

Qualifying Paper Report for Adjusting COVID-19 Seroprevalence Survey Results to Account for Test Sensitivity and Specificity

Naitong Chen

March 7, 2023

1 Summary

Over the past few years, tracking the spread of COVID-19 has been crucial to developing of a scientific understanding the disease, which ultimately guided public health policies aimed at controlling the spread of the disease across the world. As with any epidemic/pandemic, reported case counts within a defined geographic region is one of the most accessible statistics indicating the scale of the spread of a disease. However, due to testing availability, there may be many individuals in a given region that have been infected but not tested. This means that the total number of infection may be much higher than that reflected from reported case counts (cite).

To more accurately estimate the cumulative number of infections over a period of time, an alternative approach is to conduct population-based seroprevalence studies. To carry out a seroprevalence study for a given region on a given disease, researchers begin by obtaining a sample representative of the population. Antibody tests are then performed for the disease of interest over each individual in the sample. A positive antibody test indicates a previous infection of the disease of interest. Therefore, the proportion of positive tests in the sample can be used as an estimate of the proportion of population previously infected with the disease, which we call the cumulative incidence. With the estimated cumulative incidence, one can estimate the total number of all individuals in the population with a previous infection. It is worth noting, however, that seroprevalence studies cannot identify previous infections whose antibodies are no longer detectable or recent infections that have yet to produce detectable antibodies. At the same time, they also do not include individuals that have died after becoming infected. As a result, a seroprevalence study as we describe it here is only informative about cumulative incidence for the average period over which antibodies are detectable, provided that the disease has a relatively low fatality rate.

We know that COVID-19 has a relatively low fatality rate (cite). We also know that an individual starts to produce detectable antibodies after an average of 25 days since infection, and that the antibodies stay detectable for months after infection (cite). Therefore, cumulative incidences estimated by seroprevalence studies conducted within the first few months of the pandemic can be relatively accurate over the period from the beginning of the pandemic until roughly a month prior to when the samples were taken.

2 Significance

3 Limitations and challenges

4 Paper-specific project

4.1 Data

The datasets used in [Lewin et al. \(2021\)](#) and [Lewin et al. \(2022\)](#) together can be viewed as one for a serial COVID-19 seroprevalence study in Quebec, Canada.

The first dataset contains numbers of antibody-positive samples as well as total number of samples for each region in Quebec (Montreal-Laval, surrounding Montreal-Laval, and other regions). These samples are collected relatively early on in the pandemic, from May 25 to July 9, 2020. The second dataset follows the same structure, but contains samples collected between January 25 and Mar 11, 2021.

In addition, [Lewin et al. \(2022\)](#) also contains results from a seroreversion substudy. Namely, we also have the number of antibody-positive samples from 2020 that remained positive in 2021.

Note that in the second study, the count of antibody-positive samples are available at a finer scale (each of Montreal-Laval, surrounding Montreal-Laval, and other regions are broken down into smaller regions), this is not the case for [Lewin et al. \(2021\)](#). Therefore the analysis will only be conducted at the bigger regional level.

4.2 Project Idea

We can apply the approach in [Meyer et al. \(2022\)](#) to the two datasets in [Lewin et al. \(2021\)](#) and [Lewin et al. \(2022\)](#) in one coherent Bayesian model that accounts for seroreversion.

Let S_{ij} denote the true seroprevalence in study i and region j ($i = 1$ and $i = 2$ correspond to first and second study, $j = 1, j = 2$ and $j = 3$ correspond to Montreal-Laval, surrounding Montreal-Laval, and other regions). Let P_{ij} denote the observed seroprevalences. Let Se, Sp, Sr denote sensitivity of test-kit, specificity of test-kit, as well as proportion of samples that has seroreverted (testing antibody-negative in 2021 but antibody-positive in 2020). Let x_{ij} denote the total number of antibody-positive samples in region j at study i , and n_{ij} denote the total number of samples in region j at study i . Finally, let x_r and n_r denote the total number of samples that seroreverted and the total number of samples in the seroreversion substudy.

The model can then be written as

$$\begin{aligned}
P_{11} &= S_{11} \times Se + (1 - S_{11}) \times (1 - Sp) \\
P_{12} &= S_{12} \times Se + (1 - S_{12}) \times (1 - Sp) \\
P_{13} &= S_{13} \times Se + (1 - S_{13}) \times (1 - Sp) \\
P_{21} &= (S_{21} - Sr \times S_{11}) \times Se + (1 - (S_{21} - Sr \times S_{11})) \times (1 - Sp) \\
P_{22} &= (S_{22} - Sr \times S_{12}) \times Se + (1 - (S_{22} - Sr \times S_{12})) \times (1 - Sp) \\
P_{23} &= (S_{23} - Sr \times S_{13}) \times Se + (1 - (S_{23} - Sr \times S_{13})) \times (1 - Sp) \\
S_{11} &\sim \text{Beta}(\cdot, \cdot) \\
S_{12} &\sim \text{Beta}(\cdot, \cdot) \\
S_{13} &\sim \text{Beta}(\cdot, \cdot) \\
S_{21} &\sim \text{Beta}(\cdot, \cdot)_{\{Sr \times S_{11}, 1\}} \\
S_{22} &\sim \text{Beta}(\cdot, \cdot)_{\{Sr \times S_{12}, 1\}} \\
S_{23} &\sim \text{Beta}(\cdot, \cdot)_{\{Sr \times S_{13}, 1\}} \\
Se &\sim \text{Beta}(\cdot, \cdot) \\
Sp &\sim \text{Beta}(\cdot, \cdot) \\
Sr &\sim \text{Beta}(\cdot, \cdot)
\end{aligned}$$

$$L(P, S, Se, Sp, Sr \mid X, N) \propto \left(\prod_{i=1}^2 \prod_{j=1}^3 P_{ij}^{x_{ij}} (1 - P_{ij})^{n_{ij} - x_{ij}} \right) (Sr^{x_r} (1 - Sr)^{n_r - x_r})$$

Note that for observed seroprevalences in the first study, we adjust for test sensitivity and test specificity (Following the setup in [Meyer et al. \(2022\)](#). We'll likely use the same prior on Se and Sp here). For observed seroprevalences in the second study, we adjust for both test sensitivity and test specificity as well as seroreversion. Namely, we also include the proportion that have seroreverted since the first study using information from the seroreversion substudy. This combines the setup in [Lewin et al. \(2022\)](#) and [Meyer et al. \(2022\)](#): we first subtract the proportion that have seroreverted from the true seroprevalence and then adjust for test-kit performance. This is because the test-kit only has a chance at detecting antibodies if the subject has not seroreverted. To ensure we do not run into negative seroprevalence estimates, we truncate S_{21}, S_{22}, S_{23} accordingly.

Altogether, this can give us an estimate of the seroprevalance in Quebec, Canada in January to March 2021 adjusting for test-kit performance as well as seroreversion.

We can compare the results from this above Bayesian model to those from [Lewin et al. \(2022\)](#) (which is not Bayesian and does not account for test-kit performance) as well a frequentist equivalence of the above Bayesian model using the same data. We can use these results to check if the claims from [Lewin et al. \(2022\)](#) still hold under this different dataset, as well as to explore potential reasons as to why they do or do not hold.

It turns out that the intervals constructed from [Rosenberg et al. \(2020\)](#) (the non-Bayesian analysis that [Meyer et al. \(2022\)](#) compared to) is just by using the 95% confidence interval endpoints for test sensitivity and specificity to correct for the true seroprevalence using $P = S \times Se + (1 - S) \times (1 - Sp)$.

5 Future Directions

References

- Antoine Lewin, Roseline Therrien, Gaston De Serres, Yves Grégoire, Josée Perreault, Mathieu Drouin, Marie-Josée Fournier, Tony Tremblay, Julie Beaudoin, Guillaume Beaudoin-Bussi res, et al. Sars-cov-2 seroprevalence among blood donors in qu bec, and analysis of symptoms associated with seropositivity: a nested case-control study. *Canadian Journal of Public Health*, 112(4):576–586, 2021.
- Antoine Lewin, Gaston De Serres, Yves Gr goire, Jos e Perreault, Mathieu Drouin, Marie-Jos e Fournier, Tony Tremblay, Julie Beaudoin, Am lie Boivin, Guillaume Goyette, et al. Seroprevalence of sars-cov-2 antibodies among blood donors in qu bec: an update from a serial cross-sectional study. *Canadian Journal of Public Health*, 113(3):385–393, 2022.
- Mark J Meyer, Shuting Yan, Samantha Schlageter, John D Kraemer, Eli S Rosenberg, and Michael A Stoto. Adjusting covid-19 seroprevalence survey results to account for test sensitivity and specificity. *American Journal of Epidemiology*, 191(4):681–688, 2022.
- Eli S Rosenberg, James M Tesoriero, Elizabeth M Rosenthal, Rakkoo Chung, Meredith A Barranco, Linda M Styer, Monica M Parker, Shu-Yin John Leung, Johanne E Morne, Danielle Greene, et al. Cumulative incidence and diagnosis of sars-cov-2 infection in new york. *Annals of epidemiology*, 48:23–29, 2020.

A Supplementary Material