Are antibodies tests accurate? Understanding predictive values and uncertainty of serology tests for the novel coronavirus

Naomi C Brownstein[1,*] **and Yian Ann Chen**[1]

[1]Moffitt Cancer Center, Department of Biostatistics and Bioinformatics, Tampa, FL, USA
[*]naomi.brownstein@moffitt.org

ABSTRACT

Antibodies testing in the coronavirus era is frequently promoted, but the underlying statistics behind their validation has come under more scrutiny in recent weeks. We provide calculations, interpretations, and plots of positive and negative predictive values under a variety of scenarios. Prevalence, sensitivity, and specificity are estimated within ranges of values from researchers and antibodies manufacturers. Illustrative examples are highlighted, and interactive plots are provided in the Supplementary Material. Implications are discussed for society overall and across diverse locations with different levels of disease burden. Specifically, the proportion of positive serology tests that are false can differ drastically from up to 3% to 88% for people from different places with different proportions of infected people in the populations while the false negative rate is typically under 10%.

1 Introduction

The SARS-CoV-2 pandemic[1] is wreaking havoc on physical[2,3], mental[4], economic[5–7], and general societal health[8,9]. Potential treatments for Covid19 currently have limited evidence of efficacy[1]. Thus, it is critical to develop agents to prevent the spread of coronavirus, such as vaccines. Although research and development on vaccine candidates is ongoing[10,11], widespread availability of a safe and effective vaccine is not expected for months or even years[12–14]. Simultaneously, there is increasing evidence of asymptomatic infection and spread[15]. With scarce testing supplies[16], many people, blind to their prior infection status or lack thereof, are self-isolating; the current situation has even consequently been derided as "Schroedinger's virus"[17].

Without a vaccine, excitement about antibodies testing is growing.[18] Theoretical benefits of identifying individuals with antibodies abound. For example, health workers with prior undetected asymptomatic infections could gain peace of mind during patient care and conserve personal protective equipment for their remaining susceptible colleagues. Daily activities such as shopping, traveling, and dining could begin to resume, alleviating currently acute social and economic effects of the pandemic. Hoping to implement these potential benefits, some politicians have considered issuing immunity passports for people who are cleared by an antibodies test[19]. Yet, serology testing is not a panacea, and is associated with concerns about its use[20–23]. Proposals for implementing serology testing programs and understanding their benefits and limitations are available[24,25]

Given the increase in testing, proper interpretation of the results is critical with implications for medicine, public policy, and personal action. The goal of this paper is to estimate, quantify and visualize uncertainty in the predictive values and false positive rates of serology testing candidates available at the time of writing. Graphical displays of predictive values feature a range of scenarios. Section 2 reviews key metrics for serology tests. Section 3 summarizes and visualizes metrics in general and for serology tests operating under an Emergency Use Authorization (EUA). Section 4 applies the concepts to specific locations in the United States (US). Section 5 discusses limitations and implications. Finally, Section 6 details our statistical and graphical methods.

2 Background

This section outlines key statistical definitions related to serology testing. Readers familiar with testing characteristics may skip to Section 3. Definitions are included in the Supplementary Material in Section 9.2, Additional details on these concepts and examples relevant to serology may be found elsewhere[25–28].

Two properties of serology tests quantify how well the tests perform in on samples in a lab with known antibody status, Sensitivity is the probability that a serology test correctly classifies a sample with antibodies for SARS-CoV-2. Specificity is the probability that a test correctly classifies uninfected samples as lacking antibodies for SARS-CoV-2. Sensitivity and

specificity are pretest quantities, or validation metrics primarily meaningful before a serology test is taken. Sensitivity and specificity are defined by equations (1) and (2) in the Supplementary material, Section 9.2.

Rather, people who confront serology tests are likely interested in post-test probabilities, including positive predictive value (PPV) and negative predictive value (NPV). PPV is the probability that a person with a positive serology test indeed has a prior infection with and antibodies for SARS-CoV-2. NPV is the probability that a person who tests negative lacks antibodies for and has not yet been infected with SARS-CoV-2. Definitions and calculations for NPV and PPV are in equations (3) through (7). PPV and NPV are more relevant to patients and clinicians in interpreting serology test results than sensitivity and specificity.

The complements of the predictive values are probabilities that test results of each type are false. The False Positive Rate (FPR), the complement of PPV, is the proportion of people who test positive that are actually lacking a prior coronavirus infection. Similarly, the false negative rate (FNR), the complement of NPV, is the proportion of people who test negative that actually had a prior infection with coronavirus. The FPR (FNR) can be interpreted as the proportion of positive (negative) serology tests are false positives (negatives). FPR and FNR are shown in equations (8) and (9).

3 Results

This section reports values for the statistics described in section 2 to help contextualize serology test results. Prevalence estimates are reported in Section 3.1. Graphical displays of NPV and PPV for the range of values under study are found in Section 3.2. Sensitivity, specificity, PPV, and FPR for tests under study are reported in 3.3.

3.1 Prevalence Estimates

Due to the lack of available diagnostic tests in the US, official counts of Covid-19 cases are likely undercounted[29,30]. Additionally, people with asymptomatic infections are unlikely to seek medical care or diagnostic testing and are likely excluded from official counts. Consequently, reliable prevalence estimates are limited. Prevalence, which affects predictive value estimates, can be considered unknown, and varies over time.

Emerging research is beginning to estimate population prevalence. Over the period from March 31 to April 7, by one estimate[31], the prevalence by state ranged from about 0.4% in Alaska, Hawaii, Kentucky, and West Virgina to 8.5% in New York with a median prevalence of 0.9%. Another estimate[30] from April 11 found infection proportions within states spanning from 0.1% in rural states to 7.0% in New York, and an overall US prevalence of 1.2%. A third group proposes a method[29] with estimates that could indicate a prevalence of up to 10% of the population as of April 4, 2020. Additional specific overall and local estimates are currently limited. A collection of case studies is highlighted in Section 4. In brief, the prevalence of specific cities, states, and counties estimates ranged from less than 1% to about 30% in especially hard hit areas such as Chelsea County in Boston and New York City.

3.2 General Interpretation in the Context of Antibodies Testing

In this section we investigate ranges of plausible values of sensitivity, specificity for antibodies tests available and and prevalence for relevant areas and compare the predictive values. Predictive values are of interest based on minima of 80% sensitivity and 90% specificity to reflect reported values for FDA-authorized serology tests and described in Section 3.3. In general, the false negative rate is low and false positive rate is highly variable for available serology tests.

Figure 1 is a plot of NPV for these specificity, and sensitivity values and prevalence ranging from 1% to 30%, the range currently reported elsewhere as discussed in Sections 3.1 and 4. Under these scenarios, the minimum NPV was 91.3%, indicating that the false negative rate was less than 10% in all scenarios. Thus, NPV should be high, and FNR should be low for all serology tests within similar ranges for sensitivity, specificity, and prevalence. In other words, negative serology tests have a high likelihood of accurately reflecting a lack of antibodies in the general population of non-infected individuals.

By contrast, figure 2 shows a corresponding plot spanning a wide range of plausible PPV values. PPV increases with prevalence and is low with lower rates of antibodies in the population. At fixed prevalence values, specificity also quickly increases PPV with prevalence. Higher sensitivity improves PPV, although the growth of PPV with increases in sensitivity is less pronounced than with increases in specificity at a given prevalence. For example, an area with 10% prevalence would have 47.1% PPV for a test with 80% sensitivity and 90% specificity, 66.7% PPV for a test with 90% sensitivity and 95% specificity, and 91.3% PPV for a test with 95% sensitivity and 99% specificity. Equivalently, the false positive rates would be 52.9%, 33.3%, and 8.7%. In areas with a 30% infection rate, the same tests would yield respective PPVs of 77.4%, 88.5% and 97.6%, and false positives of 22.6%, 11.5%, and 2.4%. Yet, if the prevalence is 1%, then PPV could reach 49.0% for 95% sensitivity and 99% specificity or fall as low as 7.5%, indicating that only 7.5% of people with positive serology tests in fact possess antibodies. Put another way, for tests with 80% sensitivity and 90% specificity in locations with 1% prevalence, about 93% of people with positive serology tests would be expected to *lack* antibodies for SARS-CoV-2!

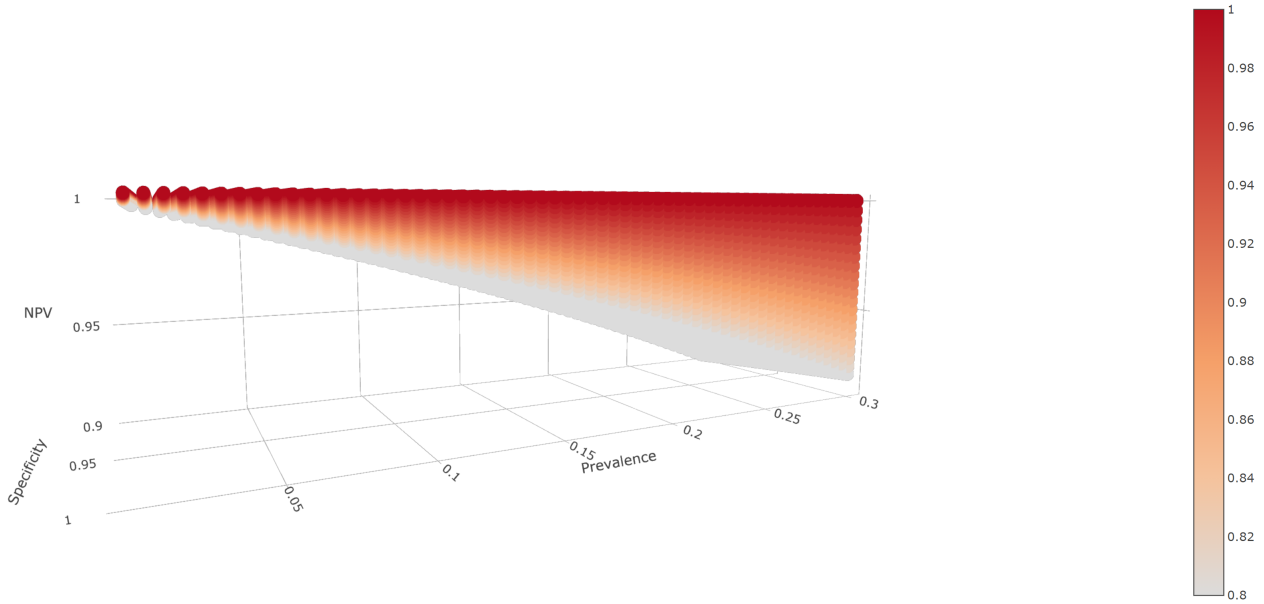


Figure 1. Plot of NPV by prevalence (0 to 0.3), specificity (0.9 to 1), and sensitivity (0.8 to 1). The bar on the right with sensitivity is denoted by color, with lighter colors denoting lower sensitivity and darker colors denoting higher sensitivity. All parameters are reported as decimals ranging from 0 to 1. NPV values exceeded 0.9 for all input parameters.

3.3 Analysis of Individual Antibodies Tests

As of May 22, 2020, the FDA has allowed 13 serology tests to operate under an Emergency Use Authorization (EUA)[32]. Characteristics of these tests appear in Table 1. Sensitivity estimates range from 83% to 100%, with 95% confidence limits ranging from about 68% to 100%. Specificity estimates range from about 94% to 100%, with confidence bounds ranging from 88% to 100%. Figures 3 and 4 highlight the relationship between PPV and prevalence for each of the 13 serology tests granted an EUA by the US FDA. The figures illustrate the known relationship that PPV should be lower in populations with lower prevalence[28], and that PPV increases more rapidly with increasing specificity than with increasing sensitivity.

At low prevalence, such as 1%, many tests exhibit low PPV and high rates of false positives (Table 1). The upper limits of false positive rates for 9 tests exceed 60% and 12 tests exceed 50% given 1% prevalence. Thus, in regions with 1% prevalence, scenarios having 50% or more percent of positive serology tests corresponding to people *lacking* SARS-Cov-2 antibodies would be within the realm of reasonable expectations. In addition, two tests have not only poor estimated false positive rates, but report high precision. Tests by Cellex and Chembio are estimated to have 80.8% and 85.6% false positives at 1% prevalence, with even the lower bounds on their FPRhovering around 70%. In such low prevalence populations, on average, anywhere between 7 and 9 out of 10 positive tests by Cellex and Chembio are expected to be false.

Results for all tests improve with prevalence, but overall false positive rates remain concerning. Although only two point estimates for false positive rates exceed 20% assuming 10% prevalence, which is higher than most currently estimated infection rates[30,31], the upper bound for false positive rates exceed 20% for 7 out of 13 tests. Moreover, only 3 tests – Abbott (Architect), Ortho-clinical (IgG), and Roche – have upper bounds on false positive rates below 10% when the prevalence is 10%. In other words, 10 out of 13 tests could reasonably produce at least 1 false positive out of 10 if the regional prevalence is 10%.

Estimates or confidence bounds of exactly 100%, reported for seven out of the thirteen tests, are shown for completeness. These estimates should be interpreted with caution, as tests establishing sensitivity and specificity were done on small samples[32], which may lack sufficient precision[20,33] for estimation when these quantities are expected to be close to 1. In fact, while specificity estimates of the tests by Ortho-Clinical were validated based on samples of around 400, estimates of 100% sensitivity by Abbott and Roche or 100% specificity by Euroimmun and Mt. Sinai were validated on samples of less than 100. Yet, as shown in equations (6) and (7), perfect (100%) specificity implies perfect PPV and perfect sensitivity implies perfect NPV. In these cases, it is especially important to consider the entire range of values for PPV and FPR. Indeed, while tests by Mt. Sinai

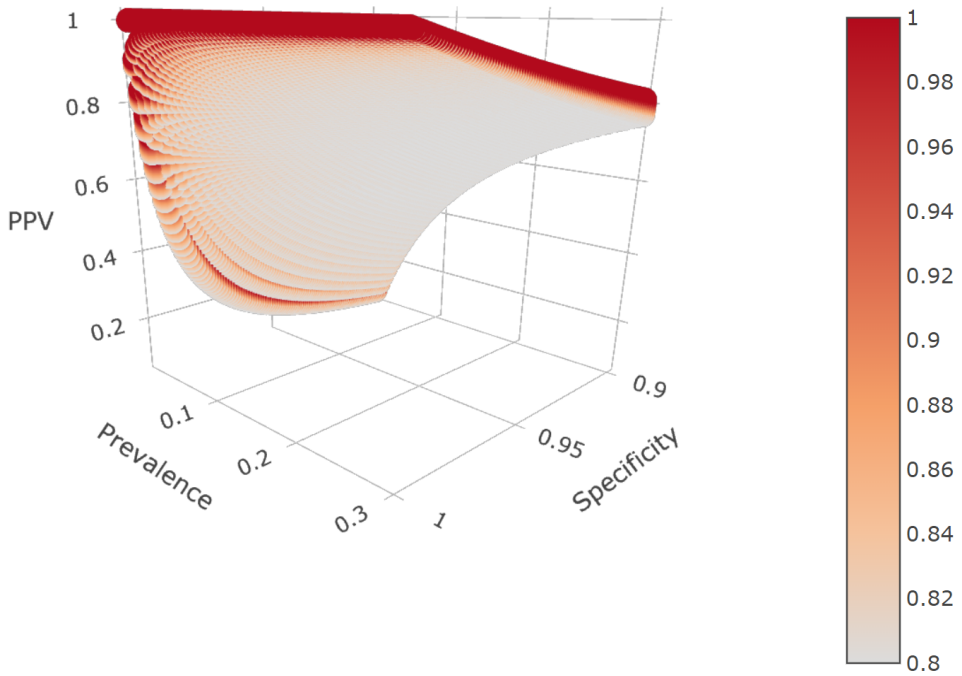


Figure 2. Plot of PPV by prevalence (0.01 to 0.3), specificity (0.9 to 1), and sensitivity (0.8 to 1). Sensitivity is denoted by color, with lighter colors denoting low sensitivity and darker colors denoting higher sensitivity. All parameters are reported as decimals ranging from 0 to 1. PPV varied widely based on different parameters, ranging from 0.07 to 1.

Test	Type	Sensitivity	Specificity	*n* (sens; spec)*	FPR at 1%	FPR at 10%
Abbot (Alinity)	IgG	100 (89.9, 100)	99.0 (94.6, 99.8)	34; 100	49.7 (16.5, 85.6)	1.8 (8.3, 35.1)
Abbott (Architect)	IgG	100 (95.8, 100)	99.6 (99.0, 99.9)	88; 1070	28.4 (9.0, 50.8)	3.5 (0.9, 8.6)
Autobio	Combined	88.1 (84.6, 90.9)	99.0 (97.2, 99.7)	302; 312	52.9 (24.6, 76.6)	9.3 (2.9, 23.0)
Bio-Rad	Pan-Ig	92.2 (81.5, 96.9)	99.6 (98.7, 99.9)	51; 687	30.0 (9.3, 61.2)	3.8 (0.9, 12.6)
Cellex	Combined	93.8 (88.2, 96.8)	96.0 (92.8, 97.8)	128; 250	80.8 (69.2, 89.0)	27.7 (17.0, 42.4)
Chembio	Combined	93.5 (79.3, 98.2)	94.4 (88.9, 97.3)	31; 125	85.6 (73.1, 93.3)	35.0 (19.8, 55.7)
DiaSorin	IgG	97.6 (87.4, 99.6)	99.3 (98.6, 99.6)	41; 1090	41.5 (28.4, 61.3)	6.1 (3.5, 12.6)
EUROIMMUN	IgG	90.0 (74.4, 96.5)	100 (95.4, 100)	30; 80	0 (0, 86.0)	0 (0, 35.8)
Mt Sinai	ELISA Comb.	92.5 (80.1, 97.4)	100 (95.1, 100)	40; 74	0 (0, 85.8)	0 (0, 35.5)
Ortho-Clinical	IgG	87.5 (75.3, 94.1)	100 (99.1, 100)	40; 407	0 (0, 54.2)	0 (0, 9.7)
Ortho-Clinical	Pan-Ig	83.3 (68.1, 92.1)	100 (99.0, 100)	49; 400	0 (0, 59.2)	0 (0, 11.7)
Roche	Pan-Ig	100 (88.3, 100)	99.8 (99.7, 99.9)	29; 5272	16.5 (9.0, 25.2)	1.8 (0.9, 3.0)
Wadsworth	Pan-Ig	88.0 (80.5, 92.8)	98.8 (97.3, 99.5)	108; 433	57.4 (34.8, 76.9)	10.9 (4.6, 23.2)

Table 1. Statistics for FDA authorized serology tests with EUA
All numbers are percentages. Interval estimates for sensitivity and specificity are 95% confidence intervals reported by the FDA[32]. Interval estimates for the false positive rates are minimum and maximum values of all possible false positive rates calculated at the specified prevalence level for the corresponding test based on all possible estimates of sensitivity and specificity within the 95% confidence intervals.
* *n* refers to the number of samples used to calculate the pretest probabilities. The number on the left refers to the number of samples with infected with SARS-COV-2 that were tested to estimate sensitivity. The number on the right refers to the number of control samples used to estimate specificity.

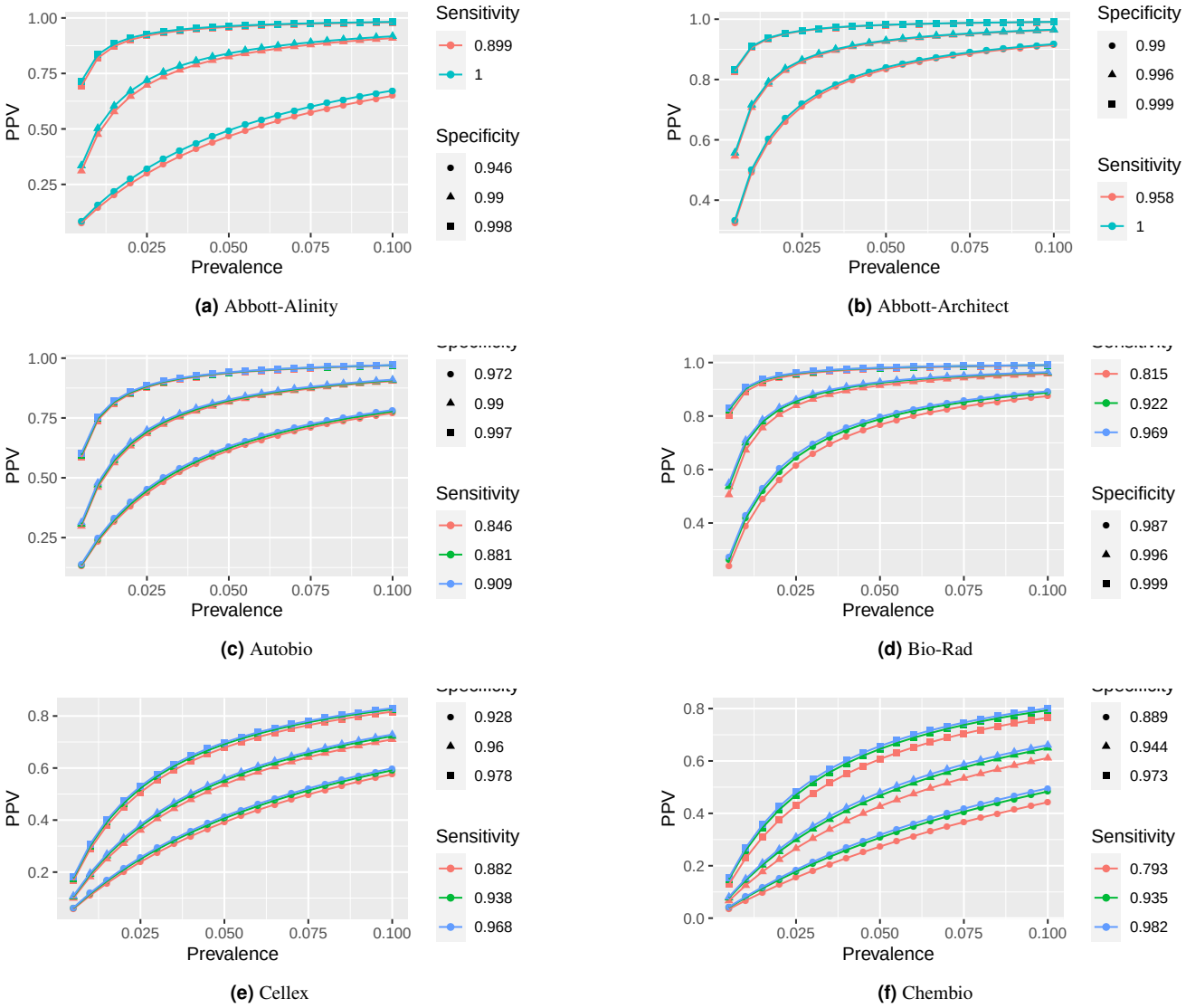
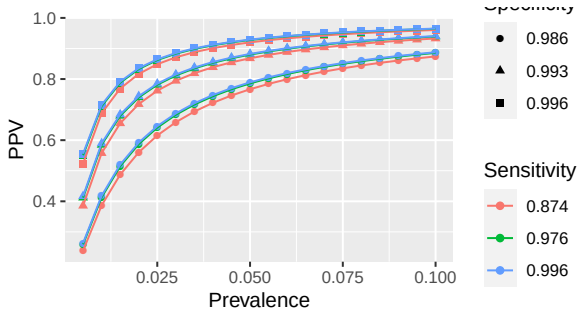
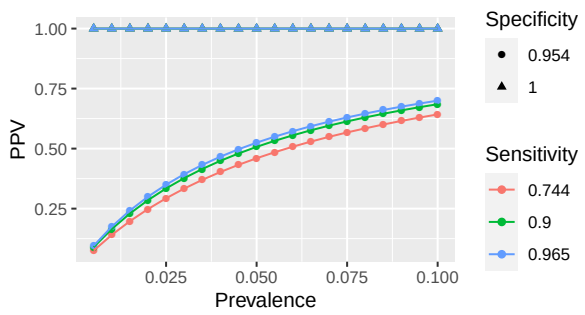


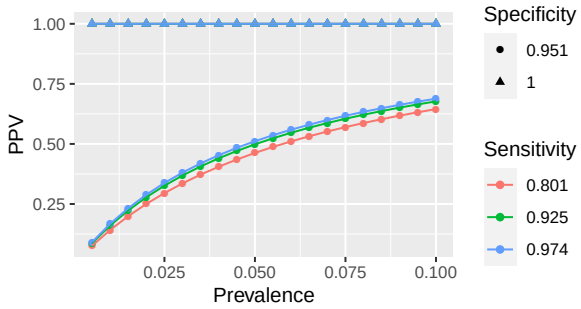
Figure 3. PPV by prevalence (up to 10%) for FDA tests (A-C)



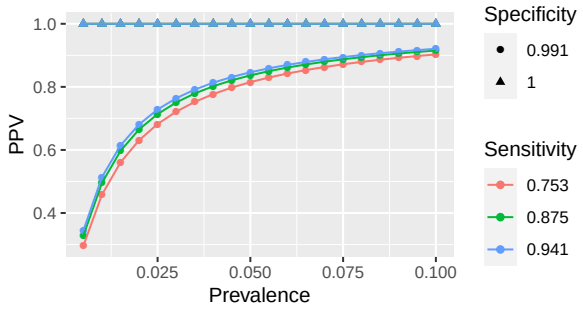
(a) DiaSorin



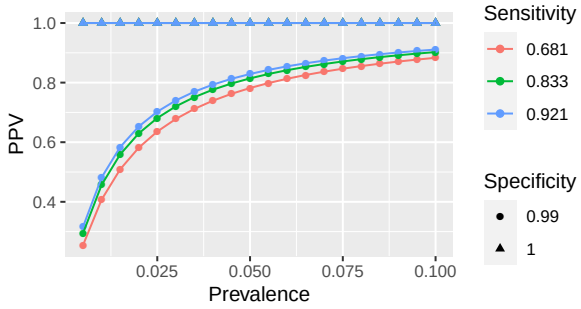
(b) EUROIMMUN



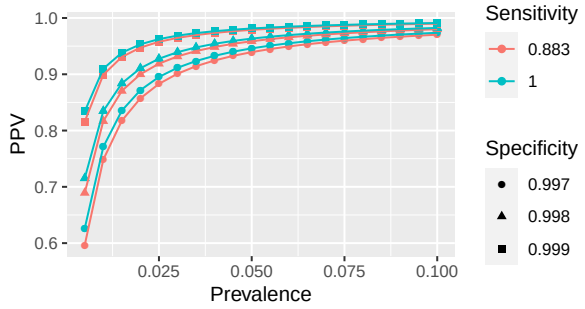
(c) Mt Sinai



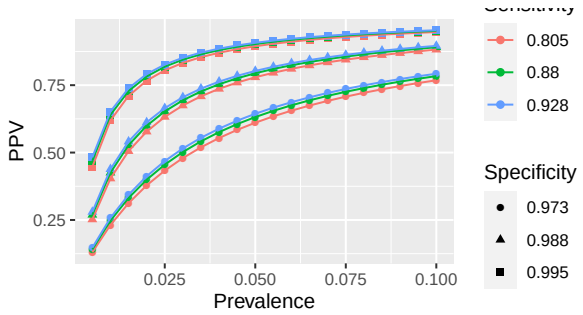
(d) Ortho-Clinical IgG



(e) Ortho-Clinic Pan IgG



(f) Roche Pan-Ig



(g) Wadsworth

Figure 4. PPV by prevalence (up to 10%) for FDA tests (D-Z)

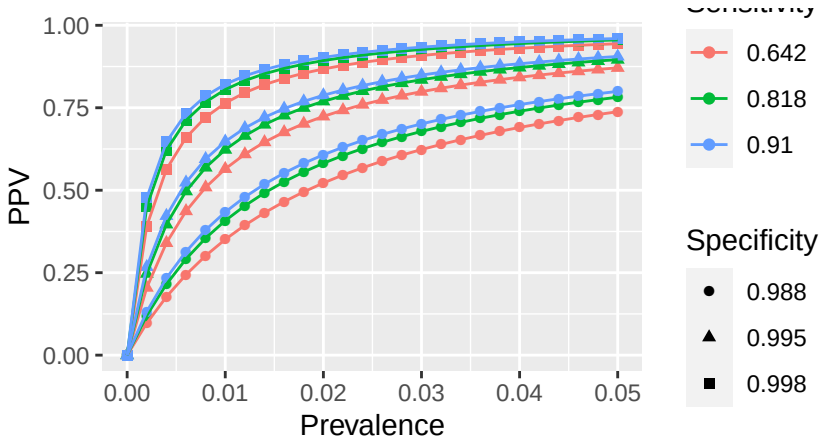


Figure 5. PPV for Antibodies test in Santa Clara County Study[35]. To reflect variation within and between references[35–37], prevalence values are included from 0% to 5% by 0.2%.

and Euroimmun have point estimates and lower interval estimates of 0% false positives, upper interval estimates climb to about 86%. Even at 10% prevalence, if the true specificity is closer to the lower 95% confidence limit, then false positive rate of Euroimmun and Mt. Sinai would be above 35%, rendering over a third of positive serology test results as likely false positives.

The test by Roche pharmaceuticals, advertised as highly accurate,[34], indeed has high PPV and low false positive rates even at 1%, with an upper limits for the false positive rate of 25.2%.The Roche test is the only test that could reasonably claim to expect more true positive results than false positive results in populations with low prevalence (1% or less). However, it would still be reasonable to expect that up to one quarter of positive test results could come from patients lacking antibodies to SARS-COV-2. Given the sample size[32] of 29 for calculating sensitivity, the associated uncertainty could be compounded in the estimates of PPV and FPR, rendering the upper bounds critical to measure and understand.

4 Application to Specific Locations

In this section, we evaluate PPV and FPR for a set of areas with three local studies of seroprevalence, in California, New York, and Boston. When rigorous prevalence estimates are unavailable, we use the proportion of positive tests as surrogates for prevalence for the purpose of estimating the rates of false positives in these studies. Rationale for and limitations of this approach along with a sensitivity analysis are discussed in the supplementary material, Section 9.3.

4.1 Santa Clara County

An early seroprevalence study was conducted in Santa Clara County.However, after extensive scrutiny, the authors provided updated estimates. Based on the updated preprint[35], the estimated prevalence adjusted for test performance characteristics were 1.2% (95%CI 0.7-1.8%) unweighted and 2.8% (95CI 1.3-4.7%) for weighted estimates based on demographic characteristics in Santa Clara County. Reanalyses[36,37] reported updated seroprevalence ranges of 0.27% and 3.21%[36] and 0% to 2.1%[37]. An estimate of prevalence in California from March 31 to April 7 is 0.9%[31]. Estimates of sensitivity and specificity vary[35–37]; we use the Bayesian posterior estimates[37] combining information from all sources as 99.5% specificity with a 95% posterior interval of (98.8%,99.8%) and 81.8% sensitivity (64.2%.91.0%)

Predictive values in the Santa Clara Study nearly spanned the entire range of probabilities. Using prevalence values reflecting prior estimates ranging from 0% to 5%, the PPV in Santa Clara County at the time of the study ranges from 0% to 96% (Figure 5). Prevalence near the low but nonzero end of the updated estimates (e.g. 0.2%)[36,37] correspond to PPV ranging from about 9.7% to to 47.7%, indicating that between about 26 and 46 of the 50 positive tests could be false. For prevalence near the high end of the updated estimates (4.7%)[35] ranges from 73.0% to 95.8%, or 2 to 14 false positives. Reanalyses estimate smaller upper bounds on prevalence (2.1%[37] and 3.2%[36]), which PPV ranges of 52.2% to 92.3% or 63.9% to 93.8%. These estimates correspond to false positive counts ranging from 2 to 24 or 3 to 18.

4.2 New York

On April 23, Govornor Andrew Cuomo announced results from a serology study in New York[38]. Seropositivity rates were 13.9% for the state overall and differed by region. New York City, Long Island, Westchester and Rockland, and elsewhere in New York, respectively, had 21.2%, 16.7%, 11.7%, and 3.6% seropositive. Specificity for the test was reported to be in the

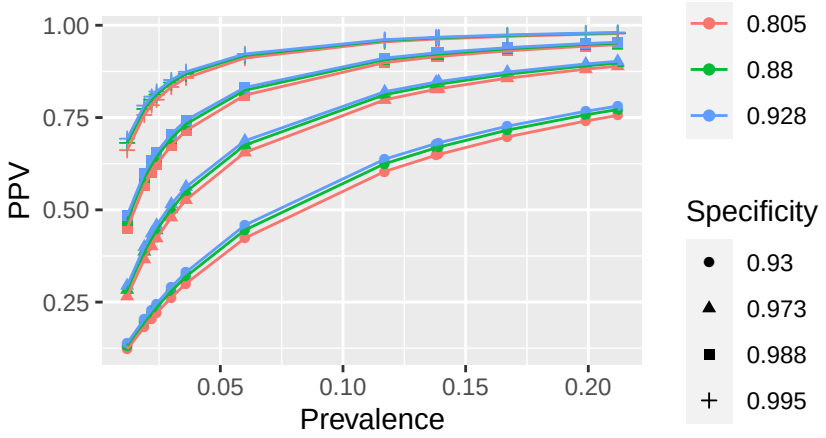


Figure 6. PPV for Antibodies test in New York Overall and by Region[32,38–40]

range of 93-100%[39], and sensitivity was not listed. However, the test was attributed to Wadsworth Center by the New York State Department of Health; the Wadsworth test parameters are reported[32] in Table 1. Results were then updated[40] on May 2. At that time, 12.3% of the population of New York state was reported to have Covid-19 antibodies based on a test of 15,000 people. By region, these figures ranged from 1.2% seropositive in North Country to 19.9% in New York City. We estimate the PPV for all combinations of values reported in all of these sources assuming the same serology tests were used in both studies.

Figure 6 shows the range of PPV based on each of these possible values of sensitivity, specificity, and prevalence. New York City and Long Island have the highest prevalence and highest PPV, ranging from 74% to 98% and 60% to 97% in all scenarios. Statewide and other areas are in the middle: 60% to 96% for Rockland, 65% to 97% for statewide. By contrast, PPV can be as low as 30% for the rest of the state, assuming a prevalence 3.6%, if the specificity is 93% or as high as 87% if sensitivity is at the upper limit of the confidence interval reported in the EUA[32]. Even worse, areas with low prevalence[40]such as North Country (1.2%), Central NY (1.9%), and Capital District (2.2%) have PPV estimates ranging from 12% to 69%, 18% to 78% and 21% to 81% In other words, the false positive rate in New York ranged from 2% to 88% depending on the region and assumed prevalence under consideration and uncertainty in the sensitivity and specificity.

4.2.1 Chelsea, Massachusetts

Researchers at Massachusetts General Hospital conducted a seroprevalence study in the city of Chelsea[22] and reported that 31.5% of participants tested positive. The researchers on the Chelsea study reported specificity exceeding 99.5%. However, the manufacturer, BioMedomics, which is not part of the labs with EUA in Section 3.3, reports a sensitivity of 88.66% and specificity of 90.63%,[41].

Assuming prevalence of 31.5% and specificity and sensitivity values reported by Biomedomics[41], PPV was only about 81%. This means that in this sample of 63 positives, about 51 are expected to be true positives and about 12 are expected to be false positives. Thus, the prevalence estimate in Chelsea reported in the press based on this sample alone is likely to be too high. In addition, participants were recruited by a convenience sample of people outside on a particular street, which may not be representative of the general population in Chelsea.

5 Discussion

Antibodies tests can yield two possible errors with different implications[25]. Consequences of false negative test results would likely relate to failing to remove negative effects of limitations during the pandemic. For example, assuming that antibodies indeed confirm protection, then people with antibodies who test negative would be safe to return to work but their negative test might convince them to remain at home. This would prolong the negative mental and physical effects of social isolation as well as economic effects to individuals and society overall. Fortunately, the false negative rate was under 10% in all scenarios.

Unfortunately, the false positive rate can be shockingly high. Based on the prevalence estimated throughout the US and serology studies in California, New York and Boston, the FPR of antibody test results range from 2% to 88%. Point estimates of tests ests with an EUA[42] reached 86% and upper limits reached 93% when the prevalence is 1%. Tests with low PPV and high FPR can be dangerous by giving patients with positive tests a false sense of security. Ironically, these people may then *increase* their risk of contracting Covid19 if they relax their use of protective measures, such as mask wearing and social distancing.

The number and implication of false positives is growing with large organizations encouraging widespread serology testing. Quest Diagnostics is offering tests by Abbott, Ortho-Clinical and Euroimmun for purchase[43]. The tests can have up to 86% FPR

in locations with 1% prevalence. Even at 10% prevalence, over one-third of positives could be false. OneBlood, a non-profit blood donation and distribution organization, is encouraging large scale blood donation and then testing samples for antibodies using the OrthoClinical total test[44]. As discussed in section 3.3, the OrthoClinical test could have a false positive rate of nearly 60% with 1% prevalence or nearly 12% if the prevalence is 10%. Further, OneBlood is planning to use blood samples testing positive for antibodies as convalescent plasma, which could inadvertently harm patients. Using all samples that test positive could mean that large numbers of samples falsely considered to contain convalescent plasma with would become inadvertent placebos for patients actively struggling with Covid-19.

One recommendation for individuals who test positive is to consider a follow-up antibody test[45]. For instance, if the FPR is 50%, assuming independent tests, the probability of two false positives drops in half (25%). The FDA includes a calculator for PPV of individual and combined tests[32]. Follow-up tests are common in other diseases with low PPV screening methods, such as mammograms for breast cancer[46]. Another idea is to test all contacts for antibodies and use their results as evidence to support or refute the original serology test. Pursuing contacts of additional seropositive individuals may increased contact tracing and testing, which can either hinder growth of future outbreaks or divert scarce resources from higher risk contacts[25]. Moreover, increased testing brings cost and feasibility concerns.

There are some additional limitations of our paper. For instance, some of the information may become outdated quickly. Tests operating under an FDA EUA will likely increase over time. For instance, one of the 13 tests was added to the original 12 tests made available by FDA under EUA while writing this paper. The FDA is aware that at least 160 serology tests were available before the FDA increased its oversight[47]. This implies some of the antibody tests on the market might even have lower sensitivity or specificity than those included in this paper and therefore have even higher FPR. Similarly, if an infected patient takes the test before antibodies are developed, then the sensitivity will be lower. An extended discussion is included in the Supplementary Material (Section 9.1). At the same time, prevalence may increase over time at least for some of the cities or towns. Importantly, neither the results nor the interpretation for serology tests generalize to diagnostic tests (See Section 9.4).

In conclusion, serology tests for the novel coronavirus generally have low false negative rates and highly volatile false positive rates. While false positive rates decrease with increasing prevalence, current prevalence estimates remain low in most areas of the US. With increasing serology testing and likely increased reporting of testing results, it is critical to understand these values and interpret test results properly. We hope that this context and interpretation can aid doctors, patients, researchers, and policy makers in informed decision making, which may even save lives.

6 Methods

We collected reported sensitivity and specificity values of serology tests with EUA approval by the US FDA. Prevalence estimates were also collected to determine an appropriate range for the plots. These values were combined to produce estimates of PPV and NPV for a variety of input parameter values. We then honed in on specific FDA tests and areas that have conducted serology tests to provide estimates and uncertainty for PPV and false positive rates.

All programs utilized R version 3.6.1. PPV and NPV were calculated using package MKmisc[48]. Plots were created with packages plotly[49] and ggplot2[50]. Figures 1 and 2, as higher dimensional plots, were designed to allow interactive visualization. Code to generate the plots and view them in an interactive mode may be downloaded from our github repository at https://github.com/nbrownst/AntibodiesPredictiveValues.

References

1. Pooladanda, V., Thatikonda, S. & Godugu, C. The current understanding and potential therapeutic options to combat covid-19. *Life Sci.* 117765 (2020).
2. Borges do Nascimento, I. J. *et al.* Novel coronavirus infection (covid-19) in humans: a scoping review and meta-analysis. *J. clinical medicine* **9**, 941 (2020).
3. do Nascimento, I. J. B. *et al.* Coronavirus disease (covid-19) pandemic: An overview of systematic reviews. *medRxiv* (2020).
4. Rajkumar, R. P. Covid-19 and mental health: A review of the existing literature. *Asian J. Psychiatry* 102066 (2020).
5. Barua, S. *et al.* Understanding coronanomics: The economic implications of the coronavirus (covid-19) pandemic. *Manuscript* (2020).
6. Fernandes, N. Economic effects of coronavirus outbreak (covid-19) on the world economy. *Available at SSRN 3557504* (2020).
7. Ozili, P. K. & Arun, T. Spillover of covid-19: impact on the global economy. *Available at SSRN 3562570* (2020).

8. Di Gennaro, F. *et al.* Coronavirus diseases (covid-19) current status and future perspectives: A narrative review. *Int. J. Environ. Res. Public Heal.* **17**, 2690 (2020).

9. Patrick, S. L. & Cormier, H. C. Are our lives the experiment? covid-19 lessons during a chaotic natural experiment a commentary. *Heal. Behav. Policy Rev.* **7**, 165–169, DOI: doi:10.14485/HBPR.7.2.10 (2020).

10. Chen, W.-H., Strych, U., Hotez, P. J. & Bottazzi, M. E. The sars-cov-2 vaccine pipeline: an overview. *Curr. tropical medicine reports* 1–4 (2020).

11. Prachar, M. *et al.* Covid-19 vaccine candidates: Prediction and validation of 174 sars-cov-2 epitopes. *bioRxiv* (2020).

12. Eyal, N., Lipsitch, M. & Smith, P. G. Human challenge studies to accelerate coronavirus vaccine licensure. *The J. Infect. Dis.* (2020).

13. Spinney, L. When will a coronavirus vaccine be ready. *The Guard. Retrieved* **18** (2020).

14. Omer, S. B., Malani, P. & Del Rio, C. The covid-19 pandemic in the us: a clinical update. *JAMA* (2020).

15. Velavan, T. P. & Meyer, C. G. The covid-19 epidemic. *Trop. medicine & international health* **25**, 278 (2020).

16. Emanuel, E. J. *et al.* Fair allocation of scarce medical resources in the time of covid-19. *New Engl. J. Medicine* **0**, null, DOI: 10.1056/NEJMsb2005114 (0). https://doi.org/10.1056/NEJMsb2005114.

17. John, B. S. Life in the time of covid-19: a crisis and a crocus. https://www.uchealth.org/today/life-in-the-time-of-covid-19-a-crisis-and-a-crocus/. Accessed: 2020-05-05.

18. Winter, A. K. & Hegde, S. T. The important role of serology for covid-19 control. *The Lancet. Infect. Dis.* (2020).

19. Organization, W. H. *et al.* "immunity passports" in the context of covid-19: scientific brief, 24 april 2020. Tech. Rep., World Health Organization (2020).

20. Farnsworth, C. W. & Anderson, N. W. Sars-cov-2 serology: Much hype, little data. *Clin. Chem.* (2020).

21. Torres, R. & Rinder, H. M. Double-edged spike: Are sars-cov-2 serologic tests safe right now? *Am. J. Clin. Pathol.* (2020).

22. Vogel, G. First antibody surveys draw fire for quality, bias. *Science* **368**, 350–351, DOI: 10.1126/science.368.6489.350 (2020). https://science.sciencemag.org/content/368/6489/350.full.pdf.

23. Phelan, A. L. Covid-19 immunity passports and vaccination certificates: scientific, equitable, and legal challenges. *Lancet (London, England)* (2020).

24. WHO. Population-based age-stratified seroepidemiological investigation protocol for covid-19 virus infection. Tech. Rep., World Health Organization (2020). Available at: https://www.who.int/publications-detail/population-based-age-stratified-seroepidemiological-investigation-protocol-for-covid-19-virus-infection.

25. Gronvall, G. *et al.* Developing a national strategy for serology (antibody testing) in the united states. Tech. Rep., Johns Hopkins Center for Health Security (2020). Available at: https://www.centerforhealthsecurity.org/our-work/publications/developing-a-national-strategy-for-serology-antibody-testing-in-the-US.

26. Pepe, M. S. *et al. The statistical evaluation of medical tests for classification and prediction* (Medicine, 2003).

27. Fleiss, J. L., Levin, B. & Paik, M. C. *Statistical methods for rates and proportions* (john wiley & sons, 2013).

28. Bonislawski, A. False positives could undermine utility of sars-cov-2 serology testing. https://www.360dx.com/infectious-disease/false-positives-could-undermine-utility-sars-cov-2-serology-testing (2020). [Online; accessed 13-May-2020].

29. Lu, F. S., Nguyen, A. T., Link, N. & Santillana, M. Estimating the prevalence of covid-19 in the united states: Three complementary approaches. *medRxiv* DOI: 10.1101/2020.04.18.20070821 (2020). https://www.medrxiv.org/content/early/2020/04/23/2020.04.18.20070821.full.pdf.

30. Breton, T. R. An estimate of unidentified and total us coronavirus cases by state on april 11, 2020. *Available at SSRN 3583941* (2020).

31. Benatia, D., Godefroy, R. & Lewis, J. Estimating covid-19 prevalence in the united states: A sample selection model approach. *medRxiv* (2020).

32. Eua authorized serology test performance. https://www.fda.gov/medical-devices/emergency-situations-medical-devices/eua-authorized-serology-test-performance (2020). Online: Accessed 05-23-2020.

33. Arya, R., Antonisamy, B. & Kumar, S. Sample size estimation in prevalence studies. *The Indian J. Pediatr.* **79**, 1482–1488 (2012).

34. Elecsys® anti-sars-cov-2. https://diagnostics.roche.com/us/en/products/params/elecsys-anti-sars-cov-2.html#productInfo. [Online; accessed 5-May-2020].

35. Bendavid, E. *et al.* Covid-19 antibody seroprevalence in santa clara county, california. *medRxiv* DOI: 10.1101/2020.04.14.20062463 (2020). https://www.medrxiv.org/content/10.1101/2020.04.14.20062463v2.full.pdf.

36. Bennett, S. T. & Steyvers, M. Estimating covid-19 antibody seroprevalence in santa clara county, california. a re-analysis of bendavid et al. *medRxiv* (2020).

37. Gelman, A. & Carpenter, B. Bayesian analysis of tests with unknown specificity and sensitivity. http://www.stat.columbia.edu/ gelman/research/unpublished/specificity.pdf (2020). Accessed 05-25-2020.

38. Video, audio, photos & rush transcript: Amid ongoing covid-19 pandemic, governor cuomo announces state health department will partner with attorney general james to investigate nursing home violations. https://www.governor.ny.gov/news/video-audio-photos-rush-transcript-amid-ongoing-covid-19-pandemic-governor-cuomo-announces-12 (2020). Online: Accessed 05-15-2020.

39. Bonislawski, A. New york, california serology studies give early estimates of covid-19 prevalence. https://www.360dx.com/infectious-disease/new-york-california-serology-studies-give-early-estimates-covid-19-prevalence (2020). [Online; accessed 13-May-2020].

40. Amid ongoing covid-19 pandemic, governor cuomo announces results of completed antibody testing study of 15,000 people showing 12.3 percent of population has covid-19 antibodies. https://www.governor.ny.gov/news/amid-ongoing-covid-19-pandemic-governor-cuomo-announces-results-completed-antibody-testing (2020). Online: Accessed 05-16-2020.

41. Covid-19 igm/igg rapid test. https://biomedomics.com/products/infectious-disease/covid-19-rt/. [Online; accessed 13-May-2020].

42. FDA. Policy for coronavirus disease-2019 tests during the public health emergeyncy (revised). Tech. Rep., U.S. Department of Health and Human Services, Food and Drug Administration (2020). Available at: https://www.fda.gov/media/135659/download.

43. Patient and healthcare provider fact sheets for covid-19 testing. https://www.questdiagnostics.com/home/Covid-19/factsheet/ (2020). Accessed 05-25-2020.

44. Covid-19 antibody testing. https://www.oneblood.org/lp/oneblood-covid-19.stml. Online: Accessed 05-23-2020.

45. Yi, G., He, W., Lin, D. K.-J. & Yu, C.-M. Covid-19: Should we test everyone? (2020). 2004.01252.

46. Nelson, H. D. *et al.* Screening for breast cancer: an update for the us preventive services task force. *Annals internal medicine* **151**, 727–737 (2009).

47. fda steps up scrutiny coronavirus antibody tests ensure accuracy. https://www.washingtonpost.com/health/2020/05/04/fda-steps-up-scrutiny-coronavirus-antibody-tests-ensure-accuracy/. Online; accessed 4-May-2020.

48. Kohl, M. *MKmisc: Miscellaneous functions from M. Kohl* (2019). R package version 1.6.

49. Sievert, C. *Interactive Web-Based Data Visualization with R, plotly, and shiny* (Chapman and Hall/CRC, 2020).

50. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).

51. Xiang, F. *et al.* Antibody detection and dynamic characteristics in patients with covid-19. *Clin. Infect. Dis.* (2020).

52. Flodgren, G. Immunity after sars-cov-2 infection, 1st update–. *Lancet Infect. Dis.* **23**, 23 (2020).

53. Westbury, C. Bayes' rule for clinicians: An introduction. *Front. Psychol.* **1**, 192, DOI: 10.3389/fpsyg.2010.00192 (2010).

54. Weisstein, E. W. Total probability theorem. https://mathworld.wolfram.com/TotalProbabilityTheorem.html. Online: Accessed 05-16-2020.

55. Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M. & Woloshin, S. Helping doctors and patients make sense of health statistics. *Psychol. science public interest* **8**, 53–96 (2007).

56. Trevethan, R. Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. *Front. Public Heal.* **5**, 307, DOI: 10.3389/fpubh.2017.00307 (2017).

7 Acknowledgements

The authors ackowledge support from the National Cancer Institute (5P30 CA076292-22) and the Biostatistics and Bioinformatics Core at the H. Lee Moffitt Cancer Center & Research Institute.

8 Additional information

Competing interests

Dr. Brownstein (NCB) served as an ad-hoc reviewer in 2020 for the American Cancer Society, for which she received sponsored travel during the review meeting and a stipend of $300. NCB received a series of small awards for conference and travel support, including $500 from the Statistical Consulting Section of the American Statistical Association for Best Paper Award at the 2019 Joint Statistical Meetings and the $500 Lee Travel Award from the Caucus for Women in Statistics to support attendance at the 2018 Joint Statistical Meetings. NCB also received a Michael Kutner/ASA Junior Faculty Travel Award of $946.60 to attend the 2018 Summer Research Conference of the Southern Regional Council on Statistics and travel support of $708.51 plus a registration waiver from the American Statistical Association (ASA) to attend and chair a session for the 2017 Symposium on Statistical Inference. Both authors (NCB and YAC) were supported by the the National Cancer Institute (5P30 CA076292-22) and the Biostatistics and Bioinformatics Core at the H. Lee Moffitt Cancer Center & Research Institute. NCB serves as the Florida Chapter Representative for the ASA and as the mentoring subcommittee chair for the Regional Advisory Board (RAB) of the Eastern North American Region (ENAR) of the International Biometrics Society (IBS). NCB was also recently elected as Section Representative for the ASA Statistical Consulting Section for 2021-2023. YAC is a member of the International Society for Computational Biology and American Association for Cancer Research. Both authors are members of the ASA and serve on the SRC at Moffitt Cancer Center.

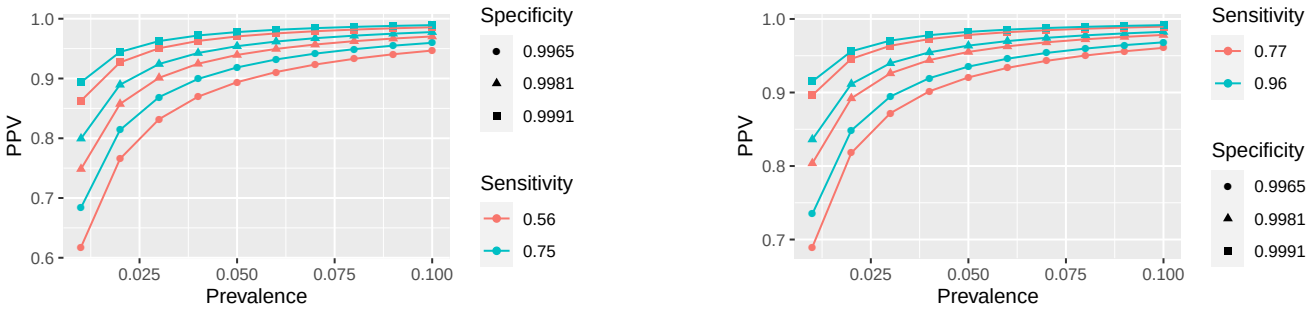
9 Supplementary Information

We include three appendices. Section 9.1 reviews Effect of Time from Infection on antibodies test results and includes a analysis of one of the FDA tests for individuals tested too early, i.e. less than two weeks after initial infection. Section 9.3 includes intermediate calculations and commentary on the difference between the test positivity rate and estimated prevalence.

9.1 Effect of Time from Infection

Seroconversion is the process during which antibodies develop after infected by Covid19 become detectable in the blood. The seroconversion duration could complicate the consideration of interpretation of serology test results. After infection, patients took about 3-40 days to develop detectable antibodies[51,52]. Typically after 14 days, most patients will develop antibodies. If the testing period is within 14 days, the sensitivities of the tests will be lower.

We provide an example of the effect of test timing on PPV. As noted in section 3.3 of the main text, the test by Roche pharmaceuticals[34], reported relatively high PPV compared to its competitors (Table 1 and Figure 4), with 95% confidence intervals of (88%,100%) for specificity and (99.61%, 99.91%) for sensitivity when the test is taken at least 14 days after infection. However, when the test is given early, i.e., within 6 days, the sensitivity is only about 65.5 % (56.1 – 74.1 %). When the test is given between 7-13 days, the sensitivity is 88.1 % (77.1 – 95.1 %). Figure 7 displays the PPV by prevalence.



(a) 0-6 days **(b)** 7-13 days

Figure 7. PPV by prevalence for Roche test taken prior to 2 weeks post infection

9.2 Definitions and equations

For the illustrative purposes of this paper, we include simplifying assumptions. First, we assume that people who have not yet had a SARS-Cov-2 infection (with or without symptoms) should lack antibodies. Conversely, we assume that a person who does not have antibodies either has not yet been infected with SARS-CoV-2 or has been infected too recently for the body to have developed antibodies. Limitations and further discussion on these items are included in Sections 5 and 9.1, and external sources[25,52] .

Statistical notation for these definitions follow. For a randomly selected person, consider the following random variables. D is the indicator of whether the person was previously infected with novel coronavirus disease (SARS-CoV-2) and has antibodies. T is the indicator of if the antibodies test result is positive. The prevalence of SARS-Cov-2 infection and antibodies posession is $p = p(D = 1)$. PPV and NPV are defined by equations (3) and (4):

$$sensitivity = P(T = 1|D = 1) \tag{1}$$

$$specificity = P(T = 0|D = 0) \tag{2}$$

$$PPV = P(D = 1|T = 1) \tag{3}$$

$$NPV = P(D = 0|T = 0) \tag{4}$$

One can invoke Bayes Rule[53] and the law of total probability[54] to show that;

$$P(T=1) = p * sensitivity + (1-p) * (1-specificity) \tag{5}$$

$$PPV = \frac{p * sensitivity}{p * sensitivity + (1-p) * (1-specificity)} \tag{6}$$

$$NPV = \frac{(1-p) * specificity}{(1-p) * specificity + p * (1-sensitivity)} \tag{7}$$

Although sensitivity and specificity clearly differ from PPV and NPV, respectively, these distinct quantities are often misinterpreted in practice[55,56]. Importantly, the PPV and NPV of serology tests depend on pretest parameters and on the prevalence of SARS-CoV-2, which is unknown and difficult to measure[29]. PPV and NPV correspond to probabilities that test results of each type are correctly classified. Their complements refer to false testing ratse. False positives refer to positive serology tests for patients lacking antibodies, while false negatives refer to negative tests for patients with antibodies.

$$FPR = 1 - PPV = P(D=0|T=1) \tag{8}$$

$$FNR = 1 - NPV = P(D=1|T=0) \tag{9}$$

9.3 Additional Calculations and Application to Serology Studies

The studies in New York and Chelsea provided only positive testing rates, not prevalence estimates. The positive testing rate is not the same as the prevalence. We conduct sensitivity analyses after including initial calculations. Results were generally similar. Often the estimated prevalence (and PPV) was slightly lower than the seropositivity rate.

First, we break down the probability of testing positive in equation (10).

$$\begin{aligned} P(T=1) &= P(T=1 \cap D=1) + P(T=1 \cap D=0) \\ &= P(T=1|D=1)P(D=1) + P(T=1|D=0)P(D=0) \\ &= P(T=1|D=1)P(D=1) + [1-P(T=0|D=0)][1-P(D=1)] \\ &= p * sensitivity + (1-p) * (1-specificity) \\ &= p * (sensitivity + specificity - 1) + (1-specificity) \end{aligned} \tag{10}$$

The proportion testing positive in a seroprevalence study can be used to estimate the true prevalence. Denote $\hat{p}_t$ as the observed proportion who have have antibodies. We can substitute $\hat{p}_t$ on the left hand side of (10) as an estimate for $P(T=1)$ and solve for the prevalence $\hat{p}$. The result is equation (11):

$$\begin{aligned} \hat{p} &= \frac{\hat{p}_t - (1-specificity)}{sensitivity + specificity - 1} \\ &= \frac{specificity + \hat{p}_t - 1}{sensitivity + specificity - 1} \\ &= \frac{specificity - (1-\hat{p}_t)}{sensitivity + specificity - 1} \end{aligned} \tag{11}$$

It is important to note that this means not every combination of sensitivity and specificity is possible for a given positive testing rate, as otherwise the estimated prevalence will be negative[37]. Given that sensitivity and specificity are usually large (e.g. each exceeding 80%), we would expect

$$sensitivity + specificity - 1 > 0$$

In order for the prevalence estimate to be non-negative, this means that the numerator must also be non-negative, and we would expect

$$\hat{p}_t \geq (1-specificity) \tag{12}$$

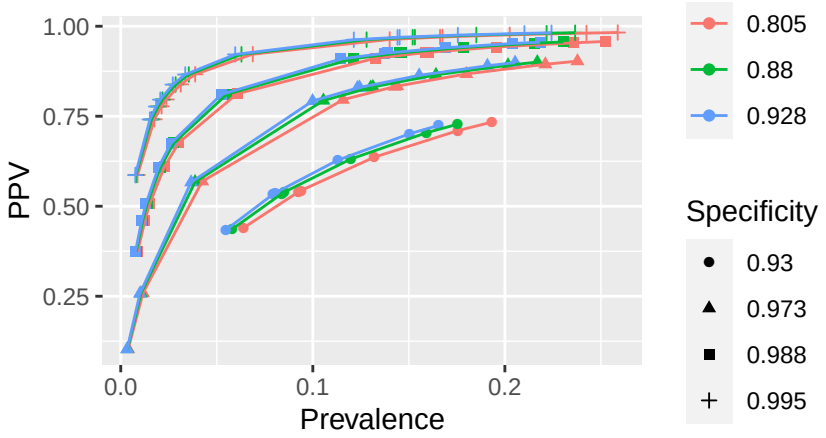


Figure 8. Sensitivity Analysis of PPV by prevalence calculated by Equation (11)

That is, the positive rates observed in a seroprevalence study, $\hat{p}_t$ should be at least be large as the false positive rate of the antibody test used in that study. Indeed, if we saw fewer positive tests than the proportion of (false) positive tests expected by chance if everyone were lacking antibodies, then we would have little evidence to suggest that the prevalence is nominally different from zero. A sensitivity analysis follows for the studies in Chelsea and New York, which reported only seropositivity estimates. While the seropositive rates are not necessarily identical to the prevalence estimate, the estimated seropositivity rate was generally either close or a slight overestimate of the prevalence.

9.3.1 Chelsea

In Chelsea, the original seropositivity rate was 31.5%. The updated prevalence estimates based on (11) was 27.9%, which is about 3.6% lower than the proportion who tested positive. Using this lower prevalence yields a lower PPV or 78.6%, which would correspond to about to expecting about 49 of the positive tests to be true positives and 14 to be false positives.

9.3.2 New York

The sensitivity analysis In the New York serology study presented in Section 4, there were 156 estimates of PPV, resulting from the product of 3 possible values for sensitivity, 4 potential values for specificity, and 13 potential seropositivity estimates from the two studies[38,40].

Prevalence estiamtes were calculated based on (11). Unlike in the sensitivity analysis for Chelsea, 36 scenarios resulted in negative estimated prevalence values. For these combinations, the estimated PPV would be zero, which is lower than the PPV in the paper. The false positive rate in the sensitivity analysis would be even higher in the sensitivity analysis than in the paper.

For the 120 remaining scenarios with estimated positive prevalence values, the difference between the prevalence and positivity rate was generally small and centered close to zero. Among the 120 scenarios, the differences between the two estimated proportions ranged from -6.2% to 4.7% with a median difference of -0.3% and a mean difference of 0.6%. While the distribution of differences in prevalence was approximately symmetric, the distribution of differences in PPV was strongly left-skewed. The median difference in PPV was -0.96% and mean difference was -5.44%, with a range from -41% to 1.4%. The large differences in absolute value of PPV correspond to values where the PPV in this sensitivity analysis is much lower than the PPV when using seropositivity to estimate prevalence. However, most differences were small in absolute value, meaning that most PPV estimates were similar regardless of which value was used for prevalence.

An updated plot of PPV by prevalence estimated with (11) is shown in Figure 8. The shape is similar to Figure 6, with noticeable missing segments for combinations with the lower bound of specificity and prevalence estimates outside of the range from about 5% to 20%. Small prevalence and specificity estimates likely violate Expression (12).

9.4 Comparison with Other Types of Tests for SARS-Cov-2

It is important to note the difference in analysis and interpretation in this paper, compared to other tests for SARS-COV-2. Our paper showed that NPV was reasonably high and PPV was low, with a lot of potential for harm, such as increasing risk for Covid19. The interpretations differ for diagnostic tests. False positives for diagnostic tests would mean that an uninfected patient would be quarantined and their contacts tested. False negative diagnostic tests would mean that an infected person could be cleared, their future contacts put at risk of exposure, and their past contacts less likely to be tested. Thus, the potential harm for a false negative likely exceeds the potential harm for a false positive, and NPV should be emphasized to mitigate this harm.

While a rigorous analysis for other types are outside of the scope of this paper, we will provide some comments. We showed that PPV was closely related to specificity. Similarly, NPV is related closely to sensitivity. Thus, it is imperative that diagnostic tests for SARS-Cov-2 have high sensitivity. It is unclear whether this is true in practice, as there have been reports and analyses of diagnostic tests having high false negatives[45].