Estimating Unobserved COVID-19 Infections in the United States

Quinn White

Submitted to the Department of Statistical and Data Sciences
of Smith College
in partial fulfillment
of the requirements for the degree of
Bachelor of Arts

Ben Baumer, Primary Faculty Advisor
Nicholas Reich, Secondary Faculty Advisor

May 2023

Acknowledgements

Will add

Preface

I am unsure as to what goes here.

Table of Contents

List of Tables

List of Figures

Abstract

As we have navigated the COVID-19 pandemic, case counts have been a central source of information for understanding transmission dynamics and the effect of public health interventions. However, because the number of cases we observe is limited by the testing effort in a given location, the case counts presented on local or national dashboards are only a fraction of the true infections. Variations in testing rate by time and location impacts the number of cases that go unobserved, which can cloud our understanding of the true COVID-19 incidence at a given time point and can create biases in downstream analyses. Additionally, the number of cases we observe is impacted by the sensitivity and specificity of the diagnostic test. To quantify the number of true infections given incomplete testing and diagnostic test inaccuracy, this work implements probabilistic bias analysis at a biweekly time scale from January 1, 2021 through February 2022. In doing so, we can estimate a range of possible true infections for a given time interval and location. This approach can be applied at the state level across the United States, as well as in some counties where the needed data are available.

Dedication

You can have a dedication here if you wish.

Chapter 1

Motivation

Placeholder

Chapter 2

Background

2.1 Probabalistic Bias Analysis

Often the focus of quantifying error about an effect estimate focuses on random error rather than the systematic error. For example, typical frequentist confidence intervals are frequent in medical and epidemiological literature, although they have faced rising criticism (Greenland et al., 2016). These confidence intervals quantify the fraction of the times we expect the true value to fall in this interval under the assumption that our model is correct. That is, if we ran an experiment 100 times and computed the effect size each time, we would expect the 95% confidence interval to contain the true value to 95 of those times, on average. Neyman stressed this in his original publication formalizing the concept of a confidence interval in 1937 (Neyman, 1937). The nuance that the confidence interval is not the probability that the true value falls within this interval, however, is often lost in the discussion of results, in part because the true meaning of a confidence interval is less intuitive.

The aim of quantitative bias analysis is to estimate systematic error to give a range of possible values for the true quantity of interest. In this sense, it is a type of sensitivity analysis. It can be used to estimate various kinds of biases, from misclassification, as is implemented in this work, as well as selection bias and unmeasured confounding (Petersen, Ranker, Barnard-Mayers, MacLehose, & Fox, 2021). Often, the goal of performing such an analysis is to see how these sources of bias affect our estimates; in particular, under what situations of bias the observed effect would be null.

There are multiple different forms of bias analysis (Lash, Fox, & Fink, 2009). The most simple case, simple bias analysis, is correcting a point estimate for a single source of error. Multidimensional bias analysis extends this to consider sets of bias parameters, but still provides a corrected point estimate rather than a range of plausible estimates. Probabilistic bias analysis, meanwhile, defines probability distributions for bias parameters to generate a distribution of corrected estimates by repeatedly correcting estimates for bias under different combinations of the parameter values. Then, via Monte Carlo we obtain a distribution of corrected estimates that reflect the corrected values under different scenarios of bias, that is, under dif-

ferent combinations of the bias parameters. This can give us a better idea for the extent of uncertainty about the corrected estimates, although this uncertainty does depend on the specification of the bias parameter distributions. Inherent in bias analysis is the dependence of our results on the specification of bias parameters, which reflect what is known from available data, literature, or theory on the extent of bias that may occur. There is uncertainty about how we define these distributions or values; otherwise, if the precise values of the bias parameters were known, we could simply correct the estimates and probabilistic bias analysis would not be useful.

Although some forms of probabilistic bias analysis can be applied to summarized data, for example, frequencies in a contingency table, the methods are most often implemented with unsummarized data in its original form, as implemented here.

In choosing specific distributions for the bias parameters, different specifications may yield density functions where most of the density is within a similar interval (MAKE PLOT WITH EXAMPLE), which means the choice of the specific distribution will not be sensitive to the particular choice of density.

2.2 Bayesian Melding

2.2.1 Theoretical Background for the Approach

The Bayesian melding approach was proposed by Poole et al. (Poole & Raftery, 2000).

The Bayesian melding approach enables us to account for both uncertainty from inputs and outputs of a deterministic model. The initial motivation for the approach was to study the population dynamics of whales in the presence of substantial uncertainty around model inputs for population growth (Poole & Raftery, 2000). However, the framework provided by Poole et al. can applied in any circumstance where we have uncertainty around some quantities θ and ϕ where there is a deterministic function $M : \theta \to \phi$. Due the utility of Bayesian melding in various contexts, since this deterministic model M could take on a wide range of forms, the approach has since been applied in various fields, including urban simulations (Ševčíková, Raftery, & Waddell, 2007), ecology (Robson, 2014), and infectious disease (Powers et al., 2011).

At this point, we can define how Bayesian melding works more formally.

Let $M : \theta \to \phi$ be the deterministic model defined by the function relating a vector of input parameters θ to an output vector ϕ, and suppose we have a prior on θ denoted $q_1(\theta)$ and a prior on ϕ denoted $q_2(\phi)$.

However, note that we actually have two distinct priors on ϕ. There is the prior formed by the distribution induced on ϕ by the prior for θ and the function M, where we denote this induced prior $q_1^*(\phi)$. If M^{-1} exists, we can write this induced prior $q_1^*(\phi) = q_1(M^{-1}(\phi))|J(\phi)|$. This result follows from the fact $M(\theta) = \phi$, so we apply a change of variables to obtain the distribution of ϕ from the distribution of

θ. This is a generalization to the multivariate case of the change of variables result often covered in probability courses in the univariate case. That is, if we have a continuous random variable X with probability density function f_X and $Y = g(X)$ for a differentiable monotonic function, then the probability density function of Y is $f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right|$. In practice, M^{-1} rarely exists exists since θ is often of higher dimensionality then ϕ, in which cases M is not invertible. This means we generally approximate $q_1^*(\phi)$ without acquiring its analytical form.

In addition to this induced prior, we have the prior $q_2(\phi)$, which does not involve M nor the inputs θ. Since these priors are based on different sources of information and may reflect different uncertainties, often it useful to use both sources of information to inform our estimates. To do so, we need to combine the distributions for $q_1^*(\phi)$ and $q_2(\phi)$ to create a pooled distribution.

Multiple pooling strategies exist for distinct distributions, but one requirement for a Bayesian analysis is that the distribution should be independent of the order in which the prior is updated and the combining of the prior distributions. That is, updating the prior distributions using Bayes' theorem and then combining distributions should yield the same result as combining distributions and then updating this combined distribution; pooling methods that have this property are deemed externally Bayesian. Logarithmic pooling has been shown to be externally Bayesian under some conditions, which are likely to hold in most settings. Furthermore, logarithmic pooling has actually been shown to be the only pooling method where this holds (Genest, McConway, & Schervish, 1986).

The logarithmically pooled prior for ϕ by pooling $q_1^*(\phi)$ and $q_2(\phi)$ is proportional to

$$q_1^*(\phi)^\alpha q_2(\phi)^{1-\alpha}$$

where $\alpha \in [0, 1]$ is a pooling weight. Commonly, a choice of $\alpha = 0.5$ is used to give the priors equal weight. In this case, logarithmic pooling may be referred to as geometric pooling since it is equivalent to taking a geometric mean.

If M is invertible, we can obtain the contrained distributions for the model inputs by simply inverting. However, this is rare, so we have to think about how to proceed in the noninvertible case.

To get intuition for a valid strategy, consider a mapping $M : \theta \to \phi$ for $\theta \in \mathbb{R}$ and $\phi \in \mathbb{R}$ defined as follows. Note the choice of q_1, q_2 does not matter here as long as they are valid densities.

θ	$q_1(\theta)$	ϕ	$q_2(\phi)$
1	0.3	1	0.4
2	0.2	2	0.6
3	0.5	2	0.6

We see that M is not invertible since $\theta = 1$ and $\theta = 2$ both map to $\phi = 2$, which implies the inverse M^{-1} would not be well defined.

We can compute $q_1^*(\phi)$ using our function M and taking $q_1^*(\phi) = q_1(M^{-1}(\phi))$; in the continuous case we need to multiply by $|J(\phi)|$, but not in the discrete case (Blitzstein & Hwang, 2019).

So we have $q_1^*(1) = q_1(1) = 0.3$ since $M(1) = 1$, and $q_1^*(2) = q_1(2) + q_1(3) = 0.2 + 0.5 = 0.7$ since $M(2) = 2$ and $M(3) = 2$.

Then, we can compute the logarithmically pooled pooled prior with $\alpha = 0.5$ by taking $q_1^*(\phi)^\alpha q_2(\phi)^{1-\alpha}$.

For $\phi = 1$, we have $q_1^*(\phi)^\alpha q_2(\phi)^{1-\alpha} = (0.3)^{0.5}(0.4)^{0.5} = 0.3464$ For $\phi = 2$, we have $q_1^*(\phi)^\alpha q_2(\phi)^{1-\alpha} = (0.6)^{0.5}(0.7)^{0.5} = 0.6481$

To make this a valid density, however, these probabilities must sum to 1, so we renormalize by dividing by (0.3464 + 0.6481). This gives us the pooled prior $q^{\sim[\phi]}(\phi)$ as

$0.3464/(0.3464+0.6481) = 0.3483$ for $\phi = 1$ and $0.6481/(0.3464+0.6481) = 0.6517$ for $\phi = 2$.

Summarizing these results, we have

ϕ	$q_2(\phi)$	$q_1^*(\phi)$	$q^{\sim[\phi]}(\phi)$
1	0.4	0.3	0.3483
2	0.6	0.7	0.6517

However, we want the pooled prior on the inputs θ, that is, $q^{\sim[\theta]}(\theta)$.

Poole et al. reasoned as follows. Since M uniquely maps $\theta = 1$ to $\phi = 1$, the probability that $\theta = 1$ should be equal to the probability $\phi = 1$. That is, we should have $q^{\sim[\theta]}(1) = q^{\sim[\phi]}(1)$.

However, the relationship for $\theta = 2$ or $\theta = 3$ to ϕ is not one to one, but since $M(2) = 2$ and $M(3) = 2$, the sum of the probabilities for $\theta = 1$ and $\theta = 2$ should be equal to that for $\phi = 2$, that is, $q^{\sim[\theta]}(2) + q^{\sim[\theta]}(3) = q^{\sim[\phi]}(2) = 0.6517$.

The challenge here is how we divide the probability for $q^{\sim[\phi]}(2)$, which is defined, among $q^{\sim[\theta]}(2)$ and $q^{\sim[\theta]}(3)$. The prior for ϕ yields no information to assist in this choice, because knowing which value ϕ takes on does not give us any information about whether $\theta = 2$ or $\theta = 3$. Thus, the information we have about θ must be taken from $q_1(\theta)$.

That is, we can assign a probability for $q^{\sim[\theta]}(2)$ by considering the probability that $\theta = 2$ relative to the probability $\theta = 3$, computing

$$q^{\sim[\theta]}(2) = q^{\sim[\phi]}(2)\Big(\frac{q_1(2)}{q_1(2) + q_1(3)}\Big).$$

That is, if the probability θ takes on the value 2 is lower in this case than the probability $\theta = 3$ which we know from the prior on θ, $q_1(\theta)$, then the pooled prior on θ, $q^{\sim[\theta]}(2)$, should reflect this.

Using this reasoning, we have $q^{\sim[\theta]}(2) = (0.7)\frac{0.2}{0.2+0.5} = 0.1862$ and $q^{\sim[\theta]}(3) = (0.7)\frac{0.5}{0.2+0.5} = 0.4655$.

The result in this simple example, using $q_1(\theta)$ to determine how to distribute the probability for values of ϕ where multiple θ map to ϕ, can be used to derive general

formulas to compute $q^{\sim[\theta]}(\theta)$ for discrete and continuous distributions (Poole & Raftery, 2000).

2.2.2 Implementation through the Sampling-Importance-Resampling Algorithm

The steps are:

1. We draw θ from its prior distribution $q_1(\theta)$, where we note θ can be multidimensional.
2. For every θ_i we compute $\phi_i = M(\theta_i)$.
3. Since $q_1^*(\phi)$ is unlikely to have an analytical form, we can compute it via a density approximation by computing $M(\theta)$ for our sampled values of θ and then estimating the density from this sample using kernel density estimation.
4. Construct weights proportional to the ratio of the prior on ϕ evaluated at $M(\theta_i)$ to the induced prior q_1^* evaluated at $M(\theta_i)$. Note that this is applying the same logic as considering $q_1(\theta)/q_1^*(\theta)$, as discussed in the previous concrete example, but representing these probabilities in ϕ space.
5. Sample values from step (1) with probabilities proportional to the weights from (4).

2.2.3 Bayesian Melding Applied to COVID-19 Misclassification

In this work, we can relate the inputs $\theta = \{P(S_1|untested), \alpha, \beta\}$ and $\phi = P(S_0|test+, untested)$ by the deterministic model $M : \theta \rightarrow \phi$ given by

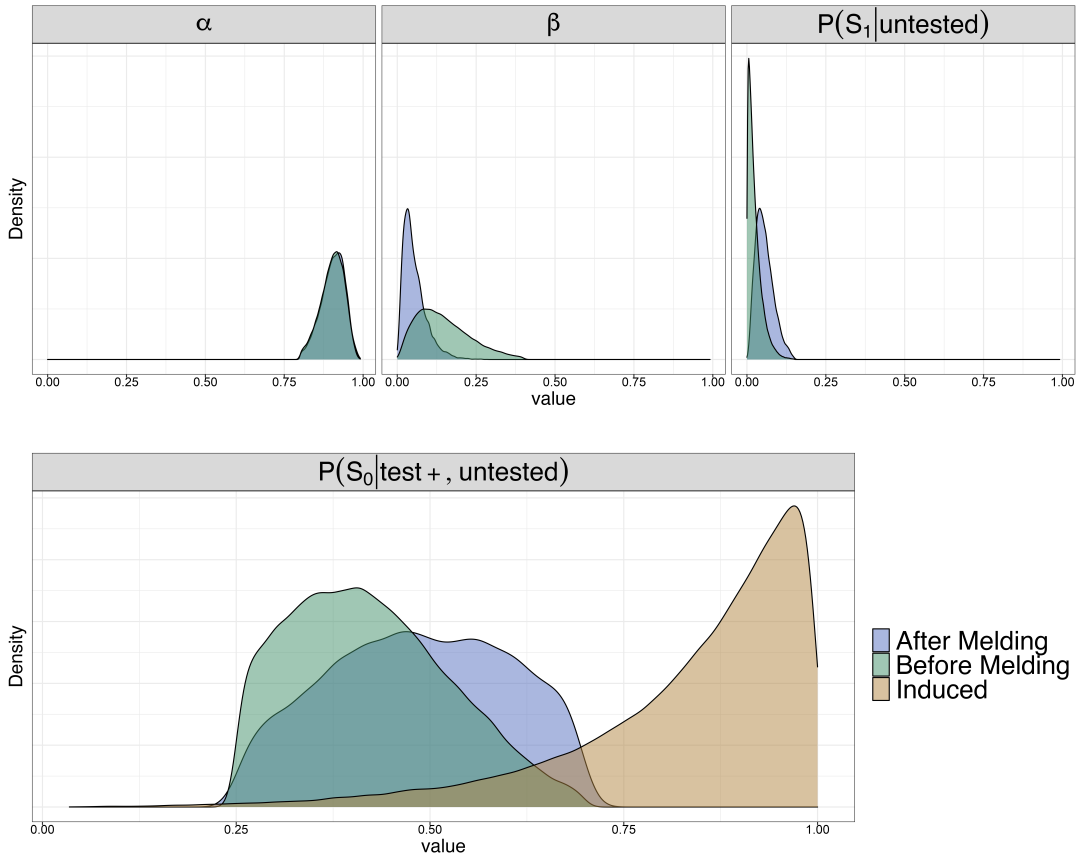
$$P(S_0|test+, untested) = \frac{\beta(1 - P(S_1|untested))}{\beta(1 - P(S_1|untested)) + \alpha P(S_1|untested)}.$$

The derivation of M is in the following section.

Now, we have two distributions on ϕ: the distribution based on data on the asymptomatic rate of infection of COVID-19, and the distribution formed by taking $M(\theta)$ where θ represents the values from the defined distributions of α, β, and $P(S_1|untested)$. With Bayesian melding, we pool these distributions using logarithmic pooling, and then implement the sampling-importance-resampling algorithm to obtain constrained distributions of the inputs θ that are in accordance with information about the asymptomatic rate of the virus.

Due to the uncertainty around our definitions of α and β, it is particularly useful to leverage the information we have about the asymptomatic rate of the virus $P(S_0|test+, untested)$ because a large collection of studies has been published in this area. In a meta-analysis pooling data from 95 studies, the pooled estimate among the confirmed population that was asymptomatic was 40.50% [95% CI, 33.50%-47.50%] (Ma et al., 2021). Another meta-analysis including 350 studies estimated the asymptomatic percentage as 36.9% [95% CI: 31.8 to 42.4%] (Sah et al., 2021).

To clarify the use of this method, we can look at the distributions before and after applying Bayesian melding.



Comparing these priors above, we see that although they have shared support, some values from the induced distribution we acquire by using M to generate values of ϕ from sampled values of θ are very unlikely to be in accordance with the information we know about SARS-CoV-2 asymptomatic infection. This is where Bayesian melding comes into play. Pooling these distributions enable us to take both the prior on $q_2(\phi)$ from published analyses on asymptomatic infection, and the induced prior, $q_1^*(\phi)$, into account to constrain the distributions of ϕ and θ to be in accordance.

2.2.4 Derivation of M

Recall that we have the function $M : \theta \rightarrow \phi$ that relates the random variables $\theta = \{P(S_1|untested), \alpha\}$ to the test positivity rate in the asymptomatic untested population, that is, $\phi = P(S_0|test_+, \text{untested})$. M is defined as

$$= P(S_0|test_+, \text{untested}) = \frac{\beta(1 - P(S_1|\text{untested}))}{\beta(1 - P(S_1|\text{untested})) + \alpha(P(S_1|\text{untested})}$$

Since we have $\alpha = \frac{P(test_+|S_1,\text{untested})}{P(test_+|tested)}$ and $\beta = \frac{P(test_+|S_0,\text{untested})}{P(test_+|tested)}$, we can write

$$= \frac{\frac{P(test_+|S_0,\text{untested})}{P(test_+|tested)}(1 - P(S_1|\text{untested}))}{\frac{P(test_+|S_0,\text{untested})}{P(test_+|tested)}(1 - P(S_1|\text{untested})) + \frac{P(test_+|S_1,\text{untested})}{P(test_+|tested)}P(S_1|\text{untested})}$$

and cancelling out the term $P(test_+|tested)$ we have

$$= \frac{P(test_+|S_0,\text{untested})(1 - P(S_1|\text{untested}))}{P(test_+|S_0,\text{untested})(1 - P(S_1|\text{untested})) + P(test_+|S_1,\text{untested})P(S_1|\text{untested})}.$$

Since $P(S_0|\text{untested}) = 1 - P(S_1|\text{untested})$,

$$= \frac{P(test_+|S_0,\text{untested})P(S_0|\text{untested})}{P(test_+|S_0,\text{untested})P(S_0|\text{untested}) + P(test_+|S_1,\text{untested})P(S_1|\text{untested})}.$$

Applying the definition of conditional probability to the term $\backslash P(test_+|S_0,\text{untested})P(S_0|\text{untest}$ in the numerator,

$$= \frac{\left(\frac{P(test_+,S_0,\text{untested})}{P(S_0,\text{untested})}\right)\left(\frac{P(S_0,\text{untested})}{P(\text{untested})}\right)}{P(test_+|S_0,\text{untested})P(S_0|\text{untested}) + P(test_+|S_1,\text{untested})P(S_1|\text{untested})}$$

$$= \frac{\left(\frac{P(test_+,S_0,\text{untested})}{P(S_0,\text{untested})}\right)\left(\frac{P(S_0,\text{untested})}{P(\text{untested})}\right)}{P(test_+|S_0,\text{untested})P(S_0|\text{untested}) + P(test_+|S_1,\text{untested})P(S_1|\text{untested})}$$

$$= \frac{\frac{P(test_+,S_0,\text{untested})}{P(\text{untested})}}{P(test_+|S_0,\text{untested})P(S_0|\text{untested}) + P(test_+|S_1,\text{untested})P(S_1|\text{untested})}$$

$$= \frac{P(test_+,S_0|\text{untested})}{P(test_+|S_0,\text{untested})P(S_0|\text{untested}) + P(test_+|S_1,\text{untested})P(S_1|\text{untested})}.$$

We can substitute this result in for the $P(test_+|S_0,\text{untested})P(S_0|\text{untested})$ term in the denominator to yield

$$= \frac{P(test_+,S_0|\text{untested})}{P(test_+,S_0|\text{untested}) + P(test_+|S_1,\text{untested})P(S_1|\text{untested})}$$

With same reasoning, we can simplify $\backslash P(test_+|S_1,\text{untested})P(S_1|\text{untested}) =$

$P(S_1, test_+|\text{untested})$, giving us

$$
\begin{aligned}
&= \frac{P(test_+, S_0|\text{untested})}{P(test_+, S_0|\text{untested}) + P(S_1, test_+|\text{untested})} \\
&= \frac{P(test_+, S_0|\text{untested})}{P(test_+|\text{untested})} \\
&= \frac{\dfrac{P(S_0, test_+, \text{untested})}{P(\text{untested})}}{\dfrac{P(test_+, \text{untested})}{P(\text{untested})}} \\
&= \frac{P(S_0, test_+, \text{untested})}{P(test_+, \text{untested})} \\
&= P(S_0|test_+, \text{untested}).
\end{aligned}
$$

Hence, we have

$$P(S_0|test_+, \text{untested}) = \frac{\beta(1 - P(S_1|\text{untested}))}{\beta(1 - P(S_1|\text{untested})) + \alpha(P(S_1|\text{untested})}$$

as desired. □

2.3 LOESS Smoothing

2.3.1 Introduction

Locally estimated scatterplot smoothing (LOESS) fits a collection of local regression models to obtain a smooth curve through the observed data (Figure 2.1). It is highly flexible in the sense that we do not have to specify the functional relationship between the predictor and response variable for the entire range of the predictor, which may be impossible in various settings. It is particularly with time series data with substantial noise.

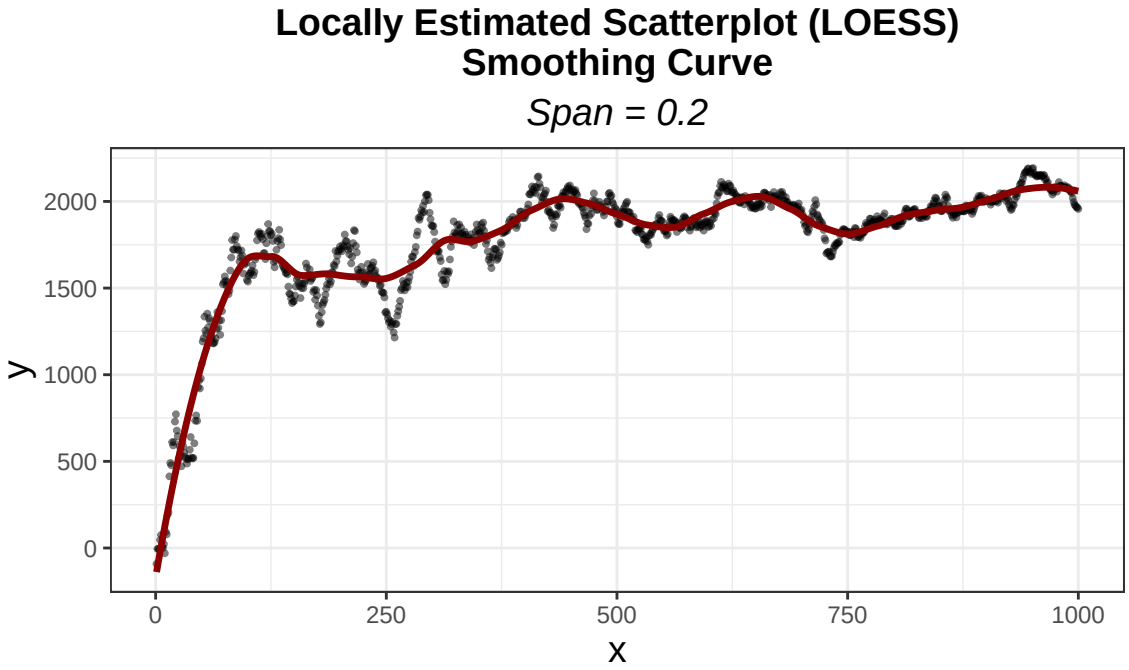


Figure 2.1

To perform LOESS smoothing, we estimate a set of local regressions (**chambers1997?**). To do this, we must specify the span; this smoothing parameter is the fraction of the data that is used for the local polynomial fit. With a smaller span, the resulting curve will fit the trends more closely, while a larger span reflects broader trends (Figure 2.2).

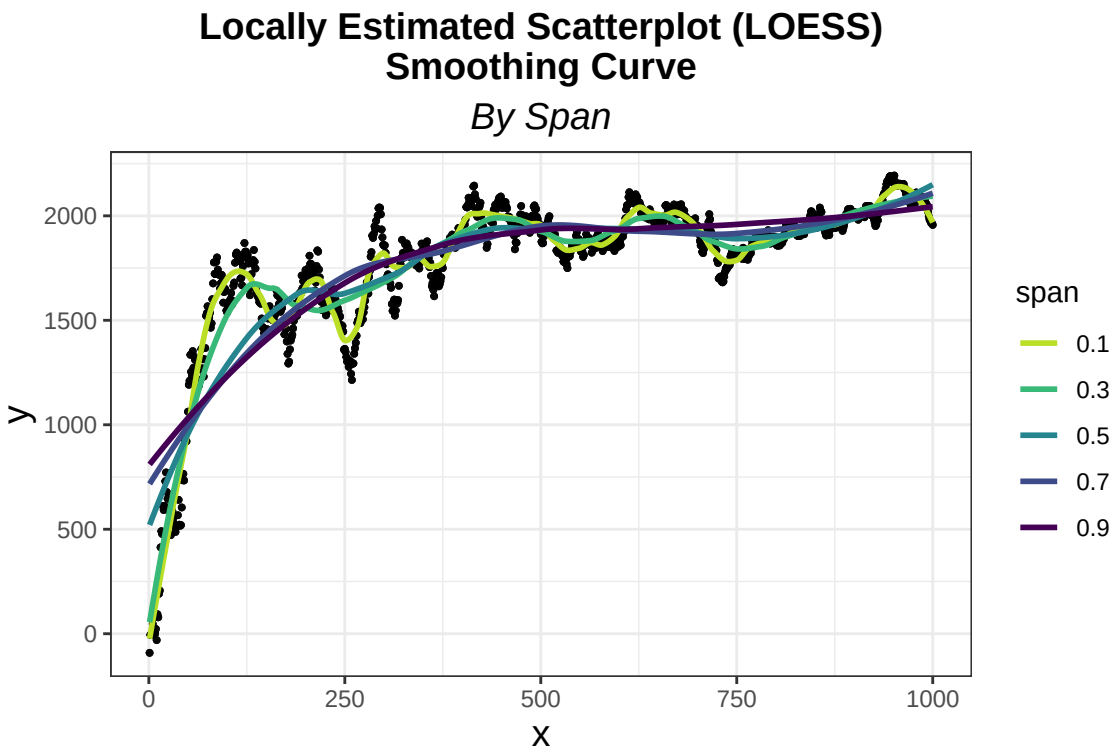


Figure 2.2

2.3.2 Fitting the LOESS Curve

To introduce some notation for the model at hand, we have a dependent variable y and independent variable x, where y and x are related by some unknown function g, that is, $y = g(x) + \boldsymbol{\epsilon}$[1]. When we want to use LOESS smoothing to estimate g, often this function is complex, so we break up the problem into estimating a set of local regressions.

To obtain a predicted value $\hat{g}(x^*)$ for a particular value of the independent variable x^*, we fit a polynomial with greatest weight placed on points in the neighborhood of x^*, where the width of this neighborhood is defined by the choice of smoothing span. Let $\alpha \in (0, 1]$ denote the chosen smoothing span.

For a particular value of x^*, we estimate the predicted value $\hat{g}(x^*)$ by fitting a local regression. We first compute the weights by computing the vector of distances from this point x^*, that is,

$$\Delta(x^*) = |\mathbf{x} - x^*|$$

We define $q = \text{floor}(\alpha n)$, and take $\Delta_q(x^*) \in \mathbb{R}$ to be the q^th smallest distance of $\Delta(x^*)$.

The vector of weights is then

$$T(\Delta(x^*), \Delta_q(x^*))$$

[1]Recall we use bold type for vectors, e.g., $\mathbf{x} \in \mathbb{R}^n$ is a vector with observations $x_i \in \mathbb{R}$.

where T is the tricube weight function given by

$$T(x) = \begin{cases} (1-(x)^3)^3 & \text{for } |x| < 1 \\ 0 & \text{for } |x| \geq 1 \end{cases}.$$

Essentially, this process gives weight to points in the neighborhood of x^*. Consider $x^* = 500$ and smoothing span $= \alpha = .2$.

Then the weights we obtain are given in Figure 2.3.

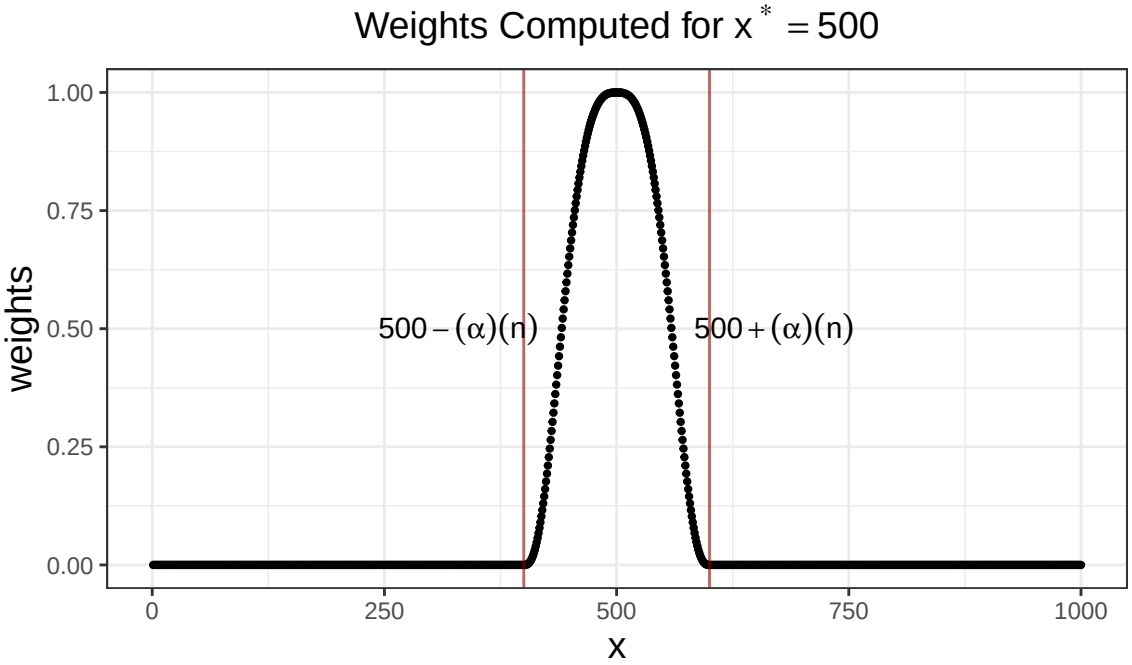


Figure 2.3: We see that the only values with nonzero weights are those within the interval $(500 - \alpha(n), 500 - \alpha(n))$, that is, the proportion α of the data points closest to x^*.

We fit a linear regression with polynomial terms, typically with degree up to 2, with these weights.

For example, fitting the model for this same $x^* = 500$, we obtain the polynomial in Figure 2.4.

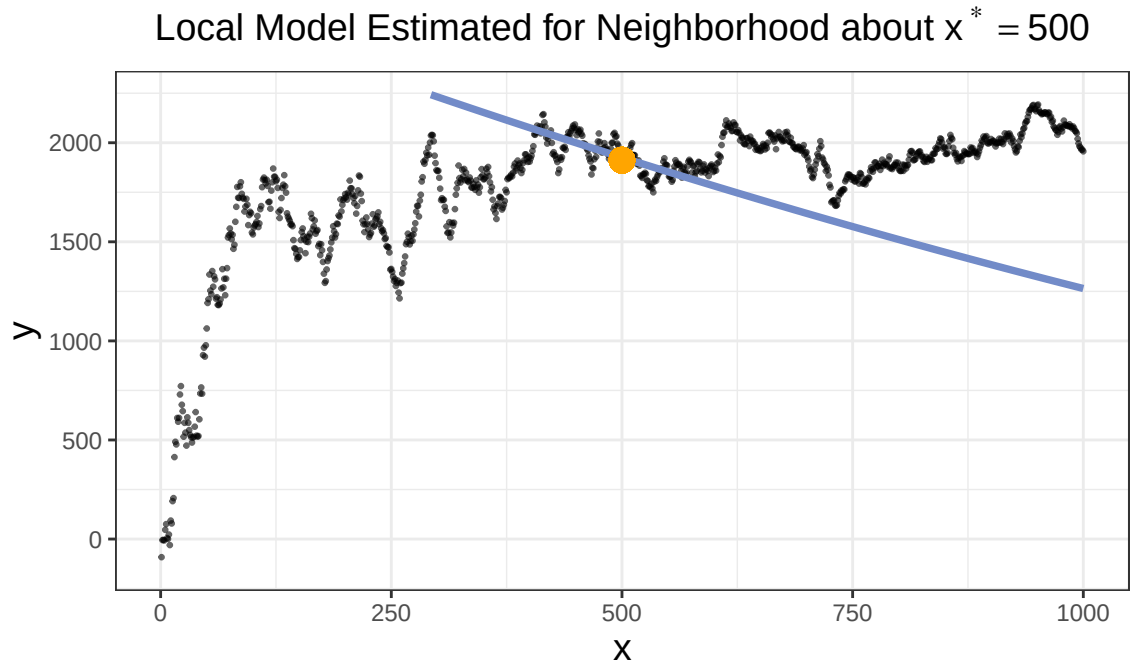


Figure 2.4

By fitting the model for every point in x, we obtain the smoothed line shown in red in Figure 2.5.

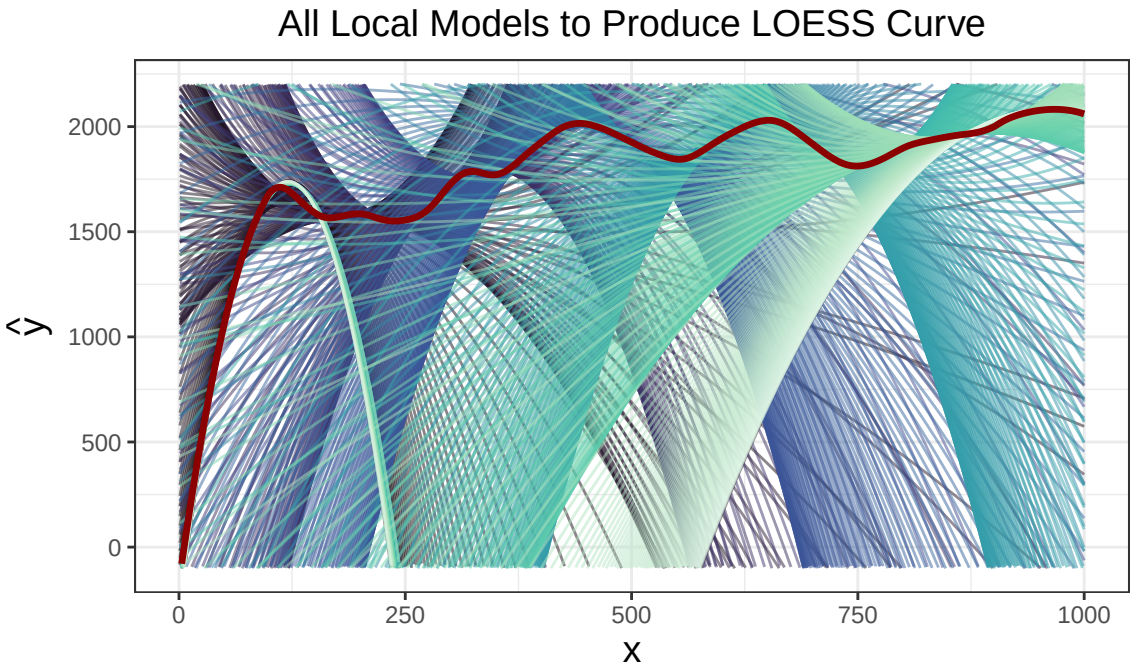


Figure 2.5

Smoothing methods are sensitive to the choice of smoothing parameter h,

which represents the fraction of the data that is used for the local polynomial fit.

Methods exist for picking the smoothing parameter h that minimizes the mean squared error between the predicted values from the estimated line and observed values of the dependent variable, for example, leave-one-out cross-validation or generalized cross-validation.

However, for this work, we used LOESS smoothing to smooth survey data from the COVID-19 Trends and Impact Survey (**reinhart2021?**). We choose the smoothing parameter for each variable based on domain knowledge regarding the level of noise present for each variable of interest. For example, there is substantial noise in the screening test positivity data that reflect trends that do not represent meaningful differences in the screening test positivity. Some trends in the screening sensitivity may be due to scheduled workplace screenings happening at regular time intervals, and some of the variation may be due to the frequency of screening testing due to other variables, such as the access and cost of testing.

Since the ratio $\frac{\text{screening test positivity}}{\text{overall test positivity}}$ is used to estimate $\beta = \frac{P(test+|S_0,untested)}{P(test+|tested)}$, the variability in the screening positivity creates substantial variability in our estimates of β.

In light of this variability and the presence of other trends regarding the screening test positivity, we set the span to $\frac{4}{12} = 0.33$ to fit the local regressions for 4-month intervals with the aim to capture the broader trends over time.

INCLUDE FIGURE OF SMOOTHED ESTIMATES HERE

There was less variabiity in the smoothing span for the weighted percentage of COVID-like Illness, the estimate of $P(S_1|untested)$. Hence, we set the smoothing parameter to 0.2 detect trends at a finer time scale.

Sensitivity analyses with modified versions of the smoothing span of β are included in the appendix in the section INCLUDE SECTION.

2.4 Kernel Density Estimation

2.5 Sampling Importance Resampling

Chapter 3

Definition of Prior Distributions for Bias Parameters

Placeholder

3.1 Background on the Beta Distribution

3.2 Background on the Gamma Distribution

3.3 Definition of Prior Distributions for Incomplete Testing Correction

3.3.1 Defining $P(S_1|Untested)$

3.3.2 Defining α

3.3.3 Defining β

3.3.4 Defining $P(S_0|test+, untested)$

3.4 Definition of Priors for Test Inaccuracy Correction

3.4.1 Defining Test Sensitivity (S_e)

3.5 Defining Test Specificity (S_p)

3.6 Summary Table of Bias Parameter Distributions

3.7 Correction for Incomplete Testing

3.8 Correction for Diagnostic Test Inaccuracy

3.8.1 Derivation of Formula for Correction for Diagnostic Test Inaccuracy

Chapter 4

Details of Implementation

- Describe each step here
- Mention reproducible workflow with make

Chapter 5

Comparison to the Covidestim Model

Placeholder

5.0.1 Overview

5.0.2 The Covidestim Model

5.0.3 Assumptions

5.1 Comparison to Other Indicators

5.2 Seropositivity Data

5.3 County-level

5.4 State-level

Appendix A

Appendix

A.1 Derivation of the Mean and Variance of the Beta Distribution

To add

References

Placeholder

Blitzstein, J. K., & Hwang, J. (2019). *Introduction to probability* (Second edition). Boca Raton: CRC Press.

Genest, C., McConway, K. J., & Schervish, M. J. (1986). Characterization of Externally Bayesian Pooling Operators. *The Annals of Statistics*, *14*(2), 487–501. Retrieved from https://www.jstor.org/stable/2241231

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. http://doi.org/10.1007/s10654-016-0149-3

Lash, T. L., Fox, M. P., & Fink, A. K. (2009). *Applying Quantitative Bias Analysis to Epidemiologic Data*. New York, NY: Springer New York. http://doi.org/10.1007/978-0-387-87959-8

Ma, Q., Liu, J., Liu, Q., Kang, L., Liu, R., Jing, W., … Liu, M. (2021). Global Percentage of Asymptomatic SARS-CoV-2 Infections Among the Tested Population and Individuals With Confirmed COVID-19 Diagnosis: A Systematic Review and Meta-analysis. *JAMA Network Open*, *4*(12), e2137257. http://doi.org/10.1001/jamanetworkopen.2021.37257

Neyman, J. (1937). Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, *236*(767), 333–380. http://doi.org/10.1098/rsta.1937.0005

Petersen, J. M., Ranker, L. R., Barnard-Mayers, R., MacLehose, R. F., & Fox, M. P. (2021). A systematic review of quantitative bias analysis applied to epidemiological research. *International Journal of Epidemiology*, *50*(5), 1708–1730. http://doi.org/10.1093/ije/dyab061

Poole, D., & Raftery, A. E. (2000). Inference for Deterministic Simulation Models: The Bayesian Melding Approach. *Journal of the American Statistical Association*, *95*(452), 1244–1255. http://doi.org/10.1080/01621459.2000.10474324

Powers, K. A., Ghani, A. C., Miller, W. C., Hoffman, I. F., Pettifor, A. E., Kamanga, G., … Cohen, M. S. (2011). The role of acute and early HIV infection in the spread of HIV and implications for transmission prevention strategies in Lilongwe, Malawi: a modelling study. *The Lancet*, *378*(9787), 256–268. http://doi.org/10.1016/S0140-6736(11)60842-8

Robson, B. J. (2014). When do aquatic systems models provide useful predictions, what is changing, and what is next? *Environmental Modelling & Software*, *61*, 287–296. http://doi.org/10.1016/j.envsoft.2014.01.009

Sah, P., Fitzpatrick, M. C., Zimmer, C. F., Abdollahi, E., Juden-Kelly, L., Moghadas, S. M., … Galvani, A. P. (2021). Asymptomatic SARS-CoV-2 infection: A systematic review and meta-analysis. *Proceedings of the National Academy of Sciences*, *118*(34), e2109229118. http://doi.org/10.1073/pnas.2109229118

Ševčíková, H., Raftery, A. E., & Waddell, P. A. (2007). Assessing uncertainty in urban simulations using Bayesian melding. *Transportation Research Part B: Methodological*, *41*(6), 652–669. http://doi.org/10.1016/j.trb.2006.11.001