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

Naitong Chen

March 11, 2023

1 Introduction

Over the past few years, tracking the spread of COVID-19 has been crucial to developing a scientific understanding the disease, which ultimately guided public health protocols aimed at controlling the spread of the disease across the world. As with any epidemic/pandemic, reported case counts wthin 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 greater than that reflected from reported case counts (cite).

To more accuratly 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 case of infection of the tested disease. Therefore, the proportion of positive tests in the sample can be used as an estimate of the porportion of population infected with the disease over some time interval, which we call the cumulative incidence. Given the population size of the corresponding region, one can estimate the total number of infected individuals in the population using the estimated cumulative incidence. 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, prior to sample collection, 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, using data from a seroprevalence study conducted within the first few months of the COVID-19 pandemic, we can estimate cumulative incidence over the period from the beginning of the pandemic until roughly a month prior to when the samples were taken.

1.1 Adjusting for test-kit performance

Since antibody tests are not 100% accurate, there may be positive cases that test negative and negative cases that test positive. Therefore, one would ideally also like to adjust cumulative incidence for test-kit performance. This is typically done as follows. We begin by defining test specificity sp as the proportion of noncases that test negative and test sensitivity se as the proportion of actual cases that test positive. Then with the true cumulative incidence being denoted as s, we can write the observed prevalence p, which describes the proportion of population that would test positive for antibodies to the virus that causes the disease of interest, as

$$p = s \times se + (1 - s) \times (1 - sp).$$

To put in words, the observed prevalence can be decomposed into the proportion of actual cases that correctly test positive and noncases that incorrectly test positive. There are multiple approaches for incorporating this test-kit performance adjustment into the analysis of seroprevalence data. (cite) proposes a novel Bayesian model

that adjusts cumulative incidence for test-kit performance. This method is then applied to a dataset from a COVID-19 seroprevalence study conducted in New York state in early 2020. We walk through the construction of this method in the following section.

2 A Bayesian Approach to Analysis of Seroprevalence Data

Given the total sample size n from some region and the number of positive tests x from the sample, we can follow the above test-kit performance adjustment and model the number of positive tests as the outcome of a Binomial distribution with the total sample size as the number of trials and observed prevalence as the probability of success, i.e.,

$$x \mid n, s, se, sp \sim \mathsf{Binom}\left(n, s \times se + (1-s) \times (1-sp)\right) = \mathsf{Binom}(n, p).$$

Using the above as the likelihood function of s, se and sp, we can construct a Bayesian model by defining a set of prior distributions on each of s, se, and sp using distributions whose support is on $[0, 1]$. These prior distributions represent our apriori knowledge about these quantities. The prior and the likelihood together lead to a posterior distribution (conditional distribution given observed data) of s, which we can use to describe our belief on the cumulative incidence updated by observing the data at hand.

2.1 Deploying the Bayesian model

(cite) applies this bayesian model to a dataset obtained from a seroprevalence study conducted in New York state between April 19 and April 28 in 2020. This dataset contains the number of positive antibody tests and the total number of tests from each of the 11 regions across New York state in the study. Full details of data can be found in (citation). With consideration of the average time between infection and when antibodies become detectable, this dataset can be used to estimate cumulative incidences from the beginning of the pandemic until Mar 29, 2020. This is because there are 25 days between Mar 29, 2020 and the seroprevalence study midpoint April 23, 2020.

Instead of directly applying the above Bayesian model where each region gets its own prior on cumulative incidence, (cite) remarks it is possible that regions close to each other geographically may share sociodemographic factors which are associated with the number of infections. As a result, (cite) groups the 11 regions into three super-regions (New York City, Westchester and Rockland Counties and Long Island, as well as rest of state), with regions from the same super-region sharing a common prior distribution on their cumulative incidences. Denoting s_{ij}, p_{ij}, n_{ij} and x_{ij} as the cumulative incidence, observed incidence, number of samples, and number of positive antibody tests from the i^{th} region in the j^{th} super-region, the final Bayesian model is defined as follows.

$$\begin{aligned}
s_{i1} &\overset{\text{i.i.d.}}{\sim} \mathsf{Beta}(2.1792, 9.8208) \quad \forall i \text{ in super-region1 (New York City)},\\
s_{i2} &\overset{\text{i.i.d.}}{\sim} \mathsf{Beta}(2.6641, 9.3359) \quad \forall i \text{ in super-region2 (Westchester, Rockland Counties and Long Island)},\\
s_{i3} &\overset{\text{i.i.d.}}{\sim} \mathsf{Beta}(1.1930, 10.8070) \quad \forall i \text{ in super-region3 (rest of state)},\\
se &\sim \mathsf{Beta}(205, 29)_{\{0.8, 0.95\}},\\
sp &\sim \mathsf{Beta}(288, 2)_{\{0.9, 1\}},\\
p_{ij} &= s_{ij} \times se + (1 - s_{ij}) \times (1 - sp),\\
x_{ij} \mid n_{ij}, p_{ij} &\overset{\text{indep}}{\sim} \mathsf{Binom}(n_{ij}, p_{ij}).
\end{aligned}$$

The priors for each region are chosen so that the mean of the prior matches the ratio between the cumulative reported case count up until March 29, 2020 and the total population of the corresponding super-region. On the other hand, the priors on test sensitivity and test specificity are based on validation studies: (cite) estimates the test specificity to be 0.9975 with a 95% confidence interval of $[0.961, 1]$, and the test sensitivity to be 0.879 with a 95% confidence interval of $[0.837, 0.921]$. The priors are then chosen so that means and variances of the priors on test specificity and sensitivity to match the results from the validation studies. Note that the subscripts denote truncation to the specified regions.

Suppose that there are r_j regions in the j^{th} super-region, we can write density of the target posterior distribution as

$$p(S, se, sp \mid X, N) = \frac{1}{Z} p(se)p(sp) \prod_{j=1}^{3} \prod_{i=1}^{r_j} p(s_{ij}) p(x_{ij} \mid sp, se, s_{ij}),$$

where $p(\cdot)$ is the prior density corresponding to each parameter of interest, $p(\cdot \mid \cdot)$ is the likelihood, and Z is the normalization constant. Also note that S, X and N denote vectors containing the cumulative incidence, number of positive tests in the sample, and total sample size for each region considered in the study. With the model specified, we can use Markov Chain Monte Carlo to obtain samples from the posterior distribution of regional cumulative incidences as well as test specificity and sensitivity. Given a set of regional cumulative incidences from the posterior distribution, we can estimate the cumulative incidence for each super-region or the entire state using the average over the corresponding regional cumulative incidences weighted by the proportion of population living in each region. (cite) uses the median values over $100,000$ posterior samples as point estimates for each parameter of interest. At the same time, equal tailed 95% credible intervals, which cover 95% of the area under the corresponding posterior densities, are used to quantify the uncertainty about the estimates.

2.2 Related frequentist approaches

(cite) compares the results from the above Bayesian model to a non-Bayesian version of the same analysis. In the non-Bayesian version of the analysis, given a specific region (or super-region, or the entire state), with the sample proportion of positive tests being denoted $\hat{p}$, we can estimate the cumulative incidence of that region by rearranging the equation that adjusts for test specificity and sensitivity. Note that here the sample proportion of positive tests plays the same role as observed prevalence in the Bayesian version of the analysis. This leads to

$$s = (\hat{p} + sp - 1)/(se + sp - 1).$$

The point estimate of the cumulative incidence is obtained by plugging in the estimated test specificity and sensitivity values from the validation studies. To quantify the uncertainty around each point estimate in terms of test-kit performance, (cite) constructs an iterval using the 95% confidence interval endpoints from the validation studies. Let $\hat{se}, \hat{sp}$ denote the point estimates for test sensitivity and specificity from the validation studies, with the corresponding 95% confidence intervals being $[se_l, se_u]$ and $[sp_l, sp_u]$. The uncertainty interval associated with s can be written as

$$[(\hat{p} + sp_l - 1)/(se_u + sp_l - 1), \quad (\hat{p} + sp_u - 1)/(se_l + sp_u - 1)].$$

In particular, the lowerbound of this above interval is obtained by plugging in the lower endpoint of test specificity and upper endpoint of test sensitivity, and vice versa. It is worth noting, however, that this interval is not a 95% confidence interval around the true cumulative incidence of the corresponding region.

Alternatively, (cite) takes on a different approach quantifying the uncertainty around each estimated cumulative incidence. The procedure can be summarized as follows. Given the number of positive tests as well as total sample size in a region, we can construct a 95% confidence interval on the observed prevalence for that given region. We denote this interval $[p_l, p_u]$. To incorporate the uncertainty of test-kit performance, we can construct the following three sets of 95% confidence intervals on the cumulative incidence

$$\begin{aligned}
&[(p_l + \hat{sp} - 1)/(\hat{se} + \hat{sp} - 1), \quad (p_u + \hat{sp} - 1)/(\hat{se} + \hat{sp} - 1)], \\
&[(p_l + sp_l - 1)/(se_u + sp_l - 1), \quad (p_u + sp_l - 1)/(se_u + sp_l - 1)], \\
&[(p_l + sp_u - 1)/(se_l + sp_u - 1), \quad (p_u + sp_u - 1)/(se_l + sp_u - 1)].
\end{aligned}$$

In the above three intervals, the first one corresponds to the average test-kit performance obtained from the validation studies, the second interval corresponds to the worst case of combined test-kit performance, and the third interval corresponds to the best case of combined test-kit performance.

Comparing the results from both studies (cite), we can see that the point estimates from both the Bayesian and non-Bayesian analyses are relatively similar, but the Bayesian credible intervals are generally narrower

than the intervals constructed using confidence interval endpoints. In addition, while some of the intervals constructed using confidence interval endpoints contain a negative lowerbound, this does not happen to any of the credible intervals constructed. We elaborate on advantages and disadvantages of the two analyses in the following sections.

3 Bayesian Seroprevalence Analysis compared to its frequentist counterpart

3.1 Significance

One of the most prominent advantages of the fully Bayesian approach to adjusting cumulative incidence for test-kit performance is that the quantified uncertainty is more interpretable compared to its non-Bayesian counterpart. This is precisely the motivation behind the development of this Bayesian procedure by (cite). In the Bayesian version of the analysis, (cite) expresses the uncertainty of the cumulative incidence for a given region using a 95% equal-tailed credible interval of the corresponding marginal posterior distribution. As an example, the credible interval of the cumulative incidence in the i^{th} region in the j^{th} super-region s_{ij} can be constructed as follows. Following the notation from the previous section, we begin by writing out the marginal posterior density for s_{ij}:

$$
\begin{aligned}
&p(s_{ij} \mid X, N) \\
&\propto \frac{1}{Z'} \int_{se} \int_{sp} p(se)p(sp) \left(\prod_{(r,q)\neq(i,j)} \int_{s_{rq}} p(s_{rq})p(x_{rq} \mid n_{rq}, sp, se, s_{rq}) \partial s_{rq} \right) p(s_{ij})p(x_{ij} \mid n_{ij}, sp, se, s_{ij}) \partial se \partial sp.
\end{aligned}
$$

Note that here the normalization constant is absored by proportionality. Given $p(s_{ij} \mid X, N)$, the 95% equal-tailed credible interval for s_{ij} corresponds to the range of possible s_{ij} values that covers the middle 95% of the area under $p(s_{ij} \mid X, N)$. In other words, if we were to sample from this marginal posterior distribution, there is a 95% chance that the sampled value is within the credible interval. Furthermore, since the prior distribution on s_{ij} ($p(s_{ij})$ in the expression of $p(s_{ij} \mid X, N)$) is Beta and has support $[0, 1]$, it is guaranteed by construction that the credible interval is contained in $[0, 1]$.

On the other hand, by following the non-Bayesian approach in (cite), we may obtain negative estimates or uncertainty intervals that cross zero. This can be seen from the cumulative incidence correction equation from the previous section. When the sample proportion of positive tests $\hat{p}$ is smaller than $1 - sp$, the resulting estimate of the cumulative incidence would be negative. At the same time, note that the interval reprensenting the uncertainty around the estimated cumulative incidence is obtained by plugging in endpoints of two independent confidence intervals (one on test specificity and one on test sensitivity). We know that each 95% confidence interval is a realization of all possible intervals that overall have a 95% chance of covering what is regarded as the underlying true value. Then the proportion of pairs of 95% confidence intervals on test specificity and test sensitivity covering the true values simultaneously is less than 95%. As a result, by constructing an interval through aggregating two independent 95% confidence intervals, the resulting interval would not be a valid 95% confidence interval. This complicates the interpretation of the resulting interval. Furthermore, uncertainty intervals constructed this way do not take into account the uncertainties around the sample proportion of positive tests. While the other approach in (cite) takes into account both the uncertainties of the sample proportion of positive tests as well as test-kit performance, the use of three intervals still makes the quantified uncertainty not as easily interpreted as that using the Bayesian credible intervals. Finally, we note that this approach may still result in negative estimated cumulative incidences.

3.2 Limitations and challenges

We begin by clarifying the notion of hierarchical priors. (cite) claims that the model they have constructed entails a hierarchical prior structure. However, this is not the case. The simplest Bayesian hierarchical models contain three layers: likelihood of the parameter of interest given data, a prior distribution that governs the parameter of interest, and a hyper-prior distribution that governs the prior distribution. However, the Bayesian model defined in (cite) does not contain any hyper-prior distributions that govens the prior distribution on

cumulative incidence, test specificity, or test sensitivity. Therefore, while the Bayesian model constructed in (cite) does leverage prior information unique to each super-region, the use of the term hierarchical prior here is not precise.

Following up on use of prior distributions that are unique to each super-region, the grouping of regions into super-regions remain somewhat arbitrary. For example, (cite) does not provide an argument for combining Westchester and Rockland Counties with Long Island into one super-region. It would be of interest to know how sensitive the Bayesian model is to the grouping of regions into super-regions. After the structure of the model is determined, we still need to consider the sensitivity of the model to prior specification. While (cite) tested a set of uninformative, weakly informative, and informative priors on the regional cumulative incidence, the same sensitivity analysis should ideally be carried out for test sensitivity and test specificity as well. This is because there are vast variabilities among estimated COVID-19 antibody test sensitivity and specificity from different studies (cite). Should there be a clear difference in the resulting estimates, further investigations should be carried out in order to produce reliable estimates of cumulative incidences for each region.

Finally, we note that both the Bayesian and non-Bayesian analyses considered here do not account for seroreversion, which is the loss of antibody detectability. As we obtain seroprevalence data from later stages of the pandemic, we run into higher risks of underestimating the proportion of population previously infected by COVID-19. This is because antibodies for the virus that causes COVID-19 may become undetectable after some period of time since infection. Individuals who have seroreverted would not count towards the sample proportions of positive antibody tests. However, since we know that individuals who have seroreverted may still be immune to the virus that causes COVID-19 (cite), it is important to account for these individuals in many cases. An example is when the goal is to understand how close a population is to achieving herd immunity. As a result, both the Bayesian and non-Bayesian analyses discussed above do not generalize to seroprevalence studies conducted later on in the pandemic as one might have hoped. In the following section, we propose a modification to the framework discussed above to account for seroreversion, and apply this modified model to a serial seroprevalence study conducted in Quebec, Canada between 2020 and 2021.

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 seroprevalances. 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

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

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

Note that for observed seroprevalences in the first study, we adjust for test sensivitity 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 sensivitity 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 substract 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-Bussières, et al. Sars-cov-2 seroprevalence among blood donors in québec, 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 Grégoire, Josée Perreault, Mathieu Drouin, Marie-Josée Fournier, Tony Tremblay, Julie Beaudoin, Amélie Boivin, Guillaume Goyette, et al. Seroprevalence of sars-cov-2 antibodies among blood donors in québec: 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