How to Determine When SARS-CoV-2 Antibody Testing Is or Is Not Useful for Population Screening: A Tutorial

Niklas Keller⬤, and **Mirjam A. Jenny**

Abstract

Extensive testing lies at the heart of any strategy to effectively combat the SARS-COV-2 pandemic. In recent months, the use of enzyme-linked immunosorbent assay–based antibody tests has gained a lot of attention. These tests can potentially be used to assess SARS-COV-2 immunity status in individuals (e.g., essential health care personnel). They can also be used as a screening tool to identify people that had COVID-19 asymptomatically, thus getting a better estimate of the true spread of the disease, gain important insights on disease severity, and to better evaluate the effectiveness of policy measures implemented to combat the pandemic. But the usefulness of these tests depends not only on the quality of the test but also, critically, on how far disease has already spread in the population. For example, when only very few people in a population are infected, a positive test result has a high chance of being a false positive. As a consequence, the spread of the disease in a population as well as individuals' immunity status may be systematically misinterpreted. SARS-COV-2 infection rates vary greatly across both time and space. In many places, the infection rates are very low but can quickly skyrocket when the virus spreads unchecked. Here, we present two tools, natural frequency trees and positive and negative predictive value graphs, that allow one to assess the usefulness of antibody testing for a specific context at a glance. These tools should be used to support individual doctor-patient consultation for assessing individual immunity status as well as to inform policy discussions on testing initiatives.

Keywords

diagnostic test interpretation, natural frequency trees, PPV/NPV graphs, SARS-CoV-2 antibody testing

ELISA-Based Antibody Tests Are Increasingly Administered

Extensive testing lies at the heart of any strategy to effectively combat the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. While the test for acute infection by use of polymerase chain reaction (PCR) has been at the forefront in the initial phases of the pandemic, in recent months, the use of enzyme-linked immunosorbent assay (ELISA)-based antibody tests has gained a lot of attention with two primary goals in mind: to identify persons who may be immune to the disease, particularly those working in critical areas of infrastructure (e.g., essential healthcare personnel),* and to assess the number of people that have been infected in a particular population including all those who have perhaps gone through the infection with no or only mild symptoms.

---

*Note that we currently do not know whether people who have recovered from COVID-19 and have antibodies are protected from a second infection and, if so, for how long.

Corresponding Author
Niklas Keller, Simply Rational—The Decision Institute, Eberhard-Roters-Platz 7a, Berlin DE 10965, Germany;
Telephone: +49 176 2041 8755 (niklas.keller@simplyrational.de).

Knowing this would provide valuable insights on hospitalization and mortality rates, which is critical for estimating the burden of the pandemic on the health care system as well as assessing the effectiveness of policy measures designed to curb the spread of the virus. Subsequently, while PCR testing to identify acute infections and disrupt transmission chains remains central, many nations and institutions are in the process of rolling out antibody testing initiatives (e.g., China, the United States, Spain, Switzerland).[1-5]

The Interpretation of These Test Results Depends on the Specific Context

The attributes of a test (specifically, its sensitivity and specificity) directly determine how well it is suited to provide a reliable immunity assessment at the individual level as well as its usefulness for population-wide screening. The sensitivity (or true positive rate) of an antibody test describes its ability to correctly identify persons that have had the disease and now have antibodies against that disease in their blood serum. Thus, if a test has a 90% sensitivity, out of 100 persons that have had a SARS-CoV-2 infection, the test will correctly identify 90 of these, while 10 would be missed and receive a false negative test result. The specificity (or true negative rate) of a test describes its ability to correctly identify those that have not had the disease. Thus, if a test has a 90% specificity, out of 100 persons not having had the disease, the test will correctly identify 90 of them as negative. The other 10 persons will receive a false positive test result.

But the number of infected and noninfected persons is rarely the same. In most cases, the part of the population that has not had SARS-COV-2 will greatly outnumber the part of the population that has had it. Continuing with the above example of a test with a 90% sensitivity and 90% specificity, we can imagine a different population in which 100 have had the disease and 10,000 had not. As above, of those having had the disease, 90 would test true positive, 10 false negative. Of those 10,000 who were not infected, 9,000 would receive a true negative test result, but a full 1,000 would receive a false positive

test result—10 times more than the entire infected population. With regard to SARS-CoV-2, prevalence differs between regions and population subgroups and depends on the point in time of the pandemic. This variation means that interpretation and usefulness of antibody testing is context specific, an idea that is not intuitive to many.[6]

Test Interpretation During a Pandemic

Several ELISA-based antibody tests for SARS-CoV-2 are currently being developed and the test characteristics, that is, the sensitivity and specificity, vary. Ideally, only tests should be employed for which independent validation studies with a sufficient sample size of participants (at least 1000 + to achieve the required resolution) are available. Prof. Drosten, a leading expert on SARS-CoV-2 whose laboratory also developed the RT-PCR (real-time polymerase chain reaction) test for acute SARS-CoV-2 infection,[7] stated that he expects at least 2% false-positives (i.e., a 98% specificity) for ELISA-based antibody tests.[8] For the purpose of the argument made in this tutorial, let us assume that current ELISA-based procedures have a sensitivity of 80%, a value that has been shown for one particular assay in a recent publication,[9] and a specificity of 98%.

At the same time, the SARS-CoV-2 prevalence can vary greatly, both in time and in space. The R0 of SARS-CoV-2 is estimated by the World Health Organization to be between 2 and 2.5.[10] R0 is one measure used to quantify the contagiousness of a disease in the absence of any countermeasures (such as social distancing) or mitigating factors (such as herd immunity) and is an estimate of the number of additional persons one infected patient will infect. An R0 between 2 and 2.5 thus means that one person infected with SARS-CoV-2 will, over the course of their infection, infect between 2 and 2.5 additional people. Figure 1 shows the spread of SARS-CoV-2 in a hypothetical population of a million people in which no countermeasures have been implemented, the R0 is 2.2, and the number of initially infected persons is 100 (0.01% prevalence). If no countermeasures are implemented, and the R0 does not change for other reasons, the virus will have infected approximately 400,000 people after 6 weeks (~ 40% prevalence). We will now use this hypothetical population to assess the utility of antibody testing both from the perspective of the individual, as well as its usefulness for population wide testing initiatives, using natural frequency trees and predictive value graphs.
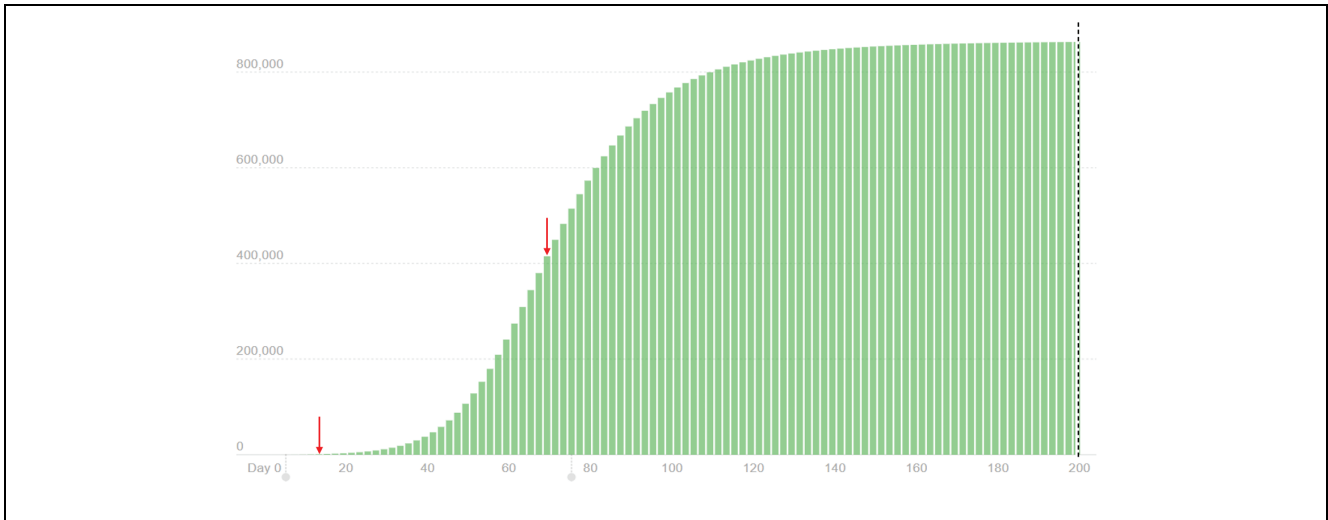
Figure 1 Number of people having undergone a SARS-CoV-2 infection in a hypothetical population of 1,000,000 persons given R0 = 2.2 and with no mitigating measures implemented. Green bars represent the number of persons having undergone a SARS-CoV-2 infection. Red arrows represent two hypothetical points in time for antibody testing during the pandemic.

Using Natural Frequency Trees to Assess the Usefulness of Antibody Testing for Individual Immunity Assessment: What Does a Positive Antibody Test Result Mean for the Person Tested?

Let us assume a person enters a clinic without having had any symptoms or other known risk factors increasing their likelihood of having had the disease and wishes to be tested for SARS-CoV-2 "immunity" via antibody testing at the beginning of the pandemic when the prevalence is 0.01%. Another person wishes to be tested 6 weeks later when the prevalence is ~40% (see red arrows in Figure 1). Both individuals test positive for SARS-CoV-2 antibodies. How certain can we be, that these individuals had the disease? That is, given a positive test result, what is the probability that the person really had the infection? This probability is what is called a "positive predictive value" and is calculated as the ratio of true positive test results to all positive test results, both true and false. Calculating such a "conditional probability" (i.e., the probability of having the disease given a positive test result) is difficult for doctors, patients, and policy makers alike.[11,12] But we can make the calculation more intuitive using *Natural Frequency Trees* (see Figure 2).[13]

Natural frequency trees (NFTs) graphically represent how a population (e.g., a million individuals) is sequentially divided into subpopulations (e.g., individuals affected or not affected by a disease). By using NFTs, one can quickly and intuitively assess the ratio of true positive test results to all (true and false) positive test results (red boxes in Figure 2), that is, the positive predictive value (PPV) of a diagnostic procedure. In the same fashion, the negative predictive value (NPV) can be arrived at by assessing the ratio of true negatives to all (true and false) negative test results (green boxes in Figure 2). The NPV is the value we are looking for when we wish to answer the question: "Given a negative test result, how likely is it, that that person really *has not* undergone a SARS-CoV-2 infection?"

Figure 2 shows that at the beginning of the pandemic, the vast majority of positively tested cases will be false positives. Only about 1 in 250 individuals tested positive in this phase will have tested positive due to an actual prior SARS-CoV-2 infection. The situation is very different for the second person being tested 6 weeks later (see Figure 3).

Six weeks later, the vast majority of positively tested cases will be true positives. In this phase, roughly 24 out of 25 positive test results will be due to a prior SARS-CoV-2 infection.

When Should Antibody Testing Be Used for Population-Based Screening?

NFTs present a simple and intuitive graphical format, which can be used to explain a test's predictive values to individuals both prior to testing ("does it make sense to get tested?") as well as afterwards ("what does the test
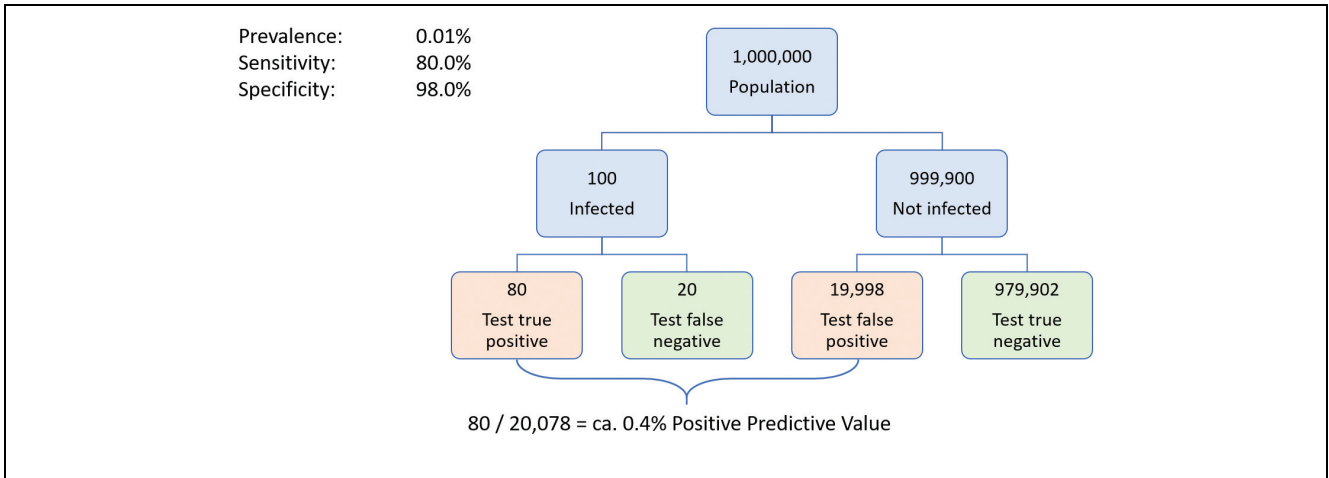
Figure 2 Natural frequency tree for antibody testing at the beginning of the pandemic with a SARS-CoV-2 prevalence of 0.01%.
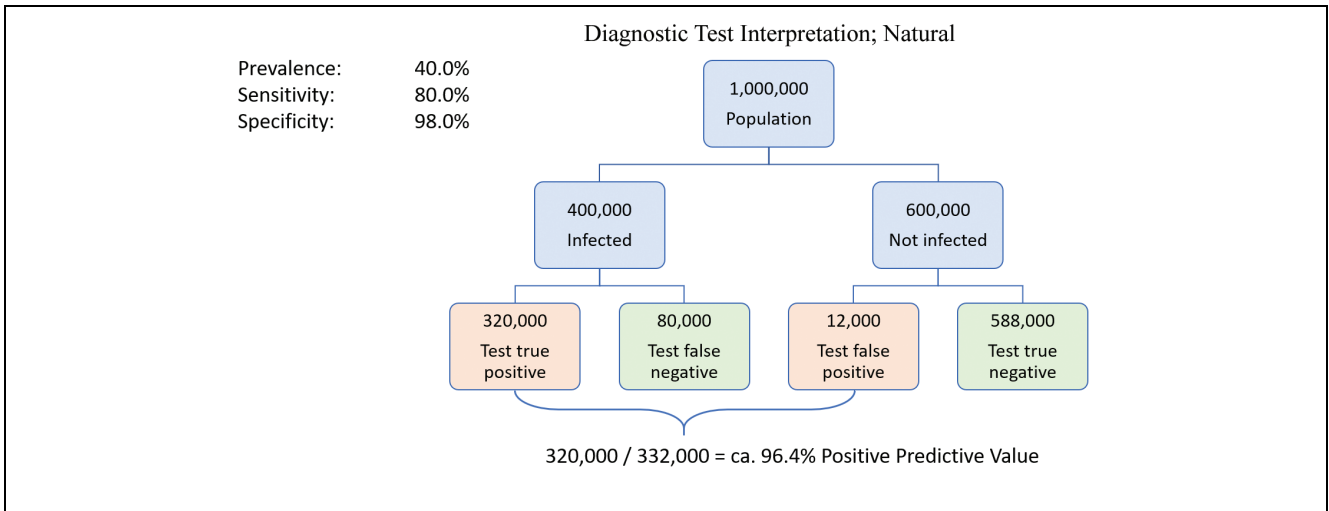

Figure 3 Natural frequency tree for antibody testing 6 weeks later with a SARS-CoV-2 prevalence of around 40%.

result imply?"). However, they only allow calculation of specific PPVs/NPVs given a specific prevalence. To decide when the prevalence of the disease is high enough for antibody tests to become useful as a screening-tool, we can use a PPV/NPV graph (Keller, Timiliotis, McDowell, and Benz, 2020, unpublished data). A PPV/NPV graph shows the likelihood of a true test result across the entire prevalence spectrum, given the specific test characteristics.

Figure 4 shows a PPV/NPV graph for antibody tests given a sensitivity of 80% and a specificity of 98%. From the graph it can be seen that population screening for SARS-CoV-2 in situations of low prevalence is not effective as the positive predictive value remains very low (Figure 4). Only once we reach about 1% prevalence of SARS-CoV-2 does the positive predictive value reach double digits and only once we reach about 10% prevalence does the PPV reach a reasonable satisfactory ~82% while maintaining a high NPV of ~98%. Note that the decision when to implement population-based screening depends on the costs and harms attributed to the test's two possible errors: false positives as well as false negatives.[14,15]
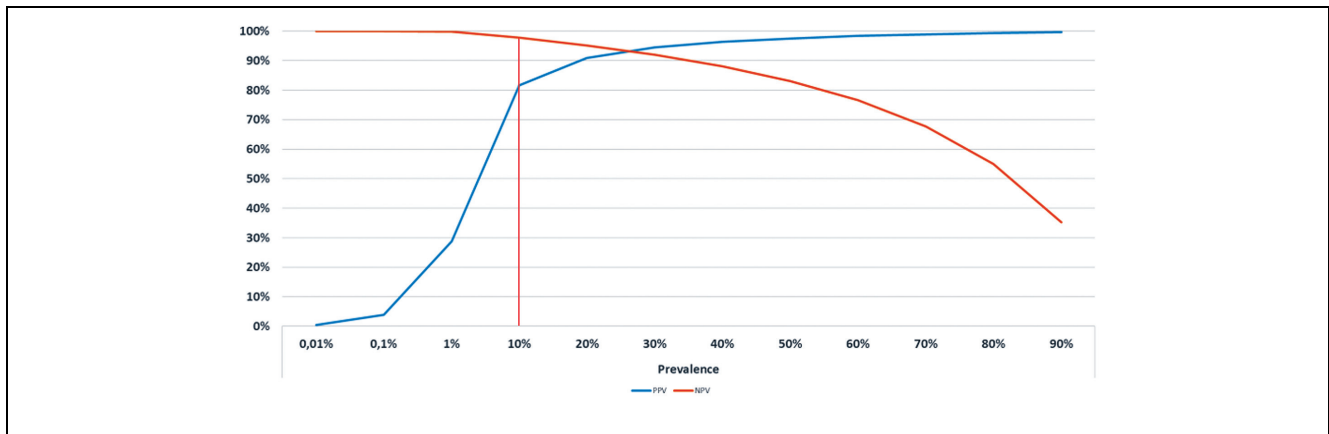
Figure 4 PPV/NPV graph for an antibody test with a sensitivity of 80% and specificity of 98% across the SARS-CoV-2 prevalence spectrum (from 0.01% to 90%). The *x*-axis is initially logarithmic (below 1%) and then continues linearly. The red line indicates a prevalence at which positive and negative test results can be assumed to be reasonably diagnostic of SARS-CoV-2 infection.

Implications for Using Antibody Testing for Population-Wide SARS-CoV-2 Screening and Individual-Level Immunity Testing

While estimates of the spread of SARS-CoV-2 differ by country and region, it can be safely assumed to be in the low single-digit range in most regions even when accounting for a large number of unknown cases. For example, in Germany, which tests more per capita than any other European nation, ~198,000 persons have tested positive for SARS-CoV-2 as of July 8, 2020, based on PCR testing.[16] This amounts to ~0.25% of the population. Even if the true number of SARS-CoV-2 infections were underestimated by a factor of 20, that is, the true rate of infected would be at 5%, it would still put the PPV of the antibody test at only about 50–50 (half of the persons testing positive would be false positives) if nationwide screening was implemented in Germany today. Even in the United States, which leads the world both in the number of daily tests per capita and the number of infected persons, only 1% of the population had been infected as of the 8th of July.[16]

One consequence of this finding is that it is very important that available ELISA-based antibody tests continue to be validated and improved. In the current situation, where prevalence is low in most places and tested groups, the tests' specificity is critical. For example, if a test has a specificity of 98%, even if no one is infected, the test would falsely indicate 2% of the persons tested as positive. If the specificity is 99.9%, on the other hand, the test would wrongly only show 0.1%

infected, that is, the number of false positives is much lower and does no longer weigh into the equation so heavily.* Furthermore, when using such tests to assess the prevalence of SARS-CoV-2 in particular locations (e.g., large hospitals), unless the results show a prevalence of ≥10% (below which the PPV drops rapidly), it is important that results are adequately adjusted downwards to account for the proportionally high number of false positives generated at low prevalence.

Put differently, at the early stages of a local outbreak or a nationwide pandemic, antibody testing should either be used only on subpopulations in which a high prevalence can be expected due to other, known factors or only be used in combination with other diagnostics with high specificity. This allows narrowing in on subpopulations with a higher disease prevalence, the effect of which is an increase in the probability with which a positive test result can be assumed to be due to an actual SARS-CoV-2 infection—a high PPV. Note that NFTs and PPV/NPV Graphs can also be used to estimate the prevalence in subgroups testing positive with other diagnostics. For example, a clinical assessment based on symptoms may act as a test for SARS-CoV-2 infection with its own sensitivity and specificity. The NFT can be used to calculate the positive and negative predictive values for this test, given an estimated prevalence in the general population.

---

*Readers are encouraged to experiment with the PPV/NPV Calculator Supplemental Material. The Calculator produces PPV/NPV curves from a given sensitivity and specificity as well as NFTs (where prevalence must additionally be provided).

The resulting PPV is nothing other than the prevalence of the disease in the subpopulation of people with these symptoms.

There is a seeming paradox here: When using antibody tests to estimate the true spread of SARS-CoV-2 in the general population, one must know the prevalence first. But it is possible to evaluate its usefulness for estimating the overall infection rate in the population in light of other, more robust indicators for the spread of the disease in a (sub)population such as the local SARS-CoV-2 hospitalization, intensive care unit admission, and mortality rates. Ideally, antibody studies should include a comparator region with low disease activity as indicated by these measures. This allows a rough estimate of the size of the over- or underdiagnosis of the prevalence assessment in the region of interest. Note that this applies similarly to assessing individuals. Comorbidities, exposure at the workplace, and other factors may influence the prior probability estimate of having had the disease (which in a population-screening context is the prevalence). A clinician seeing a patient with, for example, a high exposure workplace or past mild symptoms (without a positive PCR result) may wish to use available evidence or her clinical intuition to update the prior probability of the patient presenting having had COVID-19. The NFT or PPV/NPV graph can then be applied as it is presented in this tutorial. However, in the individual context, the behavioral consequences of test results need careful consideration. There is a high potential to elicit extremely risky behaviors from those who tested positively for immunity status, including those with false positives. Such persons may be more likely to expose themselves to SARS-CoV-2 infected persons, may preferentially work with vulnerable groups, or reduce hygiene and other protective measures, perhaps in order to save equipment at their place of work, thus creating situations of high SARS-CoV-2 transmission risk for themselves and others.

Individuals as well as the larger public should be informed about the possibilities and likelihood of false positives of antibody testing given estimates of local disease prevalence. Similar calls have been made to inform patients about the accuracy of PCR testing for acute COVID-19 infection[17] and the principles and tools discussed here apply to all diagnostic testing procedures. To this end, we suggest using natural frequency trees and PPV/NPV graphs to support both, doctor-patient consultations as well as policy discussions and public communication.

ORCID iD

Niklas Keller ⬤ https://orcid.org/0000-0003-4394-9788

References

1. Xu X, Sun J, Nie S, et al. Seroprevalence of immunoglobulin M and G antibodies against SARS-CoV-2 in China. *Nat Med.* 2020;26(8):1193–5. doi:10.1038/s41591-020-0949-6
2. Sood N, Simon P, Ebner P, et al. Seroprevalence of SARS-CoV-2-specific antibodies among adults in Los Angeles county, California, on April 10–11, 2020. *JAMA.* 2020;323(23):2425–27.
3. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *Lancet.* 2020;396(10250):535–44. doi:10.1016/S0140-6736(20)31483-5
4. Stringhini S, Wisniak A, Piumatti G, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *Lancet.* 2020;396(10247):313–9. doi:10.1016/S0140-6736(20)31304-0
5. Ng D, Goldgof G, Shy B, et al. SARS-CoV-2 seroprevalence and neutralizing activity in donor and patient blood from the San Francisco Bay area. *medRxiv.* Published online May 27, 2020. doi:10.1101/2020.05.19.20107482
6. Austin LC. Physician and nonphysician estimates of positive predictive value in diagnostic v. mass screening mammography: an examination of Bayesian reasoning. *Med Decis Making.* 2019;39(2):108–18. doi:10.1177/0272989X18823757
7. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill.* 2020;25(3):2000045. doi:10.2807/1560-7917.ES.2020.25.3.2000045
8. Drosten C, Hennig C. The coronavirus update from NDR info [in German; cited April 21, 2020]. *Norddeutscher Rundfunk.* Available from: https://www.ndr.de/nachrichten/info/podcast4684.html
9. Bendavid E, Mulaney B, Sood N, et al. COVID-19 antibody seroprevalence in Santa Clara County, California. *medRxiv.* Published online April 30, 2020. doi:10.1101/2020.04.14.20062463

10. World Health Organization. *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)*. World Health Organization; 2020.

11. McDowell M, Jacobs P. Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychol Bull*. 2017;143:1273–312. doi:10.1037/bul0000126.

12. Jenny MA, Keller N, Gigerenzer G. Assessing minimal medical statistical literacy using the Quick Risk Test: a prospective observational study in Germany. *BMJ Open*. 2018 8:e020847. doi:10.1136/bmjopen-2017-020847

13. Hoffrage U, Gigerenzer G. Using natural frequencies to improve diagnostic inferences. *Acad Med*. 1998;73(5): 538–40. doi:10.1097/00001888-199805000-00024

14. Brodersen J, Siersma VD. Long-term psychosocial consequences of false-positive screening mammography. *Ann Fam Med*. 2013;11:106–15. doi:10.1370/afm.1466

15. Bleyer A, Welch HG. Effect of three decades of screening mammography on breast-cancer incidence. *N Engl J Med*. 2012;367(21):1998–2005. doi:10.1056/NEJMoa1206809

16. John Hopkins University, Coronavirus Resource Centre [cited September 16, 2020]. Available from: https://coronavirus.jhu.edu/map.html

17. Watson J, Whiting PF, Brush JE. Interpreting a covid-19 test result. *BMJ*. 2020;369:m1808. doi:10.1136/bmj.m1808