ExtimatingCOVID-19 Prevalence and Seroprevalence in the United States

Weihsueh A. Chiu, Martial L. Ndeffo-Mbah

Presenter: *Martial Ndeffo-Mbah, Ph.D. Texas A&M University*
Email: *m.ndeffo@tamu.edu*

TEXAS A&M UNIVERSITY

Dr. Ndeffo short Resume

I am currently an Assistant Professor in Epidemiology at Texas A&M College of Veterinary Medicine and Biomedical Sciences and an Adjunct Assistant Professor of Epidemiology at the Texas A&M School of Public Health. Previously, I was a Research Scientist at the Yale School of Public Health and associate director of the Yale Center for Infectious Disease Modeling and Analysis. I use a range of mathematical, statistical, socio-economic, and optimization modeling approaches to address scientific questions and public policy challenges for infectious diseases transmission, prevention, and control. I have expertise in developing, calibrating, and validating data-driven disease transmission models, and developing optimization and cost-effectiveness frameworks for infectious disease surveillance and control. A focus of my research program has been the development and application of these modeling frameworks to improving our understanding of the spread and control of wide range of infectious diseases including Influenza, Neglected Tropical diseases, Ebola, Vector-borne diseases, and more recently COVID-19.

Estimating COVID-19 Prevalence and Seroprevalence in the USA

Timely and reliable estimates of COVID-19 prevalence and seroprevalence are paramount for evaluating the spread and control of the pandemic in different US states. Relying on reported cases and test positivity rates individually can result in incorrect inferences as to the spread of COVID-19 and ill-informed public health decision-making. On the other hand, Mathematical models are generally complex models that require substantial quantitative skills and extensive data and are generally perceived as “black boxes” by many public health practitioners and decision makers.

Our study developed a simple semi-empirical model for estimating state-level prevalence and seroprevalence of COVID-19 in the United States using reported case and test positivity rates data. We found that due to the preferential nature of diagnostic COVID-19 testing in the US, the geometric mean of reported case and test positivity rates is an accurate predictor of undiagnosed COVID-19 prevalence and trends.

Conceptual Basis of the Semi-empirical Model

1: Observations

a) Reported cases and test positivity rate have been widely used to infer prevalence. But they provide highly biased estimates.

b) Test positivity rate is correlated to the lagged prevalence of undiagnosed COVID-19 infection in the population with a time-dependent bias parameter

2: Assumption

We assume large-scale passive testing as a baseline testing rate, in US states. So, increase of testing rate (e.g. active testing) will preferentially increase the infected population testing rate relative to the general population testing rate. This results in a "diminishing return" $b(t) = \left[\frac{N_{\text{test},r}(t)}{N}\right]^{-n}$, with $n \in [0,1]$, N is population size, N_{test} is number of tests administered. $b(t)$ is the bias parameter correlation between lagged undiagnosed prevalence and test positivity rate.

3: Objective

Develop a simple semi-empirical model to estimate state-level prevalence and seroprevalence of COVID-19 in the US based only on reported cases and test positivity rate.

4: Model equations

$$\frac{I_U(t-t_{\text{lag}})}{N} = P_{+,c}(t)b(t)^{-1} = P_{+,c}(t)^{1-n}\left[\frac{N_{+,c}(t)}{N}\right]^n \equiv P_{+,c}(t)^{1-n} \times C_{+,c}(t)^n$$

Where

I_U/N is the undiagnosed prevalence, $P_{+,c}$ t is the positivity rate

$N_{+,c}$ is the number of reported cases, $C_{+,c}(t) = N_{+,c}(t)/N$.

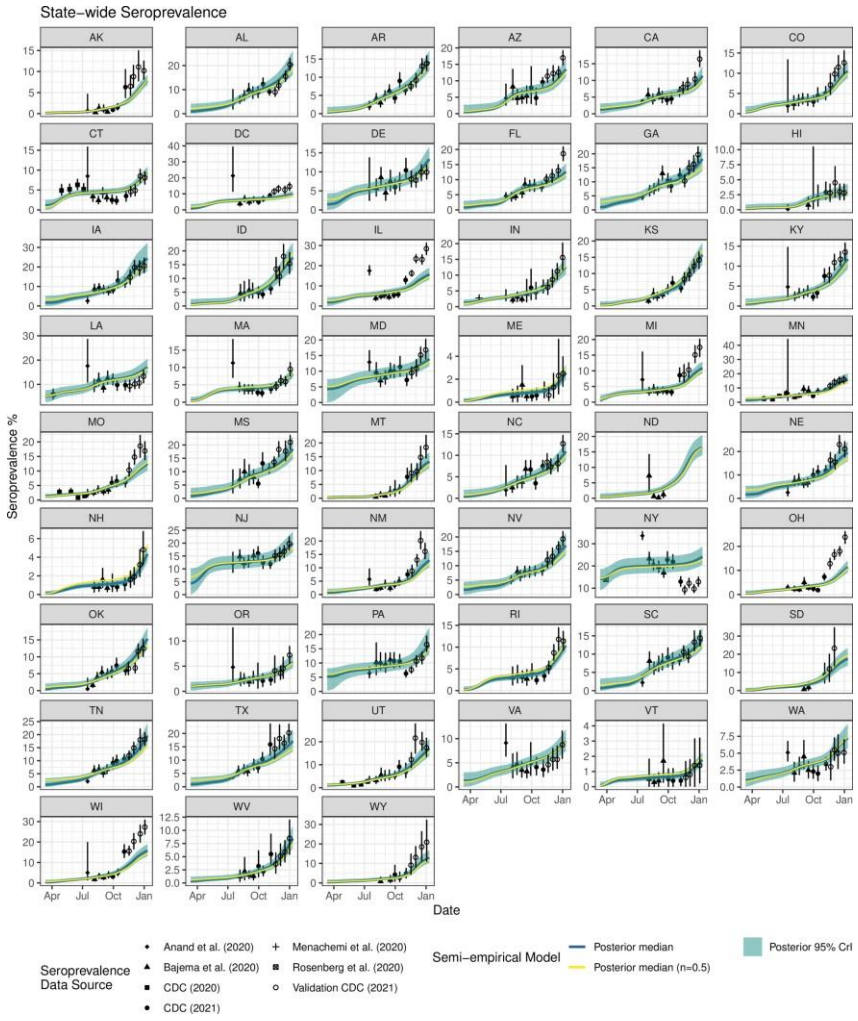
$$\frac{\text{SP}_\text{U}(t-t_{\text{lag}})}{N} = \frac{\text{SP}_\text{o}}{N} + \sum_{t'<t} \frac{P_{+,c}(t')^{1-n} \times C_{+,c}(t')^n}{T_{\text{etf}}(t')}.$$

where SP_U is undiagnosed seroprevalence, T_{eff} depends on time from infection to seropositivity, the power parameter n, and the initial seroprevalence SP_0.

Chiu WA, Ndeffo-Mbah ML (2021). PLoS Comput Biol 17(9): e1009374

Model Fitting to Empirical State-level Seroprevalence data

Model fitting and validation to data



States may differ in testing rate, as well as in the extent to which the initial surge of cases went unreported:

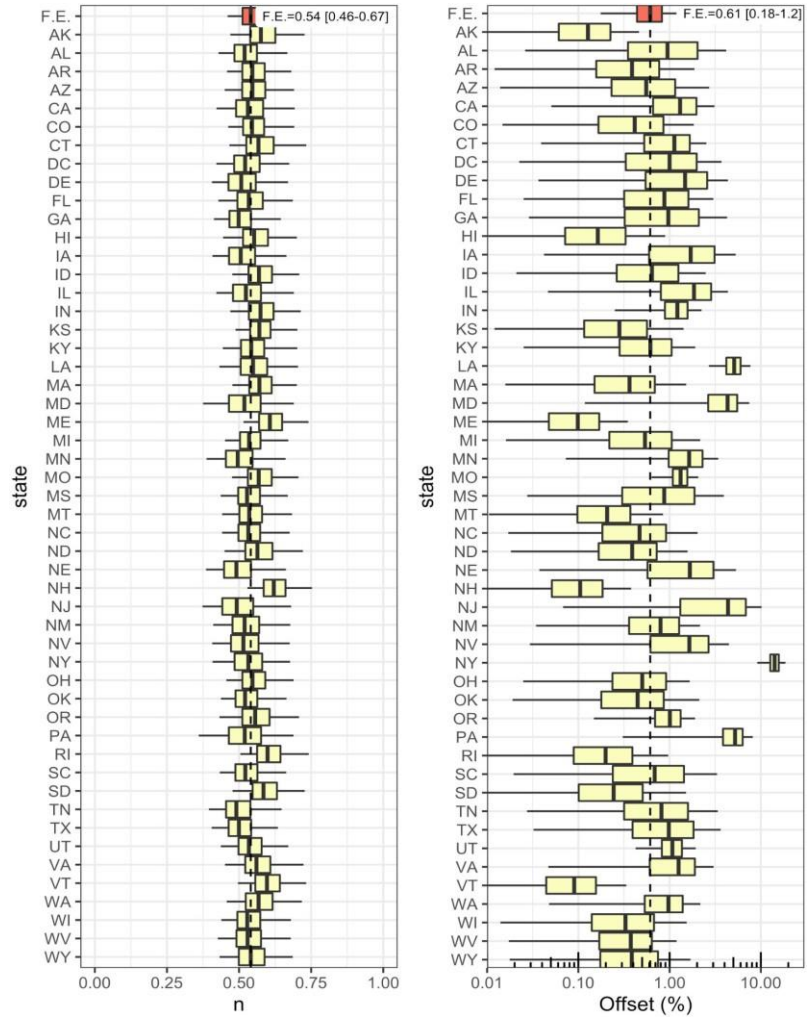
we assume random effects for the power parameter *n* and SP_o.

SP_o for most states was < 1%, but three states: New York (NY), Pennsylvania (PA), Louisiana (LA) > 5%.

NY = 14% [95% CrI: 8.9%-18.6%],
PA = 5.2% [95% CrI: 0.3%-8.1%],
LA = 5.1% [95% CrI: 2.6%-7.7%].

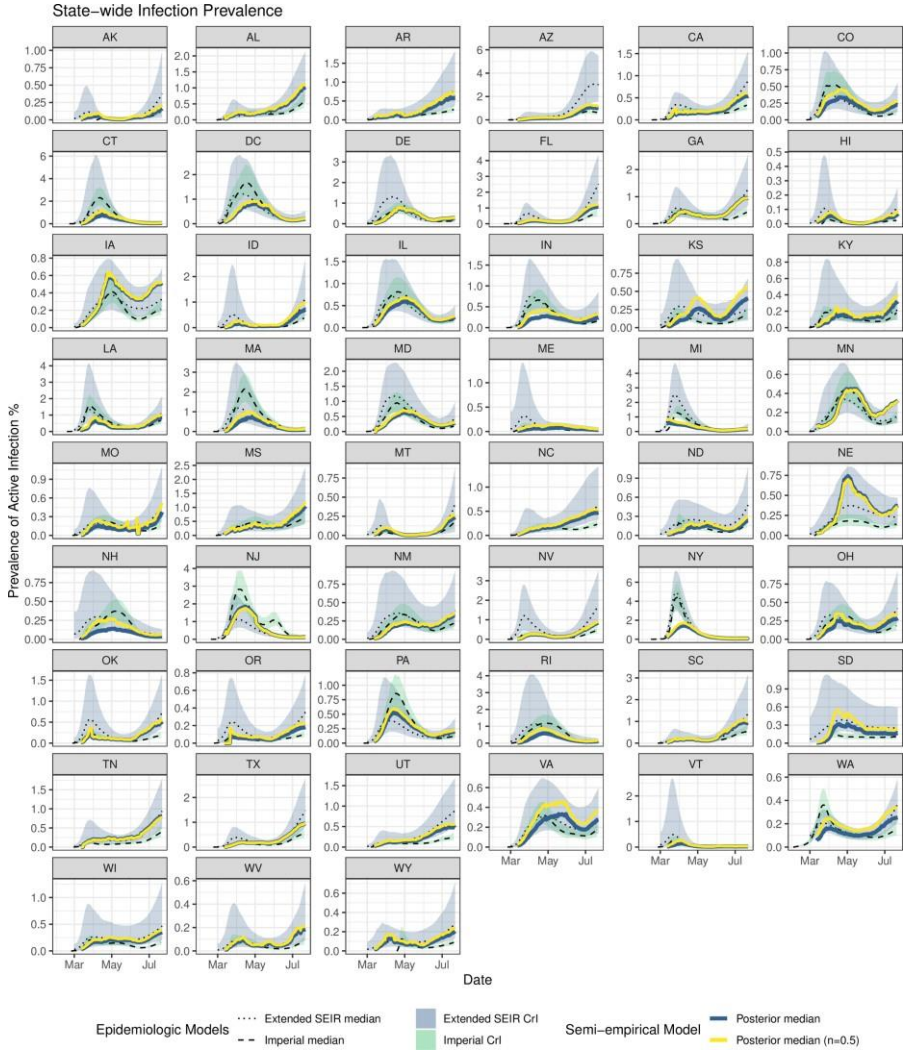
High SP_o are consistent with large surges of cases and death-to-case ratio in the early days of the pandemic when COVID-19 testing was highly limited and likely to have missed large surge of cases.

Estimates of *n* and SP_o parameters



Chiu WA, Ndeffo-Mbah ML (2021). PLoS Comput Biol 17(9): e1009374

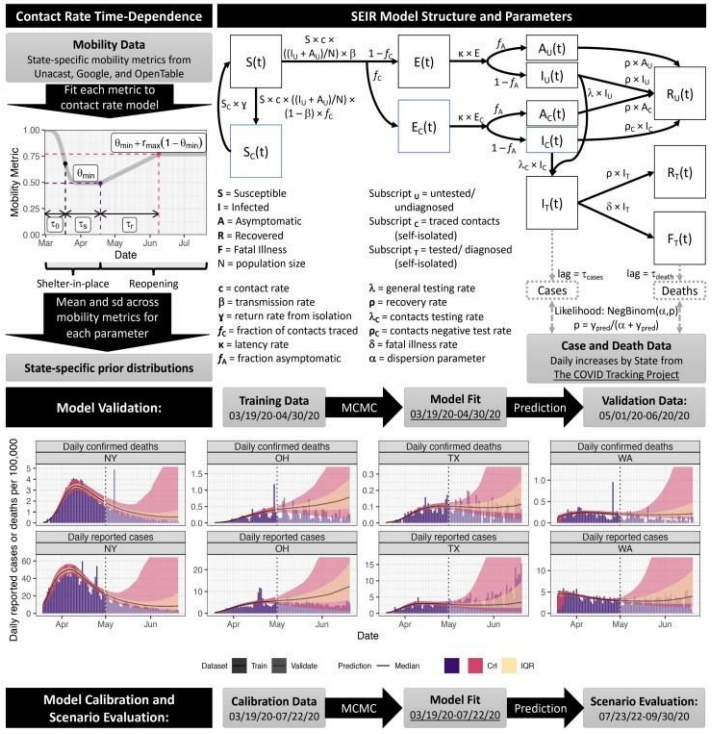
Comparison of Semi-empirical Model Against two Mathematical Models



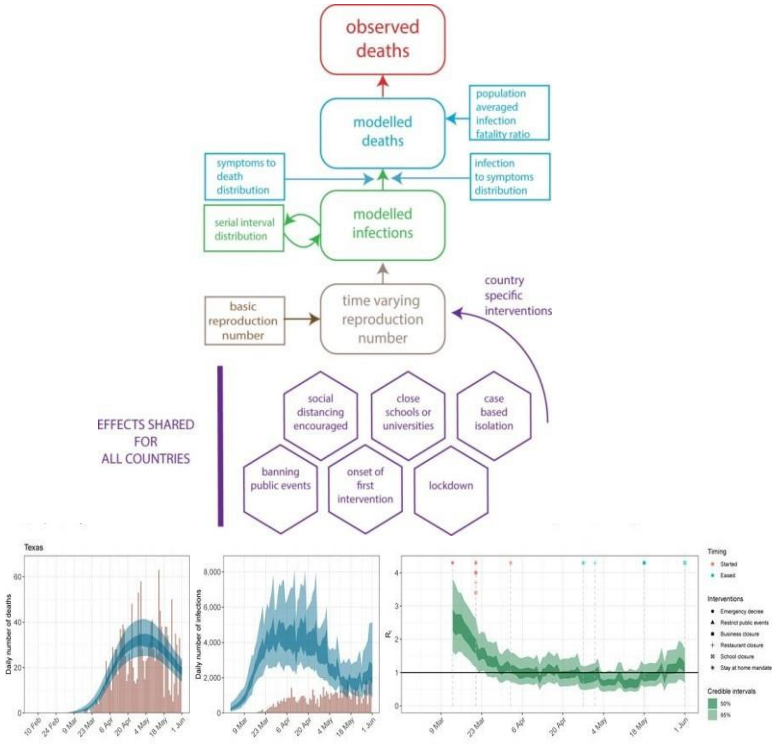
The semi-empirical estimate of infection prevalence is consistent with that of mathematical models

Extended-SEIR (Susceptible-Exposed-Infectious-Recovered) Model

Nat Hum Behav **4,** 1080–1090 (2020)



Imperial model *Nat Commun* **11,** 6189 (2020)

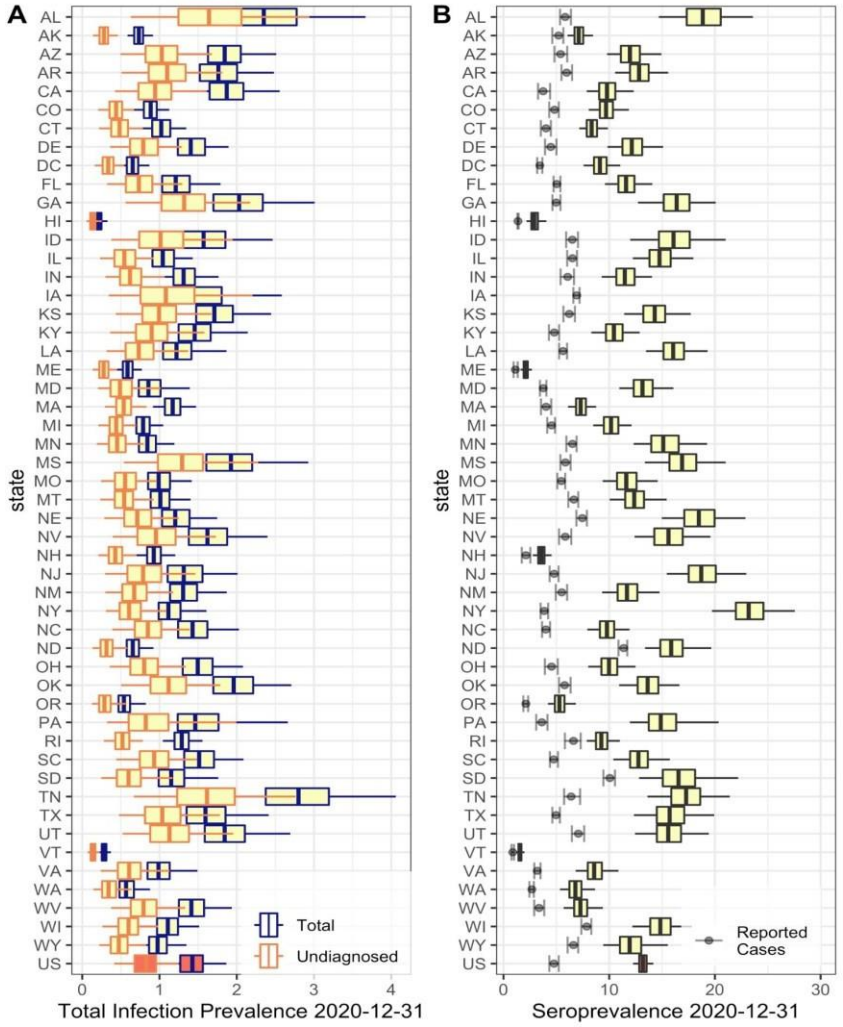


The Residual Standard Error (RSE) difference between the semi-empirical estimates and the extended-SEIR model has a 75% coefficient of variation and a R^2 of 0.68. A 74% coefficient of variation and 0.68 R^2 between the semi-mechanistic model and the Imperial model, and a 63% coefficient of variation between the extended-SEIR and Imperial model.

Chiu WA, Ndeffo-Mbah ML (2021). PLoS Comput Biol 17(9): e1009374

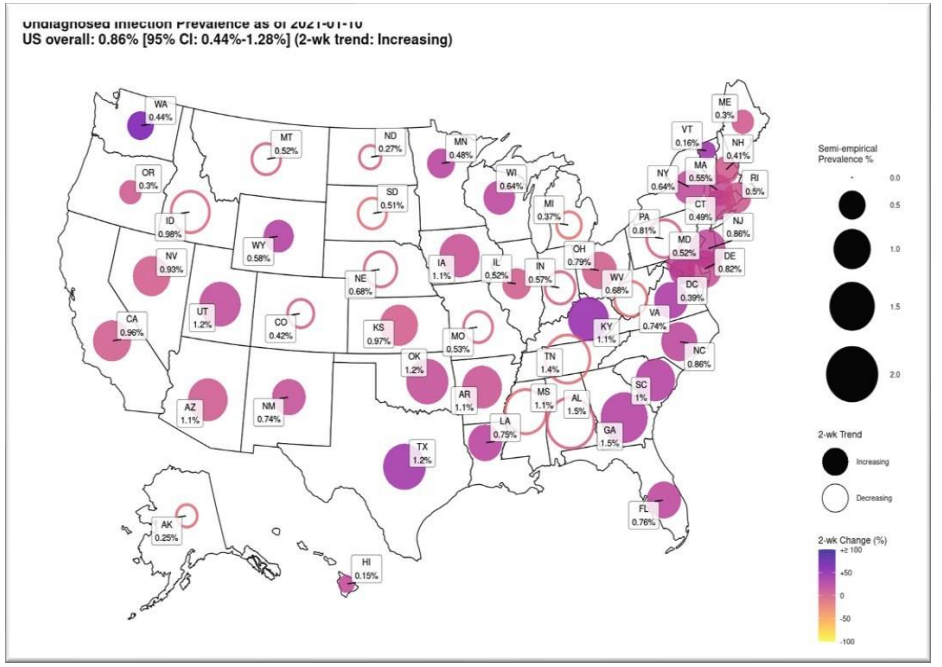
Estimates State-level Prevalence and Seroprevalence

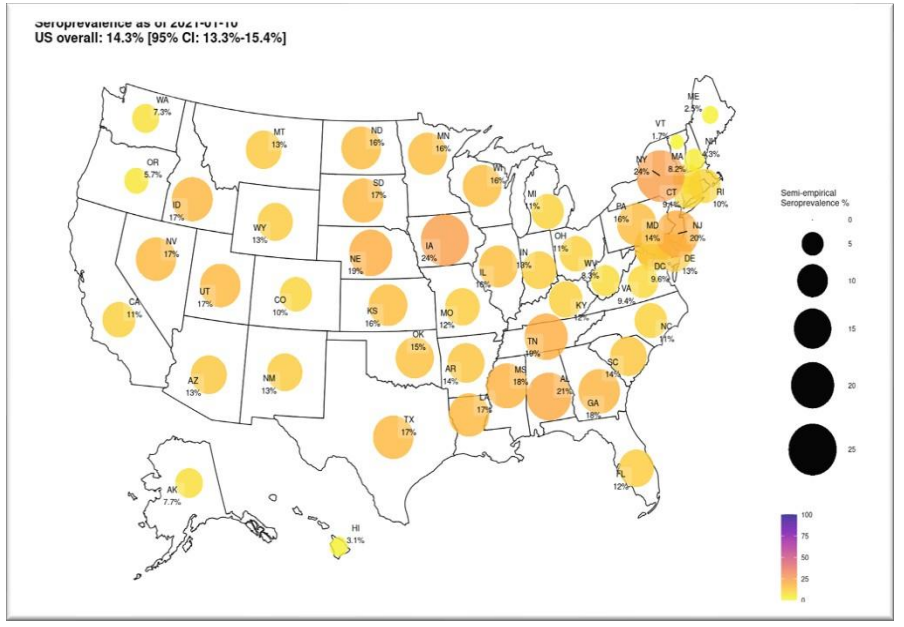
As of December 31th 2020, our semi-empirical model estimates that COVID-19 prevalence in the USA was 1.43% [CrI: 0.99%-1.86%], with more than half undiagnosed (0.83% [0.41%-1.25%]), and a seroprevalence of 13.2% [CrI: 12.3%-14.2%].



State-level undiagnosed prevalence estimates ranged from 0.15% in Hawaii to 1.5% in Georgia

State-level seroprevalence estimates ranged from 1.7% in Vermont to 24% in New York





Take-home Messages

- Test positivity and case reported rates are not good indicators of disease prevalence and transmission
- A simple and easy-to-communicate semi-empirical model can be used to reliably inform real-time state-level COVID-19 prevalence and seroprevalence in the US
- This simple modeling approach can be applied to other settings conditioned upon data availability to train the model
- The model needs to be extended to account for the impact of vaccination

The code is on
Github (https://github.com/wachiuphd/COVID-19-US-Semi-Empirical)
GitLab (https://gitlab.com/steveguillouzic/covid-wqm)