Preliminary thoughts on

Interpreting Diagnostic Tests for Covid-19

Rev. Thomas Bayes Can Help

Yakov Ben-Haim[1], PhD and Shai Linn[2], MD, DrPH

Clinical response to the covid-19 epidemic depends, in part, on interpreting swab tests of body fluids to evaluate incident cases, and serological surveys to assess disease prevalence.

The Positive Predictive Value of a test (PPV) answers the question: What is the probability that a positive test indicates presence of disease? In statistical terms it is a conditional probability: the probability that a person is infected given that the person's test is positive. The Negative Predictive Value of a test (NPV) is the conditional probability that the person is not infected, given a negative test result.

The PPV and NPV are crucial in clinical decision making. If the PPV is large, then clinical intervention is strongly indicated for people whose tests results are positive. Similarly, if the NPV is large, then intervention is not indicated for people with negative test results.

This is where Rev. Thomas Bayes comes in. Bayes introduced a precise formulation of the concept of conditional probability, which was published in 1763, two years after his death. The PPV and NPV are each conditional probabilities, and they depend on three other probabilities. The *sensitivity* of a test is the conditional probability that the test will yield a positive result in an infected person. The *specificity* of a test is the conditional probability that the test will yield a negative result in an uninfected person. Sensitivity and specificity are properties of the test and its mode of application to the subject, and their values are usually (though not always) pretty well known. Finally, PPV and NPV also depend on the *prevalence:* the fraction of the population that is infected with the disease. The prevalence of covid-19 is poorly known in most countries at present.

Carver and Jones[1] estimate sensitivity of swab tests ranging from 0·60 to 0·70, or sometimes less. Sensitivity and specificity of serological tests for covid-19 are estimated by Bendavid and colleagues[2] (2020) at about 0·803 and 0·995, respectively. They concluded that the prevalence of covid-19 ranged from about 2·5% to 4·2% in the populations and region they studied at that time. The prevalence can be much different – lower or higher – as the disease emerges.

Neither the sensitivity nor the specificity of a test are immediately relevant to interpreting a test result for a specific individual. Large sensitivity and specificity does not mean that a positive (or negative) test result for patient X means that X is infected (or not). Only the PPV (or NPV) reveals that clinical insight. And if the prevalence is poorly known, then the PPV and NPV are poorly known, and the test result cannot be confidently interpreted.

Consider some examples. We focus on PPV; analogous arguments apply to NPV.

---

[1] Yitzhak Moda'i Chair in Technology and Economics, Technion – Israel Institute of Technology, Haifa 32000 Israel. yakov@technion.ac.il

\papers\corona-diag-prob2020\viewpoint010.docx

[2] School of Public Health, University of Haifa, Israel, linnshai@gmail.com

Suppose sensitivity and specificity are both 0·98, and suppose the prevalence is 3%, that is, 0·03. Then the PPV is 0·60, indicating 60% chance of illness given a positive test result. This is 20 times the naïve pre-test value of prevalence of 3%, so intervention would seem to be indicated. Even sensitivity of 70%, with the same prevalence and specificity, yields a PPV of 0·52; still 17 times the pre-test value. If both sensitivity and specificity are 70%, then the PPV with 3% prevalence is a mere 0·067; perhaps not a strong basis for clinical intervention. Thus, clinical diagnoses are highly dependent on our knowledge of the prevalence.

Thank you Rev. Bayes for explaining conditional probability. That is important, but there's more.

The estimated covid-19 infection prevalence is highly uncertain, so we ask: how much error in the estimated prevalence can we tolerate, if we want the estimated PPV of covid-19 to confidently support a clinical decision? For example, how much can the true prevalence deviate from an estimated 3%, if we can accept error in the PPV of no more than plus or minus ten percentage points? More generally, how robust to uncertainty in the prevalence, is the clinical decision to intervene or not? Info-gap theory suggests some answers (Ben-Haim[3,4]).

Suppose sensitivity is 0·70 and specificity is 0·98. Preliminary evidence suggests that prevalence is about 0·03, but this is uncertain and may be only 1% or 2%. We have very limited information about the prevalence, and cannot establish statistical confidence intervals. The estimated PPV is 0·52 with 0·03 prevalence, but it is 0·26 if the prevalence is 0·01. If a PPV of 0·52 is sufficient to justify intervention, then one asks if 0·26 also justifies intervention. If the answer is yes, then clinical intervention is indicated, and this recommendation has substantial robustness to uncertainty. If a PPV of 0·26 would not indicate intervention, then the decision to intervene is not robust to uncertainty in the prevalence, and then one might suggest additional testing and perhaps advise watchful waiting.

Our examples so far focused on PPV at low prevalence, and have shown the importance of considering robustness to uncertainty in the prevalence. In contrast, the NPV is very insensitive to uncertainty in the prevalence, when the prevalence is low, but the NPV becomes very sensitive at higher values of prevalence. For instance, at levels of prevalence that are thought to represent herd immunity – values around 60% prevalence or more – the NPV is quite sensitive to uncertainty in the prevalence. At high prevalence it is necessary to analyze robustness of the NPV, just as we have done for the PPV at low prevalence.

In summary, we have augmented the probabilistic concept of conditional probability, with the non-probabilistic concept of robustness to lack of knowledge. PPV and NPV are conditional probabilities of undisputed importance in clinical decision making. However, they depend on quantities – prevalence and sensitivity in particular – that may be poorly known. When this is the case, then clinical recommendation based on the best estimate of the PPV or NPV is not the final word. We must ask if the recommendation is robust to gaps in our information.

Mathematical details are available. See working paper:
Quantitative Analysis for Interpreting Diagnostic Tests for Covid-19: Rev. Thomas Bayes Can Help.
File: papers\corona-diag-prob2020\cdp005.pdf

References

1. Carver, C. and Nick Jones, N. Comparative accuracy of oropharyngeal and nasopharyngeal swabs for diagnosis of COVID-19, evidence-cov.id/swabs, 20 March 2020. Accessed 23.4.2020.
2. Bendavid E, Mulaney B, Sood N, et al., COVID-19 Antibody Seroprevalence in Santa Clara County, California. 11 April 2020. https://www.medrxiv.org/content/10.1101/2020.04.14.20062463v1 Accessed 23.4.2020.
3. Ben-Haim Y. 2006, *Info-Gap Decision Theory: Decisions Under Severe Uncertainty*, 2nd edition, Academic Press.
4. Ben-Haim Y. 2018, *Dilemmas of Wonderland: Decisions in the Age of Innovation,* Oxford University Press.