



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/191877/>

Version: Published Version

---

**Article:**

Kritsotakis, E.I. (2020) On the importance of population-based serological surveys of SARS-CoV-2 without overlooking their inherent uncertainties. *Public Health in Practice*, 1. 100013. ISSN: 2666-5352

<https://doi.org/10.1016/j.puhip.2020.100013>

---

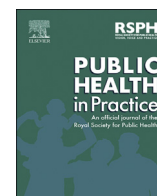
**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



## Commentary

# On the importance of population-based serological surveys of SARS-CoV-2 without overlooking their inherent uncertainties



Evangelos I. Kritsotakis

School of Medicine, University of Crete, 71003, Heraklion, Crete, Greece

## ARTICLE INFO

## Keywords:

Seroepidemiology  
Sero-prevalence  
Random sampling  
Diagnostic validity  
Antibodies  
Coronavirus

## ABSTRACT

The SARS-CoV-2 epidemic has caused an unprecedented public health situation and more than ever it is important to be well informed on methods to monitor and analyse the progression of the epidemic. This brief note aims to explain the scope in conducting large-scale serological surveys of SARS-CoV-2 to define the landscape of population immunity, without overlooking the inherent uncertainty stemming from sampling design and diagnostic validity. The note completes with a succinct appendix of simple statistical methods for estimating prevalence from random population samples using imperfect diagnostic tests.

## 1. The problem

To date we know little about the SARS-CoV-2 virus spread into the general population. Our great uncertainty stems from the fact that the virus spreads easily between people but many COVID-19 infections are mild or subclinical [1] and therefore go unnoticed. The actual number of people already exposed to SARS-CoV-2 may be much higher than the number of confirmed COVID-19 patients who have been seriously ill and/or tested positive for SARS-CoV-2. Most experts would agree that it is reasonable to assume that we are at least 10 times off in reported numbers, but a recent report suggests that the actual number of infections may be as much as 85 times higher than that reported [2].

From a public health standpoint, knowing how many and who have already been exposed to SARS-CoV-2 gives a clearer picture of how widespread the virus is in local populations. This is extremely useful because public health measures depend on how far Coronavirus has already penetrated into the general population. In the absence of precise estimates from a random sample of the general population, we are essentially operating in the dark and likely to continue taking restrictive measures without being able to assess their effectiveness.

## 2. Sero-prevalence surveys

Population-based serological surveys can generate much needed data [3]. They use serological tests to examine a large number of blood samples from people without a confirmed SARS-CoV-2 infection to detect signs that they were once infected with the virus. That is, serological tests

detect our body's response to the virus but not the virus itself (as opposed to molecular tests). Therefore, they cannot be used early in infection before the patient's body has already developed an antibody response. Thus, serological tests are not much helpful for clinicians to diagnose infection in individual persons. However, they are extremely useful for epidemiological purposes to understand the immunity landscape of the population at large.

Estimating the true rate of SARS-CoV-2 infection allows epidemiologists to predict the likely future course of the epidemic in specific locations or populations and helps public health authorities to better design interventions to control the epidemic. This is because we expect, although no one is entirely certain yet, that once we have antibodies to the virus, they will provide us with immunity, that is, we will be protected for some period of time. Detecting people who are potentially immune to SARS-CoV-2 could even play an important role in when and how social distancing restrictions are lifted. The results of serological surveys can also be useful in guiding strategic decisions on essential staffing in hospitals and other health care facilities - for example, by assigning to the forefront those who are probably immune. It is therefore desirable to conduct targeted serological studies of healthcare workers.

## 3. Inherent uncertainties

The results of serological surveys come with uncertainty, but it is important to note that this can be assessed. Uncertainty stems from two main sources: (a) sampling variability, that is, from the fact that we examine only a small part of the overall population, and (b) diagnostic

E-mail address: [e.kritsotakis@uoc.gr](mailto:e.kritsotakis@uoc.gr).

<https://doi.org/10.1016/j.puhip.2020.100013>

Received 8 May 2020; Accepted 16 May 2020

Available online 22 May 2020

2666-5352/© 2020 The Author. Published by Elsevier Ltd on behalf of The Royal Society for Public Health. This is an open access article under the CC BY-NC-ND

license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

validity, that is, imperfect accuracy of the immunoassay test in detecting the presence or the absence of antibodies. Therefore, it is critical that serological surveys are based on both appropriate sampling designs assuring population representation and accurate serological tests. Due to urgency and demand, several serological tests have been developed and placed on the market recently. Manufacturer's own data [4] and independent evaluations [5] indicate that accurate enough tests are currently available: their probability of successfully detecting people exposed to SARS-CoV-2 (sensitivity) exceeds 90% a few days after the infection and their success in detecting non-infected individuals (specificity) reaches 99%.

#### 4. An example

Available serological tests are not perfect but are acceptable for use in the context of surveying populations for SARS-CoV-2 antibodies, because survey estimates can be corrected for imperfect diagnostic performance. For example, let us assume that a serological survey of  $n = 1,000$  people found that  $a = 100$  are positive for SARS-CoV-2 antibodies, meaning that  $P_A = a/n = 10\%$  were infected. The test used was imperfect, say with known sensitivity  $S_e = 92\%$  and specificity  $S_p = 98\%$ , but we can correct

#### Statistical Appendix

This appendix provides a summary of simple methods to estimate prevalence using imperfect diagnostic tests.

Assume that the prevalence of infection ( $\pi_T$ ) in the target population is a fixed, but unknown quantity. To estimate  $\pi_T$ , we do a diagnostic test on  $n$  randomly sampled individuals from the target population and  $a$  individuals test positive. However, the test is imperfect, with sensitivity ( $S_e$ ) and/or specificity ( $S_p$ ) that are below 100%. Thereby, the apparent prevalence  $P_A = a/n$  is a biased estimate of  $\pi_T$ .

Let  $P_T$  denote the true prevalence proportion that we would observe if the diagnostic test was perfect. It is easy to confirm that the apparent prevalence  $P_A = a/n$  and the true prevalence  $P_T$  are related by:

$$P_T = \frac{P_A + S_p - 1}{S_e + S_p - 1}$$

$P_T$  is known as the Rogan–Gladden-estimator. [6] Assuming  $S_e$  and  $S_p$  are known with certainty,  $P_T$  is an unbiased estimate of the true population prevalence  $\pi_T$ . It is also a maximum likelihood estimate of  $\pi_T$ . [7] Note that  $P_T$  is meaningful under the reasonable requirement that the diagnostic test is better than the flip of a coin ( $S_e + S_p > 1$ ). Nevertheless,  $P_T$  is not guaranteed to lie between 0 and 1 (especially when  $P_A$  is very small) and a “clipped” estimate may need to be used:  $P_{TC} = \min[\max(P_T, 0), 1]$ .

The standard error of  $P_T$  is:

$$SE(P_T) = \frac{SE(P_A)}{S_e + S_p - 1}$$

where  $SE(P_A)$  depends on the sampling design used. For a simple random sample from a large population:

$$SE(P_A) = \sqrt{\frac{P_A(1 - P_A)}{n}}$$

For large  $n$ , the statistic  $(P_T - \pi)/SE(P_T)$  can be treated as a standard normal variate. Thus, an approximate 95% confidence interval for  $\pi$  is obtained as:

$$P_T \mp 1.96 \cdot SE(P_T)$$

The “clipped” estimate  $P_{TC} = \min[\max(P_T, 0), 1]$  is asymptotically equivalent to  $P_T$ , [8] so the large sample theory is valid in that case too.

Essentially, for fixed  $S_e$  and  $S_p$ , a 95% confidence interval  $[l, u]$  for the apparent prevalence  $\pi_A$ , can be converted to a 95% confidence interval for the true prevalence  $\pi_T$  by

$$\left[ \frac{l + S_p - 1}{S_e + S_p - 1}, \frac{u + S_p - 1}{S_e + S_p - 1} \right]$$

Consequently in situations where asymptotic assumptions are not met (e.g. small sample size and/or very low prevalence), exact methods (e.g. Clopper-Pearson) can be applied to calculate confidence limits for the apparent prevalence that can be converted to confidence limits for the true prevalence using the formula above. [9]

If  $S_e$  and  $S_p$  are not known with certainty, but independent binomial estimates are available from a validation study on persons whose infection status is known, then  $P_T$  is biased but to a much lesser degree than  $P_A$  [6]. In that case, a more valid quantification of standard error that captures the uncertainty in  $S_e$  and  $S_p$  is given by:

our estimate for these inaccuracies. The corrected estimate of the true prevalence of SARS-CoV-2 turns out to be  $P_T = 8.9\%$ . We can express the uncertainty associated with this estimate using a 95% confidence interval, which in this case is from 6.7% to 11.1%. In this way, we get a fairly precise idea of the extent of the virus spread into the population.

#### 5. Conclusion

Large-scale seroprevalence surveys are an important tool in combating COVID-19 disease as they can provide much-needed estimates of the fraction of the population with antibodies against SARS-CoV-2. The quality of the antibody prevalence estimates depends on the sampling design and the diagnostic accuracy of serological tests.

#### Declaration of competing interest

The author is an Editorial Board member of Public Health in Practice. The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

$$SE(P_T) = \frac{1}{S_e + S_p - 1} \sqrt{\left[ SE(P_A) + \frac{S_e(1 - S_e)}{n_1} P_T^2 + \frac{S_p(1 - S_p)}{n_2} (1 - P_T)^2 \right]}$$

where  $n_1$  and  $n_2$  denote the numbers of infected and non-infected individuals in the validation study. [6] A double sampling design that partly utilises a more definitive diagnostic test can also be used [10]. Using a binomial distribution model for the number of positive tests  $a$  out of the  $n$  individuals tested, a Bayesian approach may also be used to estimate  $\pi_T$  that does not yield explicit formulae but is computationally easy [11,12].

## References

- [1] D. Wu, T. Wu, Q. Liu, Z. Yang, The SARS-CoV-2 outbreak: what we know, *Int. J. Infect. Dis.* 94 (2020) 44–48.
- [2] E. Bendavid, et al., COVID-19 antibody seroprevalence in Santa Clara County, California, *medRxiv* 2020 (2020), <https://doi.org/10.1101/2020.04.14.20062463>, 04.14.20062463.
- [3] C.J.E. Metcalf, et al., Use of serological surveys to generate key insights into the changing global landscape of infectious disease, *Lancet* 388 (2016) 728–730.
- [4] E. Cairns, Covid-19 antibody tests face a very specific problem, *Evaluate Vantage* (2020).
- [5] P.I. Kontou, G.G. Braliou, N.L. Dimou, G. Nikolopoulos, P.G. Bagos, Antibody tests in detecting SARS-CoV-2 infection: a meta-analysis, *medRxiv* 2020 (2020), <https://doi.org/10.1101/2020.04.22.20074914>, 04.22.20074914.
- [6] W.J. Rogan, B. Gladen, Estimating prevalence from the results of a screening test, *Am. J. Epidemiol.* 107 (1978) 71–76.
- [7] P.S. Levy, E.H. Kass, A three-population model for sequential screening for bacteriuria2, *Am. J. Epidemiol.* 91 (1970) 148–154.
- [8] J.L. Gastwirth, The statistical precision of medical screening procedures: application to polygraph and AIDS antibodies test data, *Stat. Sci.* 2 (1987) 213–222.
- [9] J. Reiczgel, J. Földi, L. Ózsvári, Exact confidence limits for prevalence of a disease with an imperfect diagnostic test, *Epidemiol. Infect.* 138 (2010) 1674–1678.
- [10] P.S. Levy, S. Lemeshow, *Sampling of Populations: Methods and Applications*, fourth ed., John Wiley & Sons, Inc., 2011 <https://doi.org/10.1002/9780470374597>. *Sampling of Populations: Methods and Applications: Fourth Edition*.
- [11] P.J. Diggle, Estimating prevalence using an imperfect test, *Epidemiol. Res. Int.* 1–5 (2011), 2011.
- [12] F.I. Lewis, P.R. Torgerson, A tutorial in estimating the prevalence of disease in humans and animals in the absence of a gold standard diagnostic, *Emerg. Themes Epidemiol.* 9 (2012) 9.