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

Naitong Chen

March 13, 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 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 greater than that reflected from reported case counts (Byambasuren et al., 2021).

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 case of previous infection of the tested disease. Therefore, the proportion of positive tests in the sample can be used as an estimate of the proportion 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 start producing 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. In fact, it is estimated to be around 2.5% in the US by Khafaie and Rahim (2020). 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 (Choe et al., 2021; Sethuraman et al., 2020). 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. Meyer et al. (2022) 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 (based on detection of SARS-Cov-2 IgG) conducted in New York state in early 2020. We describe the construction of this method in the following section.

## 2 A Bayesian Approach to Analyzing 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 a priori 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

Meyer et al. (2022) 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 conducted in each of the 11 regions across New York state in the study. Full details of data can be found in Rosenberg et al. (2020). 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 the same prior on cumulative incidence, Meyer et al. (2022) remarks it is possible that regions close to each other geographically may share sociodemographic factors that are associated with the number of infections. As a result, Meyer et al. (2022) 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 $i^{\text{th}}$ region in the $j^{\text{th}}$ super-region, the final Bayesian model is defined as follows.

$$s_{i1} \overset{\text{i.i.d.}}{\sim} \mathsf{Beta}(2.1792, 9.8208) \quad \forall i \text{ in super-region 1 (New York City)},$$
$$s_{i2} \overset{\text{i.i.d.}}{\sim} \mathsf{Beta}(2.6641, 9.3359) \quad \forall i \text{ in super-region 2 (Westchester, Rockland Counties and Long Island)},$$
$$s_{i3} \overset{\text{i.i.d.}}{\sim} \mathsf{Beta}(1.1930, 10.8070) \quad \forall i \text{ in super-region 3 (rest of state)},$$
$$se \sim \mathsf{Beta}(205, 29)_{\{0.8, 0.95\}},$$
$$sp \sim \mathsf{Beta}(288, 2)_{\{0.9, 1\}},$$
$$p_{ij} = s_{ij} \times se + (1 - s_{ij}) \times (1 - sp), \quad \forall i, j$$
$$x_{ij} \mid n_{ij}, p_{ij} \overset{\text{indep}}{\sim} \mathsf{Binom}(n_{ij}, p_{ij}) \quad \forall i, j.$$

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: Rosenberg et al. (2020) 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 match the results from the validation studies. Note that the subscripts denote truncation to the specified regions to correspond to the widths of the confidence

intervals.

Suppose that there are $r_j$ regions in the $j^{\text{th}}$ super-region, we can write the 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 (MCMC) 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 an average over all regional cumulative incidences under consideration weighted by the proportion of population living in each region. Meyer et al. (2022) uses the median values over $400,000$ 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

Meyer et al. (2022) 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), we can estimate the cumulative incidence of that region by rearranging the equation that adjusts for test specificity and sensitivity.

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

The point estimate of the cumulative incidence is obtained by plugging in the sample proportion of positive tests $\hat{p}$ as well as the estimated test specificity $\hat{sp}$ and sensitivity $\hat{se}$ from the validation studies. To quantify the uncertainty around each point estimate in terms of test-kit performance, Meyer et al. (2022) constructs an interval using the 95% confidence interval endpoints from the validation studies. Let 95% confidence intervals for test sensitivity and test specificity be denoted $[se_l, se_u]$ 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 lower bound 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. This is discussed in more detail in the following sections.

Alternatively, Rosenberg et al. (2020) 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 $x$ as well as total sample size $n$ in a region, we can construct a 95% confidence interval on the observed prevalence for that given region

$$[p_l, \quad p_u] = \left[ \hat{p} - Q(0.975) \times \sqrt{\frac{\hat{p}(1 - \hat{t})}{n}}, \quad \hat{p} + Q(0.975) \times \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right],$$

where $\hat{p} = \frac{x}{n}$ and Q is the quantile function of the standard normal distribution. To incorporate the uncertainty of test-kit performance, we can construct the following three sets of 95% confidence intervals on the cumulative incidence

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

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, 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 lower bound, this does not happen to any of the credible intervals constructed. A full comparison between the two models can be found in Meyer et al. (2022). In the following section, we discuss the advantages and disadvantages of the Bayesian seroprevalence analysis.

# 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 Meyer et al. (2022). In the Bayesian version of the analysis, Meyer et al. (2022) 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 $i^{\text{th}}$ region in the $j^{\text{th}}$ 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 $s_{ij}$:

$$p(s_{ij} \mid X, N)$$
$$\propto \int_{se} \int_{sp} p(se)p(sp) \left( \prod_{(r,q)\neq(i,j)} \int_{s_{rq}} p(s_{rq})p(x_{rq} \mid n_{rq}, sp, se, s_{rq})\partial s_{rq} \right) p(s_{ij})p(x_{ij} \mid n_{ij}, sp, se, s_{ij})\partial se\partial sp.$$

Note that here the normalization constant is absorbed 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 Meyer et al. (2022), we may obtain negative estimates or uncertainty intervals that cross zero. This can be seen from the cumulative incidence correction equation in Section 2.2. When the sample proportion of positive tests $\hat{p}$ is smaller than $1 - \hat{sp}$, the resulting estimate of the cumulative incidence would be negative. At the same time, note that the interval representing 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 Rosenberg et al. (2020) 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 under the Bayesian framework. 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. Meyer et al. (2022) 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 Meyer et al. (2022) does not contain any hyper-prior distributions that govern the prior distribution on cumulative incidence, test specificity, or test sensitivity. Therefore, while the Bayesian model constructed in Meyer et al. (2022) 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, Meyer et al. (2022) 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 Meyer et al. (2022) tested a set of non-informative, weakly informative, and informative priors on the regional cumulative incidences, 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. As an example, test sensitivity is estimated to be as low as 28% in Noordin et al. (2022). Should there be a clear difference in the resulting estimates, further investigations need 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, it is important to account for these individuals in many cases. An example is when the goal is to estimate the case fatality rate of COVID-19, where it is crucial to include those who have seroreverted to avoid overestimating the case fatality rate (Brazeau et al., 2022). 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 Bayesian model to a serial seroprevalence study conducted in Quebec, Canada between 2020 and 2021.

# 4   Bayesian Seroprevalence Analysis with Seroreversion Correction

In a serial seroprevalence study conducted in Quebec, Canada by Lewin et al. (2021) and Lewin et al. (2022), a procedure for seroreversion correction is applied in a non-Bayesian analysis. Here we begin by describing the dataset used in this analysis before illustrating the procedure used to adjust estimated cumulative incidences for seroreversion.

## 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. For both phases of the study, numbers of antibody-positive samples as well as total number of samples are recorded for each of the three regions in Quebec: Montreal-Laval region, region surrounding Montreal-Laval, as well as other regions. Phase I of the study was conducted relatively early in the pandemic, and it collected samples obtained between May 25 and July 9, 2020. Phase II of the study contains samples collected from January 25 to Mar 11, 2021.

While we can use the data from phase I of the study to estimate regional cumulative incidences in Quebec between the beginning of the pandemic until around May 23, 2020 (accounting for the average 25 days between infection and when antibodies become detectable), simply repeating the same analysis using data from

|                              | **Phase I** |        | **Phase II (unvaccinated)** |        | **Seroreversion substudy** |        |
|------------------------------|:-----------:|:------:|:---------------------------:|:------:|:--------------------------:|:------:|
|                              | reactive    | tested | reactive                    | tested | seroreverted               | tested |
| Montreal-Laval               | 90          | 3061   | 48                          | 1925   | -                          | -      |
| Surrounding Montreal-Laval   | 48          | 1925   | 128                         | 1422   | -                          | -      |
| Other                        | 35          | 2705   | 372                         | 4304   | -                          | -      |
| Total                        | 173         | 7691   | 715                         | 7304   | 32                         | 109    |

Table 1: Seroprevalence data from Lewin et al. (2021) and Lewin et al. (2022). Reactive indicates the number of positive tests against SARS-Cov-2 IgG.

phase II of the study likely would not result in valid estimates of cumulative incidences between the beginning of the pandemic and the beginning of 2021. This is because there had already been almost a year since the beginning of the pandemic by the time when phase II samples were collected. Some of the individuals may have had a previous infection but had seroreverted by the time of sample collection. Consequently, the raw estimate may be lower than the true cumulative incidence of a given region. As a result, to account for seroreversion when analyzing data from phase II of the seroprevalence study, Lewin et al. (2022) also conducted a seroreversion substudy. In particular, a number of the individuals who tested positive for SARS-Cov-2 antibodies during phase I of the study were tested again in 2021. The number of these individuals as well as the number of negative tests (indicating seroreversion) were recorded. The data from both phases of the study are summarized in Table 1.

It is important to point out that the phase II data presented here only concern the unvaccinated population. While data for both those who were and were not vaccinated are presented in Lewin et al. (2022), we have chosen to focus on the unvaccinated population since none of the vaccinated individuals had seroreverted at the time and so it is not necessary for adjust for seroreversion. Finally, we note that for the remainder of the report, any notation that is redefined from the previous sections are within the context of the seroprevalence study in Quebec.

## 4.2   Adjusting for seroreversion

As mentioned above, not correcting for seroreversion when conducting seroprevalence analysis on data from phase II of the above-mentioned study may result in unreliable estimates of regional cumulative incidences. This is because such analyses fail to account for those who have had a previous infection but had already seroreverted by the time of sample collection. Luckily, with the data from phase I of the study, for each region in Quebec, we can produce a valid estimate of the cumulative incidence between the beginning of the pandemic and May 2020. Additionally, from the seroreversion substudy, we can estimate the proportion of the antibody-positive population from phase I of the study that had seroreverted by phase II of the study. As a result, Lewin et al. (2022) adjusts the raw estimate of cumulative incidence from phase II by adding the product of the estimated cumulative incidence from phase I and the estimated proportion that have seroreverted. Mathematically, let $\hat{s}_{prev}$ be the estimated cumulative incidence of a region in Quebec from phase I, $\hat{p}$ be the estimated observed prevalence (proportion of positive samples) from phase II, and $\hat{sr}$ be the observed proportion of seroreversion, we can estimate the cumulative incidence between the beginning of the pandemic and the beginning of 2021 using

$$\hat{s} = \hat{p} + \hat{s}_{prev} \times \hat{sr}.$$

To quantify the uncertainty around $\hat{s}$, we begin by noting that there are three sources of uncertainty: the estimated cumulative incidence from phase I, the observed prevalence from phase II, and the proportion of seroreversion. Note that here the cumulative incidence from phase I is directly estimated using the observed prevalence from phase I. We can construct individual 95% confidence intervals for the cumulative incidence from phase I $s_{prev}$ and observed prevalence from phase II $p$, as well as the proportion of seroreversion $sr$. We write these intervals as

$$[s_{prev,l}, \quad s_{prev,u}], \quad [p_l, \quad p_u], \quad [sr_l, \quad sr_u].$$

Then to construct an uncertainty interval for $s$, we can follow Meyer et al. (2022) and aggregate the above three 95% confidence intervals using combinations of the confidence interval endpoints that give the lowest and highest

estimated cumulative incidence

$$[p_l + s_{prev,l} \times sr_l, \quad p_u + s_{prev,u} \times sr_u].$$

Again, the above interval is not a valid 95% confidence interval, but it nonetheless quantifies the scale of uncertainty.

We can easily extend the model above and incorporate adjustments for test-kit performance. This is achieved by writing, with $p_{prev}$ denoting the observed prevalence in phase I,

$$p_{prev} = s_{prev} \times se + (1 - s_{prev}) \times (1 - sp) \implies s_{prev} = \frac{p_{prev} + sp - 1}{se + sp - 1},$$

$$p = (s - sr \times s_{prev}) \times se + (1 - (s - sr \times s_{prev})) \times (1 - sp) \implies s = \frac{p + 1 - sp}{se + sp - 1} + sr \times s_{prev}.$$

In this above formulation of observed prevalence from phase II $(p)$, we first subtract from the true cumulative incidence 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 yet seroreverted. Using the same approach to aggregate over all sources of uncertainty, we arrive at the estimated cumulative incidence for a given region between the beginning of the pandemic and the beginning of 2021

$$\hat{s} = \frac{\hat{p} + 1 - \hat{sp}}{\hat{se} + \hat{sp} - 1} + \hat{sr} \times \hat{s}_{prev}$$

with a corresponding uncertainty interval adjusting for both seroreversion and test-kit performance

$$\left[ \frac{p_l + 1 - sp_l}{se_u + sp_l - 1} + sr_l \times s_{prev,l}, \quad \frac{p_u + 1 - sp_u}{se_l + sp_u - 1} + sr_u \times s_{prev,u} \right],$$

Note that

$$s_{prev,l} = (p_{prev,l} + sp_l - 1)/(se_u - 1 + sp_l), \quad s_{prev,u} = (p_{prev,u} + sp_u - 1)/(se_l - 1 + sp_u)$$

correspond to the 95% confidence interval endpoints for $s_{prev}$. Finally, we note that if we had data on the unvaccinated population from each of the three regions in Quebec, we could obtain a province-wide estimate of the cumulative incidence by averaging the regional cumulative incidences weighted by the proportion of population in each region. However, this information is not included in Lewin et al. (2022), and so our analysis here remains regional. We also do not adjust for age and sex of individuals due to data unavailability.

## 4.3   Model formulation

While the previous section outlines an approach to conducting seroprevalence analysis that accounts for both seroreversion and test-kit performance, we note that the resulting uncertainty intervals are not valid 95% confidence intervals, and that it is still possible to get negative estimates in certain cases. As demonstrated in Meyer et al. (2022), this can be addressed by constructing a Bayesian version of the same analysis. We outline this model below.

Recall that $sr$ denotes the test seroreversion proportion. Following the same notations as established in Section 2, with $i = 1, 2$ indicating the phase of the study, and $j = 1, 2, 3$ representing Montreal-Laval, surrounding

Montreal-Laval, and other regions, the final Bayesian model can be written as

$$s_{1j} \sim \mathsf{Beta}(1,1) \quad \forall j \in \{1,2,3\}$$
$$s_{2j} \sim \mathsf{Beta}(1,1)_{\{sr \times s_{1j},1\}} \quad \forall j \in \{1,2,3\}$$
$$se \sim \mathsf{Beta}(205,29)$$
$$sp \sim \mathsf{Beta}(288,2)$$
$$sr \sim \mathsf{Beta}(32,77)$$
$$p_{1j} = s_{1j} \times se + (1 - s_{1j}) \times (1 - sp) \quad \forall j \in \{1,2,3\}$$
$$p_{2j} = (s_{2j} - sr \times s_{1j}) \times se + (1 - (s_{2j} - sr \times s_{1j})) \times (1 - sp) \quad \forall j \in \{1,2,3\}$$
$$x_{ij} \mid n_{ij}, p_{ij} \overset{\mathrm{indep}}{\sim} \mathsf{Binom}(n_{ij}, p_{ij}), \quad \forall i \in \{1,2\}, \quad j \in \{1,2,3\}.$$

The density of the target posterior distribution is then given by

$$p(S, se, sp, sr \mid X, N) = \frac{1}{Z} p(se)p(sp)p(sr) \prod_{i=1}^{3} \prod_{j=1}^{2} p(s_{ij})p(x_{ij} \mid sp, se, sr, s_{ij}).$$

In this above model, $p_{1j}$ is modeled in the same way as Meyer et al. (2022) for all $j$, since there is no need to adjust for seroreversion. $p_{2j}$, on the other hand, accounts for both seroreversion and test-kit performance. Since the population data by the three regions in Quebec are not provided in Lewin et al. (2022), non-informative priors are used here. To ensure we do not run into negative seroprevalence estimates, we truncate $s_{21}, s_{22}, s_{23}$ accordingly. For the priors on test sensitivity and test specificity, we use the untruncated version of the same priors as in Meyer et al. (2022). This allows for more room for uncertainty given that these tests were conducted at a different location and at different times. Finally, the prior on seroreversion is chosen so that the mean reflects the sample proportion of seroreversion.

Similar to Meyer et al. (2022), we use MCMC to obtain samples from the target posterior distribution. Specifically, samples from $s_{2j}$ for all $j$ allow us to estimate the regional cumulative incidences in Quebec between the beginning of the pandemic and the beginning of 2021, where the estimates are adjusted for both seroreversion and test-kit performance. Figs. 2 and 3 show that the MCMC algorithm has converged (this is also confirmed by checking the Gelman-Rubin diagnostic). We follow Meyer et al. (2022) and samples from four individual chains (100,000 samples each with 100,000 burn-in samples discarded) to obtain the median as a point estimate as well as the equal-tailed 95% credible interval for uncertainty quantification.

To compare the Bayesian model output, we also conduct a number of frequentist seroprevalence analyses on the same dataset. In particular, we include the raw estimates of regional cumulative incidence from phase II directly from the sample proportion of positive tests, an analysis that adjusts for seroreversion, as well as an analysis that adjusts for both seroreversion and test-kit performance. The last two analyses correspond to those described in Section 4.2. The code used to generate all results and figures can be found at `https://github.com/NaitongChen/QP-4`.

## 4.4  Discussion of analysis results

The estimated regional cumulative incidences as well as their corresponding uncertainty intervals from each of the four analyses are summarized in Table 2. We see that from each of the four analyses, the estimated cumulative incidence in the Montreal-Laval region is the highest, and those of the other two regions are similar to each other, with the region surrounding Montreal-Laval having a slightly higher estimated cumulative incidence. We now look at the estimate for the same region across the three non-Bayesian analyses. Going from the raw results to those adjusted for seroreversion, we see that there is an increase in the estimated cumulative incidence. By further adjusting for test-kit performance, we see yet another increase in the estimated cumulative incidence for the same region. This two-step increase is consistently observed for each of the three regions in Quebec. This pattern is sensible as with each additional adjustment, we are inherently adding to the cumulative incidence estimate by including parts of the population that previously would not have been considered. However, as we increase the number of items that we adjust for (seroreversion and test-kit performance), we also incur more sources of uncertainty. Given that we have constructed our uncertainty intervals using the combination of individual 95% confidence interval endpoints that yield the lowest and highest estimates, the resulting intervals

| | Raw results | | |
|---|---|---|---|
| | mean | lower bound | upper bound |
| Montreal-Laval | 0.136 | 0.120 | 0.154 |
| Surrounding Montreal-Laval | 0.090 | 0.076 | 0.106 |
| Other | 0.086 | 0.078 | 0.095 |
| | **Adjusted for seroreversion** | | |
| | mean | lower bound | upper bound |
| Montreal-Laval | 0.145 | 0.125 | 0.168 |
| Surrounding Montreal-Laval | 0.097 | 0.080 | 0.119 |
| Other | 0.090 | 0.080 | 0.102 |
| | **Adjusted for seroreversion and test-kit performance** | | |
| | mean | lower bound | upper bound |
| Montreal-Laval | 0.161 | 0.088 | 0.201 |
| Surrounding Montreal-Laval | 0.107 | 0.037 | 0.142 |
| Other | 0.100 | 0.037 | 0.122 |
| | **Bayesian analysis adjusted for seroreversion and test-kit performance** | | |
| | median | lower bound | upper bound |
| Montreal-Laval | 0.159 | 0.137 | 0.183 |
| Surrounding Montreal-Laval | 0.105 | 0.085 | 0.125 |
| Other | 0.096 | 0.082 | 0.110 |

Table 2: Summary of regional cumulative incidence point estimates in Quebec along with uncertainty intervals for all methods discussed. The first three non-Bayesian analyses contain the mean estimate as well as an uncertainty interval constructed from multiple 95% confidence intervals, and the last Bayesian analysis contains the median and equal-tailed 95% credible interval of the corresponding posterior distribution.

grow wider and wider. Take the Montreal-Laval region as an example, the width of the uncertainty interval increases from 0.34 in the raw results to 0.113 in the analysis that accounts for both seroreversion and test-kit performance. The width of the interval more than triples. It is worth mentioning that the uncertainty intervals for $s_{prev}$ for each region (or equivalently $s_{11}, s_{12}, s_{13}$) cross zero. In other words, they contain a negative lower bound.

We now turn our attention to the output of the Bayesian model. Since the priors for each parameter of interest are properly set up, none of the estimates are negative (Figs. 2 and 3). Compared to the non-Bayesian analysis that also adjusts for both seroreversion and test-kit performance, we see that the point estimates for each region are very similar to each other, with the Bayesian point estimates being slightly lower. At the same time, the Bayesian 95% credible intervals are consistently narrower than the corresponding non-Bayesian uncertainty intervals. Again taking the Montreal-Laval region as an example, the Bayesian credible interval has a width of 0.046, which is less than half of the width of the non-Bayesian interval (0.113). Fig. 1 visualizes the uncertainty intervals between the Bayesian and non-Bayesian analyses that adjust for both seroreversion and test-kit performance. We see that the non-Bayesian uncertainty intervals are not centred around the point estimate. This is likely due to the non-linearity introduced by the adjustments on estimated cumulative incidences. To summarize, the Bayesian seroprevalence analysis, as noted by Meyer et al. (2022), produces similar point estimates compared to its non-Bayesian counterpart, but with narrower and more interpretable uncertainty intervals.
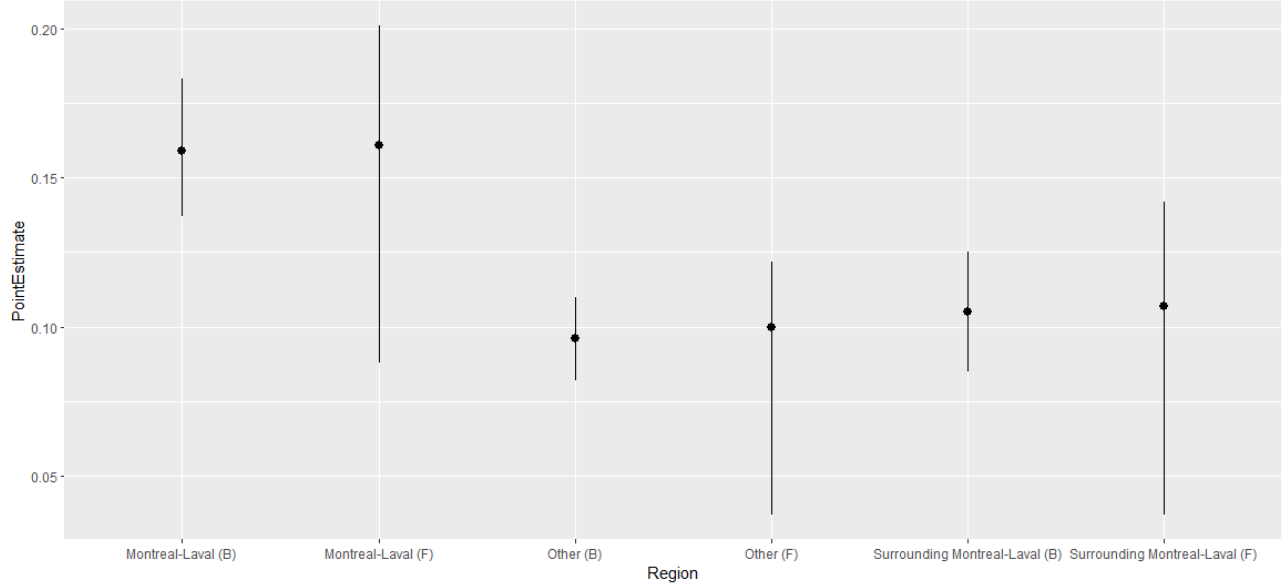
Figure 1: Comparison between uncertainty intervals for regional cumulative incidences between the Bayesian (B) and non-Bayesian (F) seroprevalence analyses adjusted for both seroreversion and test-kit performance.

# 5   Conclusion and Future Directions

Meyer et al. (2022) proposes a Bayesian model for conducting seroprevalence analysis that accounts for test-kit performance. By applying this model to a dataset from a seroprevalence study conducted in New York state in early 2020 and comparing the results to the non-Bayesian version of the same analysis, we see that both analyses provide similar point estimates of regional cumulative incidences. However, the Bayesian credible intervals used to quantify uncertainties around each estimate, compared to their non-Bayesian counterparts, are narrower, more interpretable, and stay well-defined (non-negative).

In this report, we extend the above Bayesian model by incorporating the approach in Lewin et al. (2022) to additionally account for seroreversion. By comparing the newly proposed model to its non-Bayesian counterpart using seroprevalence data from Quebec, Canada, we see that the claims in Meyer et al. (2022) largely hold. Namely, the point estimates from both models are similar, with the Bayesian credible intervals being narrower and more centred around the point estimates. In fact, we do come across instances of the non-Bayesian uncertainty intervals having a lowerbound below zero.

One of the major motivations for approaching seroprevalence analysis in a Bayesian way is to improve the interpretability of the uncertainty intervals associated with each estimated cumulative incidence. In particular, with the way that uncertainty intervals are constructed in the non-Bayesian analysis in Meyer et al. (2022), they are not valid confidence intervals. Therefore, as a future direction of research, it is of interest to develop more theoretically justified ways to aggregate individual confidence intervals to produce uncertainty intervals that remain as valid confidence intervals. One idea is to use the Bonferroni correction to adjust the confidence level of each individual confidence interval so that the resulting interval has a desired confidence level. The degree Bonferroni correction here depends on the number of individual confidence intervals (or sources of uncertainty) that we use to produce the final interval. This is based on the observation that all cumulative incidence estimators discussed in this report are strictly monotone in each parameter of interest. However, this approach does not rule out the possibility of obtaining negative point estimates or negative uncertainty interval endpoints.

Another direction for future research involves more formally testing the sensitivity of the proposed Bayesian models to the specification of prior distributions. As discussed before, the priors on cumulative incidences are based on cumulative reported case counts, and the priors on test sensitivity, test specificity, as well as seroreversion proportion are based on (external) validation studies. These sources of prior distributions contain

variabilities within themselves, and so it is crucial to assess prior sensitivity beyond merely testing a number of plausible prior specifications. Should the model be sensitive to prior specification, more investigation may be conducted to help stabilize the output of the corresponding Bayesian models.

# References

Nicholas F Brazeau, Robert Verity, Sara Jenks, Han Fu, Charles Whittaker, Peter Winskill, Ilaria Dorigatti, Patrick GT Walker, Steven Riley, Ricardo P Schnekenberg, et al. Estimating the covid-19 infection fatality ratio accounting for seroreversion using statistical modelling. *Communications medicine*, 2(1):54, 2022.

Oyungerel Byambasuren, Claudia C Dobler, Katy Bell, Diana Patricia Rojas, Justin Clark, Mary-Louise McLaws, and Paul Glasziou. Comparison of seroprevalence of sars-cov-2 infections with cumulative and imputed covid-19 cases: Systematic review. *PloS one*, 16(4):e0248946, 2021.

Pyoeng Gyun Choe, Kye-Hyung Kim, Chang Kyung Kang, Hyeon Jeong Suh, EunKyo Kang, Sun Young Lee, Nam Joong Kim, Jongyoun Yi, Wan Beom Park, and Myoung-don Oh. Antibody responses 8 months after asymptomatic or mild sars-cov-2 infection. *Emerging infectious diseases*, 27(3):928, 2021.

Morteza Abdullatif Khafaie and Fakher Rahim. Cross-country comparison of case fatality rates of covid-19/sars-cov-2. *Osong Public Health and Research Perspectives*, 11(2):74, 2020.

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.

Noorliza Mohamad Noordin, Aziyati binti Omar, Ishmah Hana Isharudin, Riasah Idris, Intan Surianne Mat Sahat, Selvanesan Sengol, Zirwatul Adilah Aziz, Zhuo-zhi Lim, and Teck-Onn Lim. Sero-prevalence of covid-19 among workers in malaysia. *medRxiv*, pages 2022–01, 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.

Nandini Sethuraman, Sundararaj Stanleyraj Jeremiah, and Akihide Ryo. Interpreting diagnostic tests for sars-cov-2. *Jama*, 323(22):2249–2251, 2020.
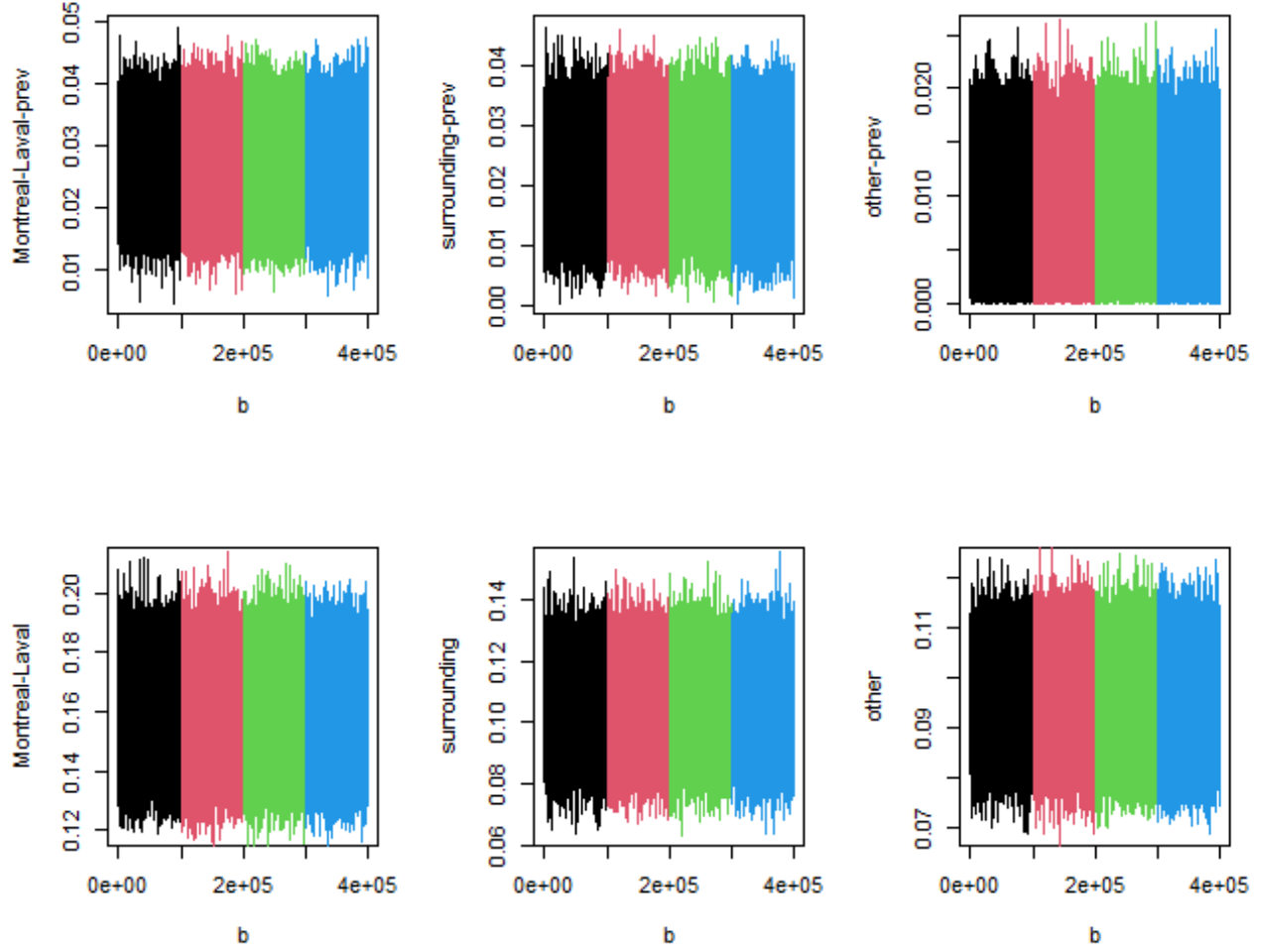
# A   Supplementary Material



Figure 2: Trace plots for regional cumulative incidences. The first row correponds to cumulative incidences after Phase I of the study, and the second row correponds to cumulative incidences after Phase II of the study. The colors indicate traces from four independent MCMC chains.
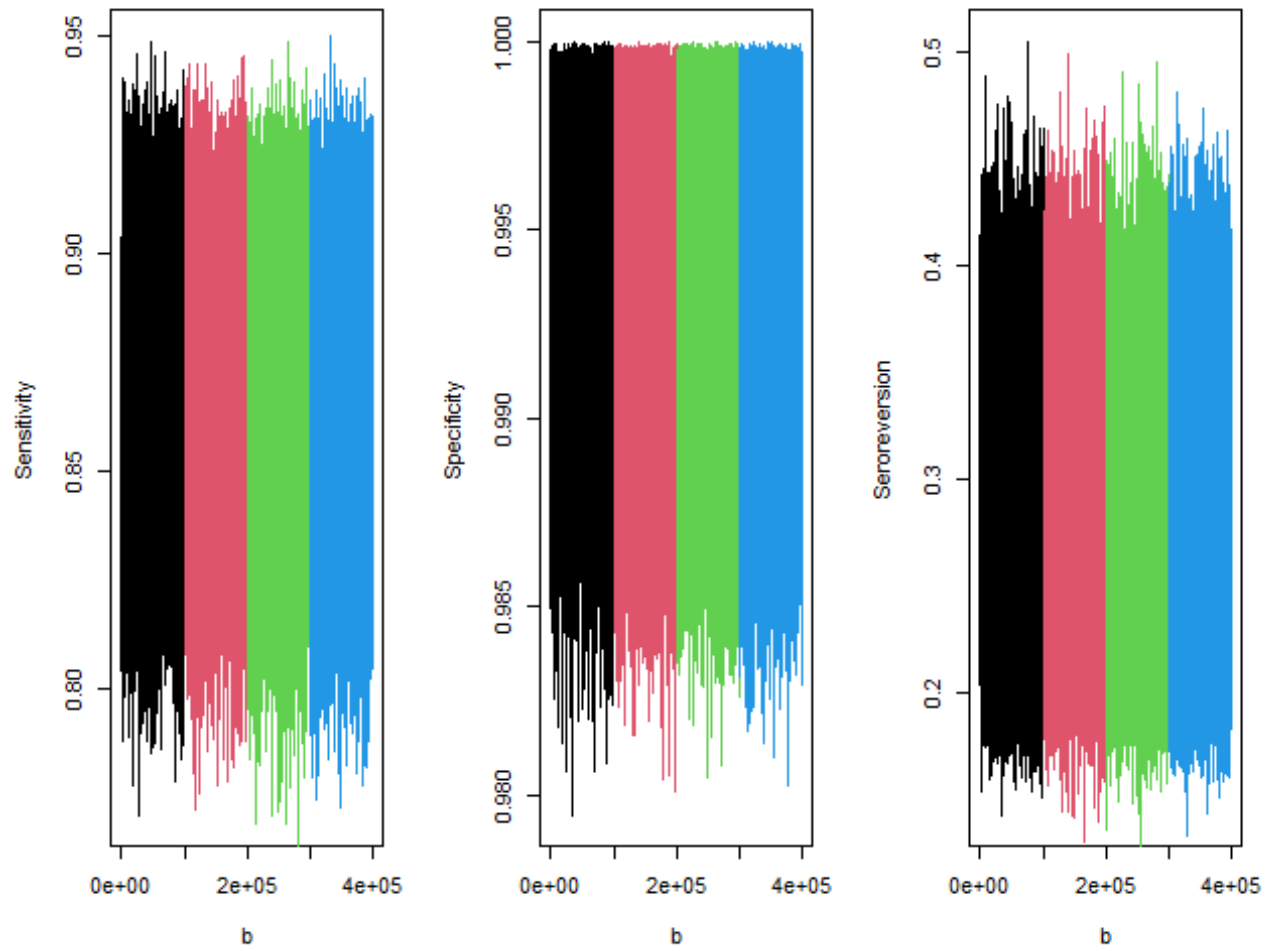
Figure 3: Trace plots for test sensitivity ($se$), test specificity ($sp$), and seroreversion rate ($sr$). The colors indicate traces from four independent MCMC chains.