mild symptoms and the serological test three months later doesn't offer a conclusion on this.

Positive Predictive Value (PPV)

However, imagine I would have tested positive on the serological test? This would be of more interest for tracking immunity levels on individual and population level. Applying the same Bayes' rule we can now calculate the positive predictive value, as the probability for truly being immune for Covid-19, after having tested positive on a serological immunity test. Different from the NPV assessment above,

when a person is considered truly immune one may assume from a logical clinical consequence that the individual indeed contracted SARS-CoV-V2 beforehand.

$$P(Immun\ posit \mid pos\ test)$$
$$= \frac{P(pos\ test \mid Immun\ posit)P(Immun\ pos)}{P(pos\ test \mid Immun\ posit)P(Immun\ posit) + P(pos\ test \mid Immun\ neg)P(Immun\ neg)}$$

Substituting some terms,

$$P(Immun\ posit \mid pos\ test)$$
$$= \frac{P(pos\ test \mid Immun\ posit)P(Immun\ pos)}{P(pos\ test \mid Immun\ posit)P(Immun\ posit) + \{1 - P(neg\ test \mid Immun\ neg)\}P(Immun\ neg)}$$

Entering the data,

$$P(Immun\ posit \mid pos\ test) = \frac{0.95 * 0.06}{0.95 * 0.06 + \{1 - 0.98\} * 0.94} \approx 0.752 \approx 75.2\%$$

Hence, the predictive value for truly being immune after having tested positive on immunity is reasonable yet not as strong as for the NPV. In order to achieve a higher degree of confidence a subsequent second test may be meaningful. Information gained from the first test will act as input for the second test and the conditional probability will further increase to 99.3%. This is called a 'double chained test'.
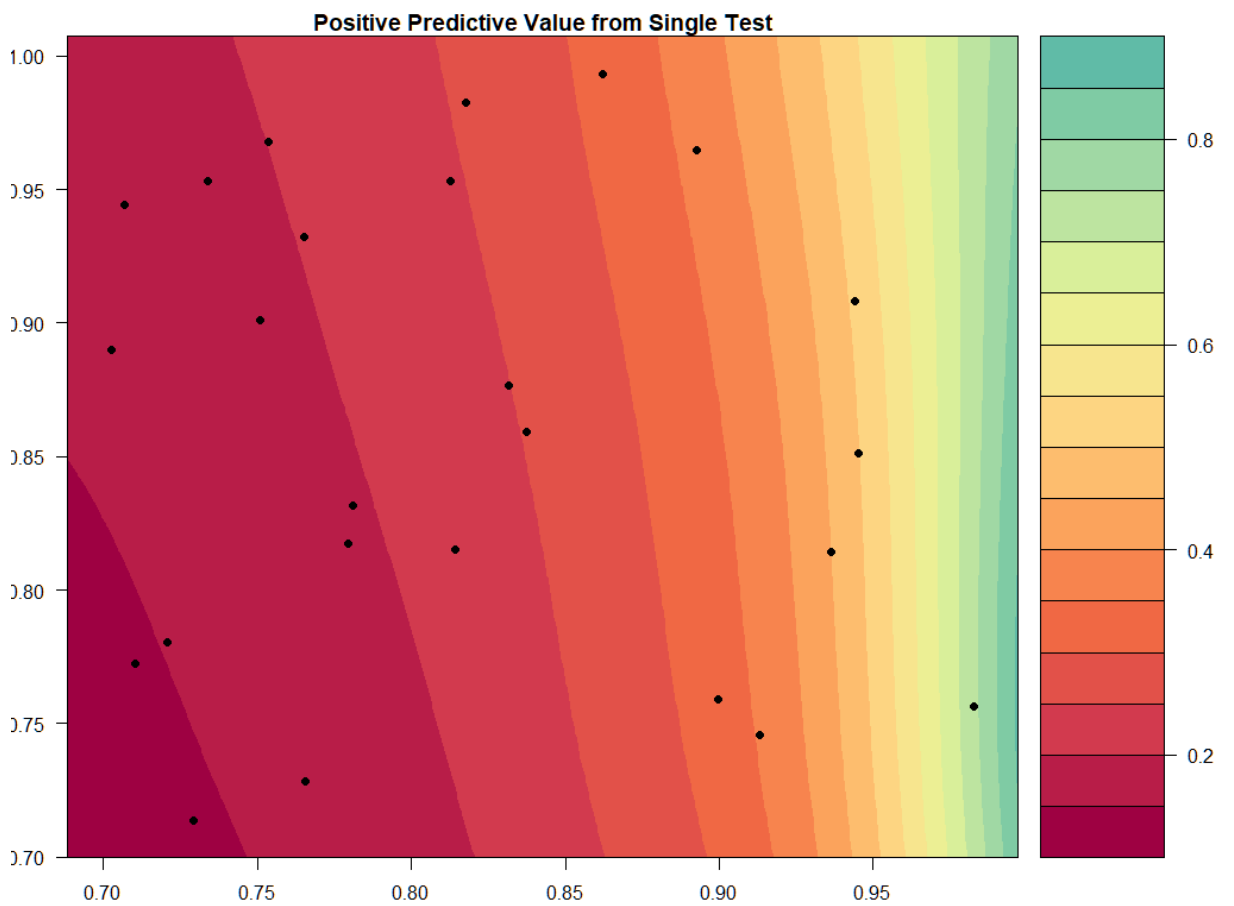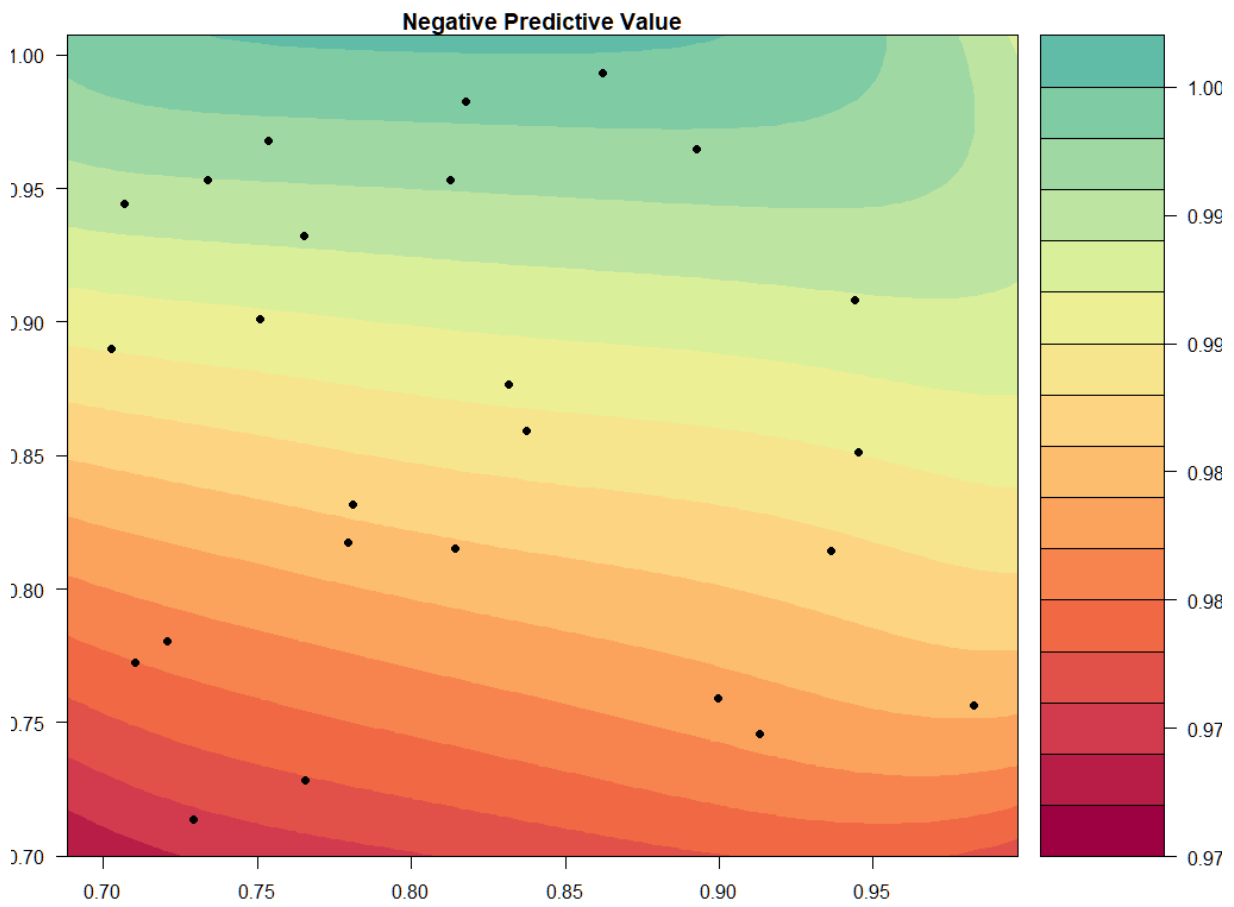
$$P(Immun\ posit \mid pos\ test\ 1 + pos\ test\ 2) = \frac{0.95 * 0.95 * 0.06}{0.95 * 0.95 * 0.06 + \{1 - 0.98\} * (1 - 0.98) * 0.94}$$
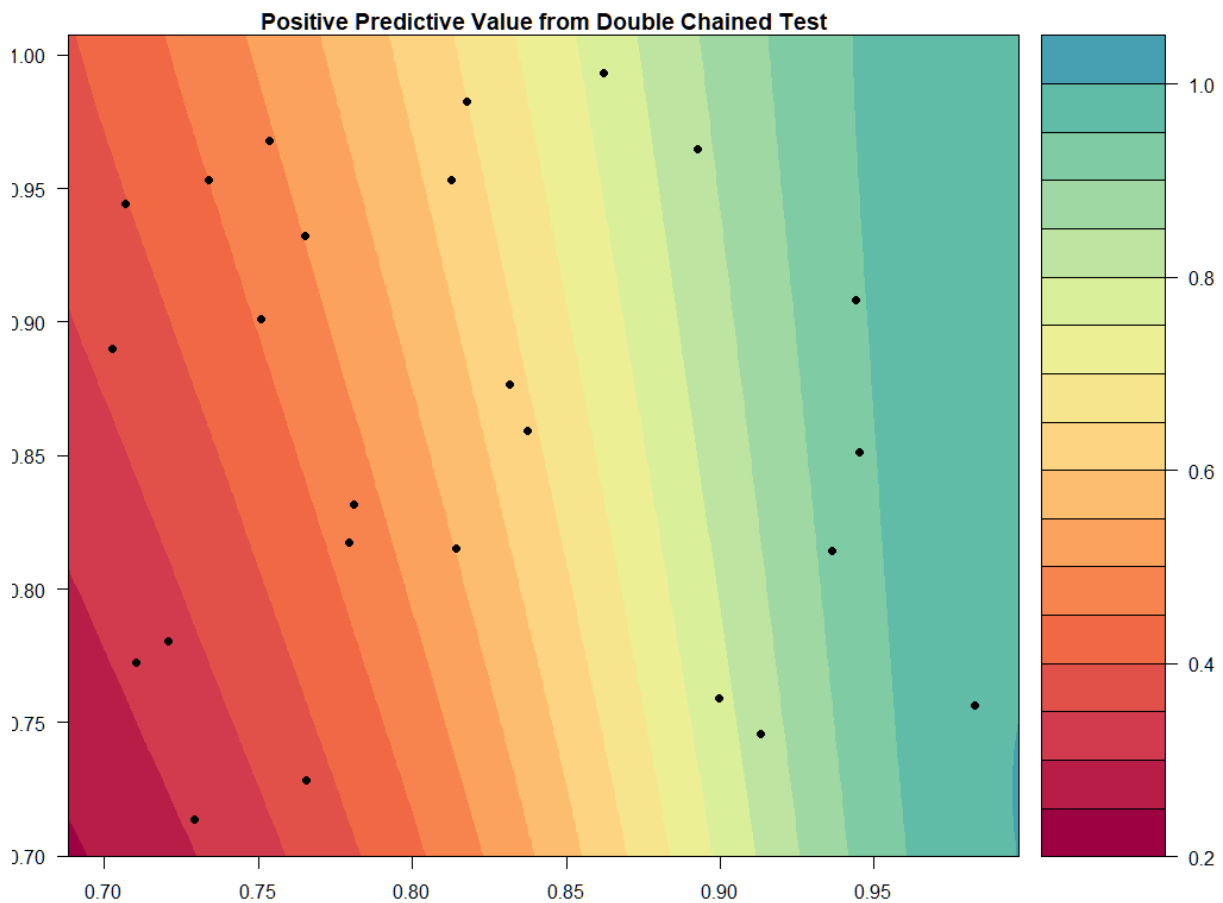$$\approx 0.993 \approx 99.3\%$$

Discussion

The concept of immunity as something different than viral infection may easily confuse people when reading out and understanding the outcome of a serological immunity test, different from a viral RNA PCR test.

As pointed out in the abovementioned calculations, the NPV turns out very high (99.7%) as probability for not-being immune, after having tested negative. This does say a lot of current non-immunity. Yet, it doesn't give any conclusive insight on past SARS-CoV-V2 infection. Only in case of a double sequential positive serological test outcome, one can state with high confidence (99.3%) that the individual was previously infected with SARS-CoV-V2 for which immunity is built up by now.

Furthermore, attention should be paid to the assumptions on specificity and sensitivity. Certainly sensitivity can vary dependent on the type of serological test method. We took an average, yet a sensitivity analysis on both variables might be of interest to understand the impact on NPV and/or PPV results. If we include test method induced variance on specificity and sensitivity then by making use of a Contour Plot we get a graphical view of the effect on the Negative or Positive Predicted Value. The charts below show on the x – axis the Specificity, on the y – axis the Sensitivity and the colour gradient gives the probability for the Negative (NPV) or Positive Predicted Value (PPV). As we see, NPV is more correlated with Sensitivity, while PPV is more dependent on Specificity.

Negative Predictive Value

Positive Predictive Value from Single Test

Positive Predictive Value from Double Chained Test

Computing Tool and Script

We made use of The R Project for Statistical Computing, R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

```
R version 3.5.1 (2018-07-02)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows >= 8 x64 (build 9200)

Matrix products: default

locale:
[1] LC_COLLATE=Dutch_Belgium.1252  LC_CTYPE=Dutch_Belgium.1252     LC_MONETA
RY=Dutch_Belgium.1252
[4] LC_NUMERIC=C                   LC_TIME=Dutch_Belgium.1252

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] plotly_4.8.0           ggplot2_3.1.1          RColorBrewer_1.1-2     Co
ntourFunctions_0.1.0
[5] RevoUtils_11.0.1       RevoUtilsMath_11.0.0

loaded via a namespace (and not attached):
 [1] Rcpp_1.0.3          later_0.7.4         pillar_1.4.3        compiler_3.5.1
plyr_1.8.4
 [6] tools_3.5.1         rpart_4.1-13        digest_0.6.17       viridisLite_0.3.
0 jsonlite_1.5
```

```
[11] tibble_2.1.3        gtable_0.2.0        pkgconfig_2.0.2    rlang_0.4.4
shiny_1.1.0
[16] rstudioapi_0.7      crosstalk_1.0.0    yaml_2.2.0         parallel_3.5.1
tgp_2.4-14
[21] httr_1.3.1          withr_2.1.2        dplyr_0.8.4        cluster_2.0.7-1
htmlwidgets_1.2
[26] grid_3.5.1          tidyselect_1.0.0   maptree_1.4-7      glue_1.3.1
data.table_1.11.4
[31] R6_2.2.2            laGP_1.5-2         tidyr_0.8.1        purrr_0.3.3
magrittr_1.5
[36] promises_1.0.1      scales_1.0.0       htmltools_0.3.6   assertthat_0.2.0
xtable_1.8-3
[41] mime_0.5            colorspace_1.3-2   httpuv_1.4.5       lazyeval_0.2.1
munsell_0.5.0
[46] crayon_1.3.4
```

R Script

```r
1   setwd("~/R/Personal_DoE_MB/Covid_19")
2   remove(list = ls())
3
4   library(ContourFunctions) # also install package 'laGP'
5   library(RColorBrewer)
6
7   # Seroprevalence P(Immun+), status May 13th 2020
8
9   # Seroprevalence amongst blood doners = 4.7% https://covid-19.sciensano.be/sites/default/files/Covid19/COVID-19_Weekly%20report_20200529%20-%20NL_0.pdf
10  # Seroprevalence amongst healthcare professionals = 8.4% https://covid-19.sciensano.be/sites/default/files/Covid19/COVID-19_Weekly%20report_20200529%20-%20NL_0.pdf
11  # Average seroprevalence = 6% https://www.vrt.be/vrtnws/en/2020/05/08/just-6-of-belgians-have-antibodies-against-covid-19/
12
13  PImmunPlus <- 0.06
14
15  PImmunNeg <- 1 - PImmunPlus # negative prevalence = 0.94 = 94%
16
17
18  # Specificity P(- | Immun-) = probability that a healthy person will test negative
19  # https://www.centerforhealthsecurity.org/resources/COVID-19/serology/Serology-based-tests-for-COVID-19.html
20
21          PNegImmunNeg <- 0.98 # assumption as best guess from several sources Johns Hopkins Center for Health Security
22
23
24
25  # Sensitivity P(+ | Immun+) = probability that the test is positive, given that the person actually has built up immunity for Covid-19
26  # https://www.centerforhealthsecurity.org/resources/COVID-19/serology/Serology-based-tests-for-COVID-19.html
27
28          PPosImmunPos <- 0.95 # assumption as best guess from several sources Johns Hopkins Center for Health Security
29
30
31  # P(Immun- | -) = probability for not having Covid 19 Immunity when tested negative
32
33
34  PImmunNegNeg <- PNegImmunNeg * PImmunNeg / ((PNegImmunNeg * PImmunNeg) + ((1 - PPosImmunPos) * PImmunPlus))
35          # so, when tested negative the probability that you truly didn't have Covid 19 Immunity was 99.7%
36
37  |
38  # P(Immun+ | -) = probability for having Covid 19 Immunity when tested negative
39
40  PImmunPosNeg <- ((1 - PPosImmunPos) * PImmunPlus) / (((1 - PPosImmunPos) * PImmunPlus) + PNegImmunNeg * PImmunNeg)
41          # so, when tested negative, the probability that you had Covid 19 Immunity was only 0.3%
42          # This is the same as 1 - P(Immun- | -)
43
44
45  # P(Immun+ | +) = probability for having Covid 19 Immunity when tested positive
46
47  PImmunPosPos <- PPosImmunPos * PImmunPlus / ((PPosImmunPos * PImmunPlus) + (1 - PNegImmunNeg) * PImmunNeg)
48          # so, when tested positive, it suggest a 75,2% probability that the patient truly had Covid 19 Immunity.
49          # This is a pretty high outcome. It could still be increased by running a subsequent second Baysian test.
50
51  PImmunPosPos12 <- PPosImmunPos * PPosImmunPos * PImmunPlus / ((PPosImmunPos * PPosImmunPos* PImmunPlus) + (1 - PNegImmunNeg) * (1 - PNegImmunNeg) * PImmunNeg)
52          # After having execute two tests, we reach a positive predictive probability of 99.3% which is extremely high.
53
54  ## Prepare for a sensitivity analysis of impact from specificity and sensitvity on conditional probability as outcome.
55
56  set.seed(12)
57
58  specificity <- runif(25, min = 0.70)
59  sensitivity <- runif(25, min = 0.70)
60
61  cols <- brewer.pal(10, "Spectral")
62  cols
63  pal <- colorRampPalette(cols)
64
65
66  ## Contourplot for the Negative Predictive Value (NPV)
67
68  NPV <- specificity * PImmunNeg / ((specificity * PImmunNeg) + ((1 - sensitivity) * PImmunPlus))
69
70  # cf_data(specificity, sensitivity, NPV)
71
72  cf_data(specificity,sensitivity,NPV, bar=T, col = pal(20), main = "Negative Predictive Value")
73
74  ## Contourplot for the Positive Predictive Value (PPV) single test
75
76  PPV1 <- sensitivity * PImmunPlus / ((sensitivity * PImmunPlus) + (1 - specificity) * PImmunNeg)
77
78  # cf_data(specificity, sensitivity, PPV1)
79
80  cf_data(specificity,sensitivity,PPV1, bar=T, col = pal(20), main = "Positive Predictive Value from Single Test")
81
82  ## Contourplot for the double tested Positivie Predictive Value (PPV12)
83
84  PPV12 <- sensitivity * sensitivity * PImmunPlus / ((sensitivity * sensitivity* PImmunPlus) + (1 - specificity) * (1 - specificity) * PImmunNeg)
85
86  # cf_data(specificity, sensitivity, PPV12)
87
88  cf_data(specificity,sensitivity,PPV12, bar=T, col = pal(20), main = "Positive Predictive Value from Double Chained Test")
89
```