Are SARS-CoV-2 seroprevalence estimates biased?

Saki Takahashi, PhD¹, Bryan Greenhouse, MD, MA^{1,2}, Isabel Rodríguez-Barraquer, MD, PhD¹

¹EPPIcenter Program; Division of HIV, ID, and Global Medicine; Department of Medicine,
University of California, San Francisco, San Francisco, CA, USA

²Chan Zuckerberg Biohub, San Francisco, CA, USA

Corresponding author:

Isabel Rodríguez-Barraquer

University of California, San Francisco

Division of HIV, ID, and Global Medicine

995 Potrero Avenue, Ward 84

San Francisco, CA 94110

isabel.rodriquez@ucsf.edu

Abstract: Growing evidence suggests that asymptomatic and mild SARS-CoV-2 infections, together comprising >95% of all infections, may be associated with lower antibody titers than severe infections. In addition, antibody levels peak a few weeks after infection and decay gradually. Yet, positive controls used for determining the sensitivity of serological assays are usually limited to samples from hospitalized patients with severe disease, leading to what is commonly known as spectrum bias in estimating seroprevalence in the general population. We performed simulations to quantify the bias potentially introduced by the choice of positive controls used. Our results suggest that assays with imperfect sensitivity will underestimate the true seroprevalence, but this can be corrected if assay sensitivity in the general population is known. If sensitivity is determined from validation sets skewed towards those with severe or recent infections and thus higher antibody levels, corrected prevalence will still underestimate the true prevalence. Correct interpretation of SARS-CoV-2 seroprevalence studies requires quantifying the extent to which the sensitivity of assays being used varies with disease severity and over time. Optimization and validation of serological assays should involve samples from across the spectrum of severity and time since infection, and performance characteristics should be stratified by these factors.

Key words: COVID-19; SARS-CoV-2 seroprevalence; antibody testing; test performance; spectrum bias

Serosurveys are needed to understand how many people have been infected by SARS-CoV-2 and where we are in the epidemic curve. Unprecedented serosurveillance efforts have been launched (e.g., Solidarity II, NIH) to generate local and global infection estimates to guide social distancing policies. Recently published results suggest that the proportion of the population that has been infected, even in places with explosive outbreaks such as Spain or New York City, is low and far from the levels required for herd immunity¹⁻³.

Multiple serological assays and rapid tests are now available and, as of May 22nd, the Food and Drug Administration has authorized the use of thirteen⁴. Assay validation requires samples from individuals with known infection status in order to determine test performance characteristics (i.e., sensitivity and specificity). Due to potential cross-reactivity of antibody responses to seasonal coronaviruses, much of the focus of assay development has been on ensuring near perfect specificity, to minimize the risk of false positive results. This is particularly important during early stages of the epidemic, when the number of true positives is expected to be very low. However, if the purpose of deploying a serological assay is to quantify the proportion of the population that has been infected by SARS-CoV-2 (i.e., serosurveillance), adequate characterization of assay sensitivity to detect prior infection in the general population is important as well. We raise this issue because in the absence of such characterization, it will not be possible to generate accurate estimates of population-level exposure to this novel pathogen⁵.

Growing evidence suggests that asymptomatic and mild SARS-CoV-2 infections, together making up over 95% of all infections, may be associated with lower antibody titers than more severe infections⁶⁻¹⁰. Similarly, it is known that antibody levels peak a few weeks after infection and then decay gradually^{11,12}. Yet, positive controls used for assay optimization and validation are usually limited to samples from hospitalized patients with severe disease, leading to what is

commonly known as spectrum bias. Sensitivities estimated from these sample sets may therefore overestimate the actual sensitivity that the assay would have when applied to the general population, leading to underestimates of the true seroprevalence.

To illustrate this point, we quantified the amount of bias in estimating population seroprevalence potentially introduced by the choice of positive controls used to evaluate assay sensitivity. We evaluated the impact of: (a) using validation sets with different proportions of severe, mild, and asymptomatic infections (Figure panel A), and (b) using validation sets with samples from recent infections only (Supplementary Figure). While the relationship between disease severity and assay sensitivity has yet to be quantified for most assays, we assumed that sensitivity was highest for severe infections and considered a range of values for asymptomatic and mild infections. Similarly, we assumed that sensitivity peaked early after infection and then decayed over time. For simplicity, we fixed test specificity at 100%.

Assays with imperfect sensitivity lead to underestimates of the true seroprevalence (Figure panel B, grey line), but can be easily corrected for if the actual sensitivity of the assay in the sampled population is known (Figure panel B, purple line)¹³. However, if test sensitivity has been determined from positive control sets skewed towards those with severe clinical outcomes (high antibody levels), the measured prevalence, even after correction, will still underestimate the true prevalence (Figure panel B, cyan and gold lines). The magnitude of the underestimate will depend on how biased the distribution of positive controls is relative to the population, and on how much assay sensitivity varies with disease severity (Figure panel C). Similarly, corrected prevalence will only equal the true prevalence if decreases in sensitivity due to waning antibody responses over time can be accounted for (Supplementary Figure).

These results have important implications for assay development and for the interpretation of SARS-CoV-2 seroprevalence studies. First, they highlight the need to quantify the extent to which the sensitivity of the assays used in ongoing serosurveillance studies varies with disease severity and over time. Incorporating loss of sensitivity with increasing time since infection will gain importance as the pandemic progresses. More importantly, these results highlight the need for detailed studies characterizing kinetics of antibody responses to SARS-CoV-2 across the severity spectrum. If antibody responses are significantly lower in milder cases, or if there is significant waning in the months following infection, assays for seroprevalence studies should be optimized to detect these lower titers. Finally, these results caution against accepting aggregate sensitivities and specificities reported by assay manufacturers at face value. Ideally, sensitivities and specificities should be stratified by disease severity and time since infection, and the characteristics of the validation set should be reported at a minimum.

Correct interpretation of SARS-CoV-2 seroprevalence studies will not be possible until we know the sensitivity of serological assays to detect mild and asymptomatic infections and infections occurring months beforehand. We urge the FDA to revisit reporting requirements for performance characteristics of SARS-CoV-2 serological assays with this in mind.

Author contributions: ST, BG, and IRB conceived the study, performed the simulations, and wrote the manuscript.

Declaration of interests: We declare no competing interests.

Role of the funding source: ST was supported by the Schmidt Science Fellows, in partnership with the Rhodes Trust. The funders had no role in the writing of the manuscript or the decision to submit it for publication.

References

- 1 de España G. Estudio ENE-COVID19: Primera Ronda.
https://www.ciencia.gob.es/stfls/MICINN/Ministerio/FICHEROS/ENECOVID_Informe_preliminar_cierre_primera_ronda_13Mayo2020.pdf.
- 2 Stringhini S, Wisniak A, Piumatti G, *et al.* Repeated seroprevalence of anti-SARS-CoV-2 IgG antibodies in a population-based sample from Geneva, Switzerland. *Infectious Diseases (except HIV/AIDS)*. 2020; published online May 6.
DOI:10.1101/2020.05.02.20088898.
- 3 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. 2020; published online May 2. <https://www.governor.ny.gov/news/amid-ongoing-covid-19-pandemic-governor-cuomo-announces-results-completed-antibody-testing> (accessed May 16, 2020).
- 4 Center for Devices, Radiological Health. EUA Authorized Serology Test Performance. U.S. Food and Drug Administration. 2020. <https://www.fda.gov/medical-devices/emergency-situations-medical-devices/eua-authorized-serology-test-performance> (accessed May 14, 2020).
- 5 Krammer F, Simon V. Serology assays to manage COVID-19. *Science* 2020; published online May 15. DOI:10.1126/science.abc1227.
- 6 Whitman JD, Hiatt J, Mowery CT, *et al.* Test performance evaluation of SARS-CoV-2 serological assays. *Infectious Diseases (except HIV/AIDS)*. 2020; published online April 29.
DOI:10.1101/2020.04.25.20074856.
- 7 Okba NMA, Muller MA, Li W, *et al.* SARS-CoV-2 specific antibody responses in COVID-19 patients. *Infectious Diseases (except HIV/AIDS)*. 2020; published online March 20.
DOI:10.1101/2020.03.18.20038059.

- 8 Wu F, Wang A, Liu M, *et al.* Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered patient cohort and their implications. *Infectious Diseases (except HIV/AIDS)*. 2020; published online April 6. DOI:10.1101/2020.03.30.20047365.
- 9 Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. DOI:10.25561/77482.
- 10 Cervia C, Nilsson J, Zurbuchen Y, *et al.* Systemic and mucosal antibody secretion specific to SARS-CoV-2 during mild versus severe COVID-19. *bioRxiv*. 2020; : 2020.05.21.108308.
- 11 Sethuraman N, Jeremiah SS, Ryo A. Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA* 2020; published online May 6. DOI:10.1001/jama.2020.8259.
- 12 Rosado J, Cockram C, Merkling S, *et al.* Serological signatures of SARS-CoV-2 infection: Implications for antibody-based diagnostics. 2020; published online May 11. <https://hal-pasteur.archives-ouvertes.fr/pasteur-02569149/document> (accessed May 14, 2020).
- 13 Diggle PJ. Estimating Prevalence Using an Imperfect Test. *Epidemiology Research International*. 2011; **2011**: 1–5.

Figure:

Figure Legend: **(A)** True distribution of clinical outcomes observed among SARS-CoV-2 infected individuals (left bar), classified as asymptomatic (blue), mild (green), or severe (red), and three simulated sets of positive controls used in assay validation. **(B)** For a range of true prevalence (x-axis), we calculated the prevalence that would be measured in the population (y-axis) under the scenarios of no correction for test characteristics (grey line), and corrections for test sensitivity as determined from each validation set (purple, gold, and cyan lines). We assumed values of symptom-dependent sensitivities (Se) as shown in the panel and a specificity of 100%, leading to crude sensitivities of 53.15% (set 1), 76.5% (set 2), and 95% (set 3) in the validation sets. For all of these scenarios, the true sensitivity in the general population would be 53.15%. **(C)** For a range of sensitivities in mild (x-axis) and asymptomatic infections (assuming sensitivity in asymptomatic is $\frac{2}{3}$ of the sensitivity in mild), we calculated the ratio of measured to true prevalence (y-axis) for each validation set used for correction. The scenarios in panel B are highlighted with the corresponding color.

Supplementary Information

Extended Methods: In validation set 1 described in the main text, the distribution of clinical outcomes in the positive controls is equal to that in the general population (with assumed values of 43% asymptomatic, 52% mild, 5% severe) as previously reported. In validation set 2, the distribution is skewed towards severe clinical outcomes (5% asymptomatic, 45% mild, 50% severe). Validation set 3 represents the extreme scenario, common practice in assay development, of including only positive controls who had severe clinical outcomes. From these sets of positive controls, the crude sensitivity of an assay is calculated as the mean of the test sensitivities of the clinical outcomes (assumed in panel B of the main figure to be 40% in asymptomatic, 60% in mild, and 95% in severe), weighted by the proportion of the positive control set that is of that clinical outcome. For example, the crude sensitivity of validation set 3 is 95%. However, the actual sensitivity of an assay when applied to the general population will be the mean of the test sensitivities of the clinical outcomes, now weighted by the proportion of the general population that is of that clinical outcome. In practice, calculating this actual sensitivity requires knowing the severity-dependent test sensitivities. A measured prevalence corrected for test characteristics can be calculated using the equation $(p_{obs} + Sp - 1) / (Se + Sp - 1)$, where p_{obs} is the observed proportion of positive test results, Se is the test sensitivity, and Sp is the test specificity (assumed to be 100%). The corrections to measured prevalence in panel B are based on using the crude sensitivities. In panel C of the main figure we calculated the ratio of the measured prevalence to the true prevalence value; this ratio is constant across all values of true prevalence, and the slope of each line in panel B corresponds to the y-value of each highlighted point in panel C. In the supplementary figure below we extend the analysis to account for the issue of time-dependent antibody test sensitivity for estimating SARS-CoV-2 seroprevalence.

Supplementary Figure:

Supplementary Figure Legend: **(A)** Epidemic curve from an SEIR transmission model with SARS-CoV-2-like parameters in a completely susceptible population, with mean latent period of 3.5 days, mean infectious period of 7 days, and R_0 of 2.1. Here, we simulated cross-sectional seroprevalence surveys performed at 60, 180, and 300 days into the epidemic (highlighted points on the curve). **(B)** The proportion of the ever-infected individuals that experienced infection within 0-60 days (black), 60-180 days (medium grey), and 180+ days (light grey) of the surveys performed at days 60, 180, and 300. The last bar shows a simulated set of positive controls used in assay validation, which includes only recent samples taken within 0-60 days after infection. We assume for these simulations that the validation controls - and all infections - represent severe clinical infections with a baseline test sensitivity of 95%. **(C)** For true prevalence values ranging from 0 to 50% (x-axis), we calculated the prevalence values that would be measured in the population at days 60, 180, and 300 (y-axis) (assuming sensitivity (Se) is reduced to 80% of the baseline level in infections that are 60-180 days old, and to 60% of the baseline level in infections that are 180+ days old, and test specificity of 100%),

correcting for test characteristics using the positive control set as described in panel B. We used the same procedure as described in the Extended Methods section to calculate the measured prevalences. We also considered prevalence values that would be measured in the population if the two sources of spectrum bias (both clinical outcomes and time since infection) are present. We calculated the prevalence that would be measured at day 180, now assuming that the distribution of true clinical outcomes and their test sensitivities is equal to that in the main figure (filled squares) (i.e., in terms of severity: 43% asymptomatic, 52% mild, 5% severe, and in terms of sensitivity: 95% in severe, 60% in mild, and 40% in asymptomatic), and that the positive controls are all severe, recent infections. The two sources of spectrum bias both contribute to the under-estimation of the true prevalence.

Code Availability: All R code used for this analysis is available at

<https://github.com/sakitakahashi/COVID-sensitivity>.