Semi-parametric modeling of SARS-CoV-2 transmission in Orange County, California using tests, cases, deaths, and seroprevalence data

Damon Bayer - University of California, Irvine - bayerd@uci.edu - @damonbayer

Motivation

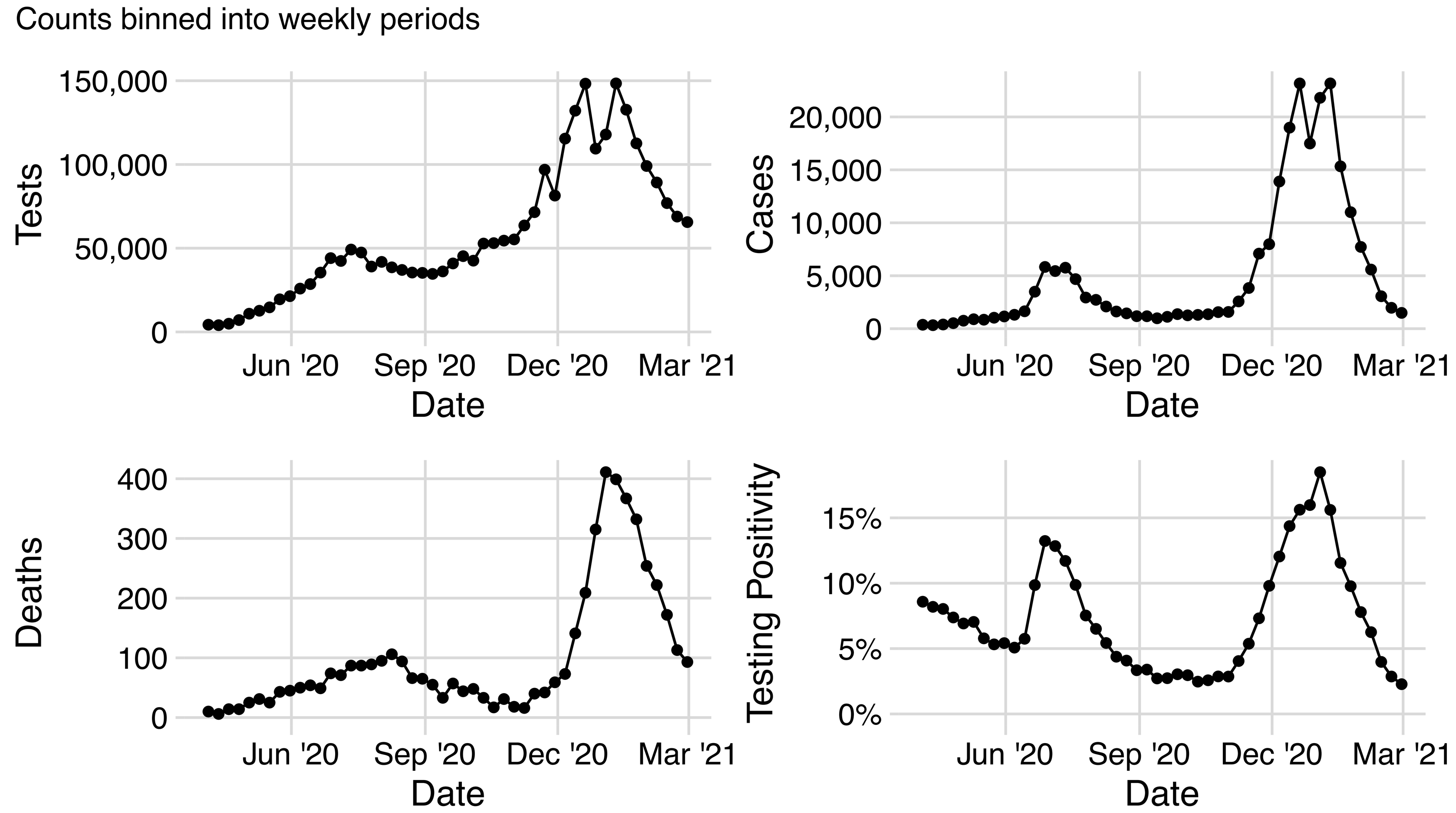
- Objectives:
  - Incorporate **multiple streams** of data into a COVID-19 model.
  - Account for **changes in testing** over time.
- Why?
  - Case data is the tip of the spear, earliest sign of increase in transmission.
  - Multiple data sources might lead to more robust inferences.

Data

- Modeling period: Mar 30, 2020 - Feb 28, 2021 (No vaccination)
- Daily cases, tests, and deaths aggregated at weekly level
- Excluded repeat positive tests
- Countywide seroprevalence study from ~Aug 2020: 11.5% positivity among 3,000 tests
- Empirically estimated reporting delay for deaths due to COVID-19

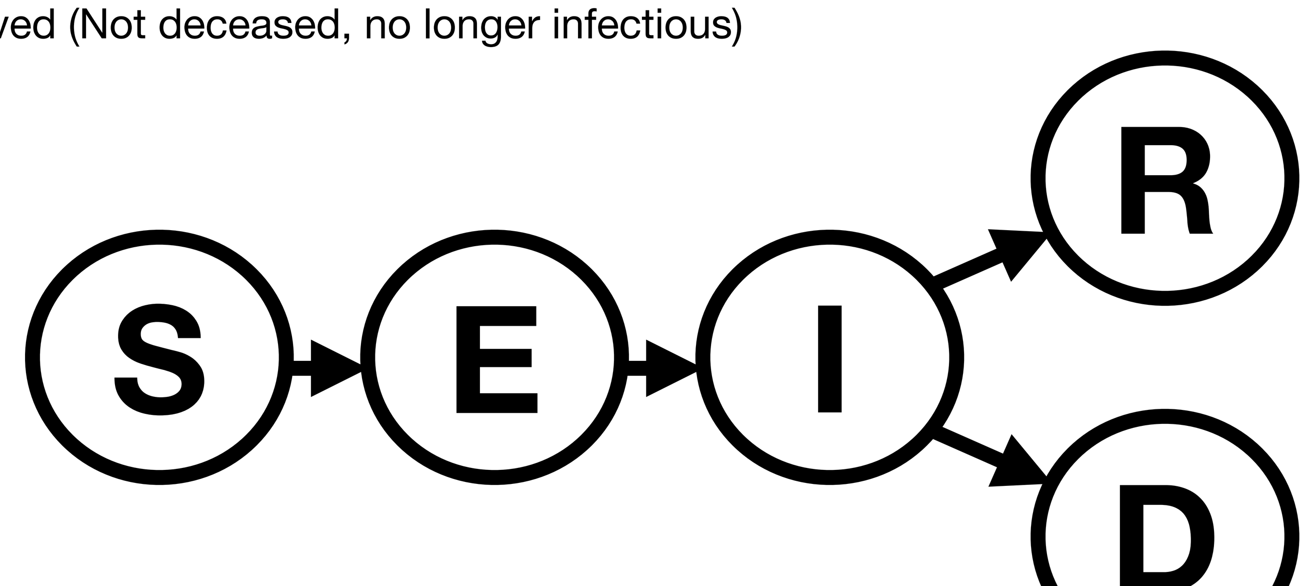
Orange County, CA data

Counts binned into weekly periods



Transmission Model

- S: Susceptible
- E: Infected but not Infectious
- I: Infectious,
- D: Death due to COVID-19
- R: Removed (Not deceased, no longer infectious)



Key Parameters:
Constant: Mean Latent Period, Mean Infectious Period
Time Varying: Basic Reproduction Number (R_0), Infection Fatality Ratio (IFR)

Surveillance Model

$$\text{Obs. Deaths}_l \sim \text{Negative Binomial}\left(\mu_l^D = (\text{New Latent Deaths})_l \cdot (\text{Report Prob.})_l, \sigma_l^2 = \frac{\mu_l^D(1+\mu_l^D)}{\text{Overdisp.}}\right)$$

$$\text{Obs. Cases}_l \sim \text{Beta-Binomial}\left(\text{Tests}_l, (\text{Overdisp.}) \cdot \mu_l^C, (\text{Overdisp.}) \cdot \left(1-\mu_l^C\right)\right)$$

$$\text{logit}\left(\mu_l^C\right) = \alpha_l + \text{logit}\left(\frac{(\text{New Latent Infections})_l}{\text{Population Size}}\right)$$

$$\text{Seroprev. Cases}_l \sim \text{Binomial}\left(\text{Seroprev. Tests}, \frac{(\text{Cumulative Latent Recoveries})_l}{\text{Population Size} - (\text{Cumulative Latent Deaths})_l}\right)$$

Bayesian Model

- $\boldsymbol{\theta}$: vector of model parameters
- $\boldsymbol{Y}$: vector of observed cases
- $\boldsymbol{M}$: vector of observed deaths
- $\boldsymbol{Z}$: vector of observed seroprev. cases
- Likelihood:

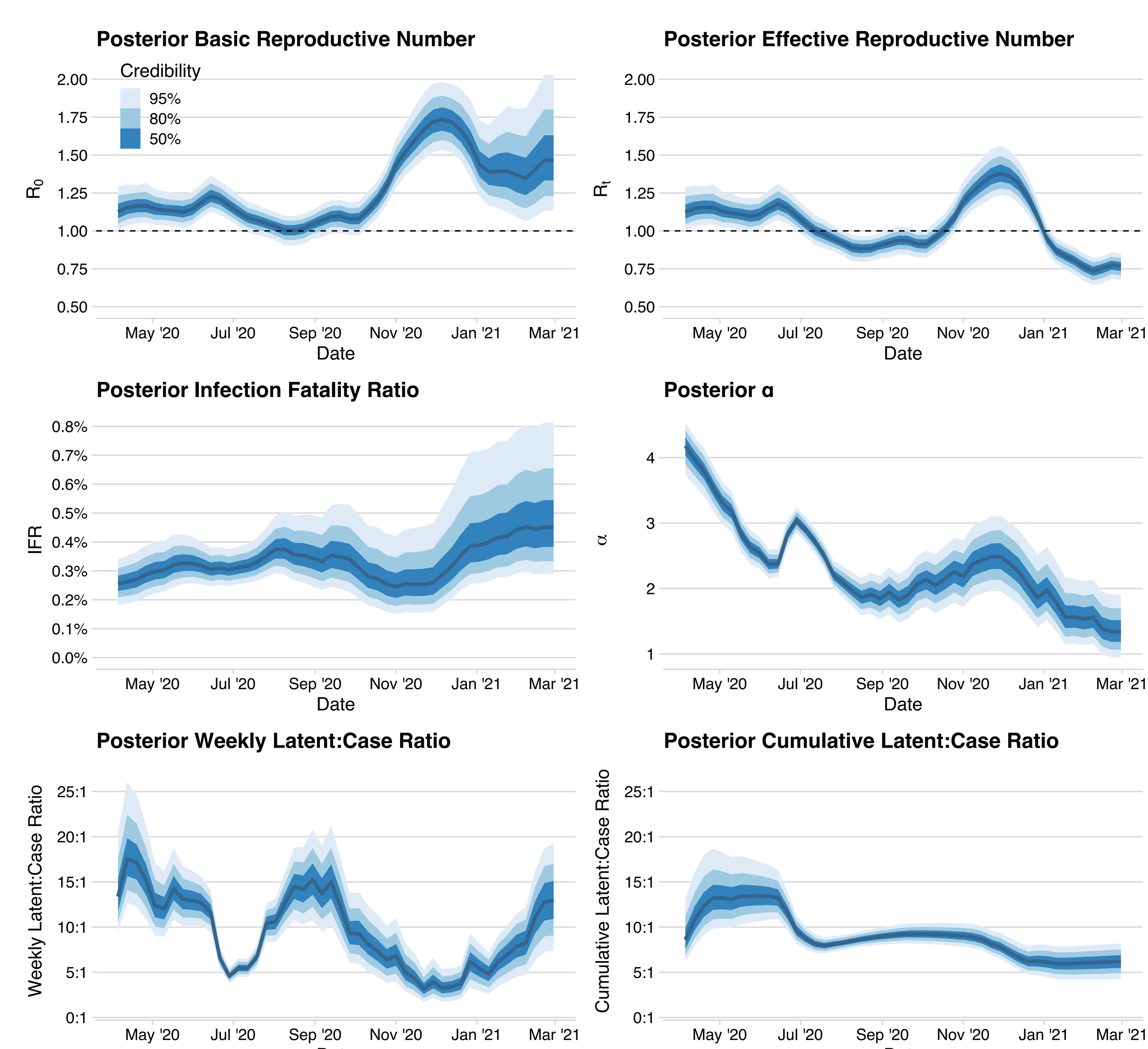
$$\Pr\left(\boldsymbol{M},\boldsymbol{Y},\boldsymbol{Z} \mid \boldsymbol{\theta}\right) = \Pr\left(\boldsymbol{M} \mid \boldsymbol{\theta}\right)\Pr\left(\boldsymbol{Y} \mid \boldsymbol{\theta}\right)\Pr\left(\boldsymbol{Z} \mid \boldsymbol{\theta}\right) = \prod_{l=1}^{L}\Pr\left(M_l \mid \boldsymbol{\theta}\right)\Pr\left(Y_l \mid \boldsymbol{\theta}\right)\Pr\left(Z_l \mid \boldsymbol{\theta}\right)$$

- Posterior:

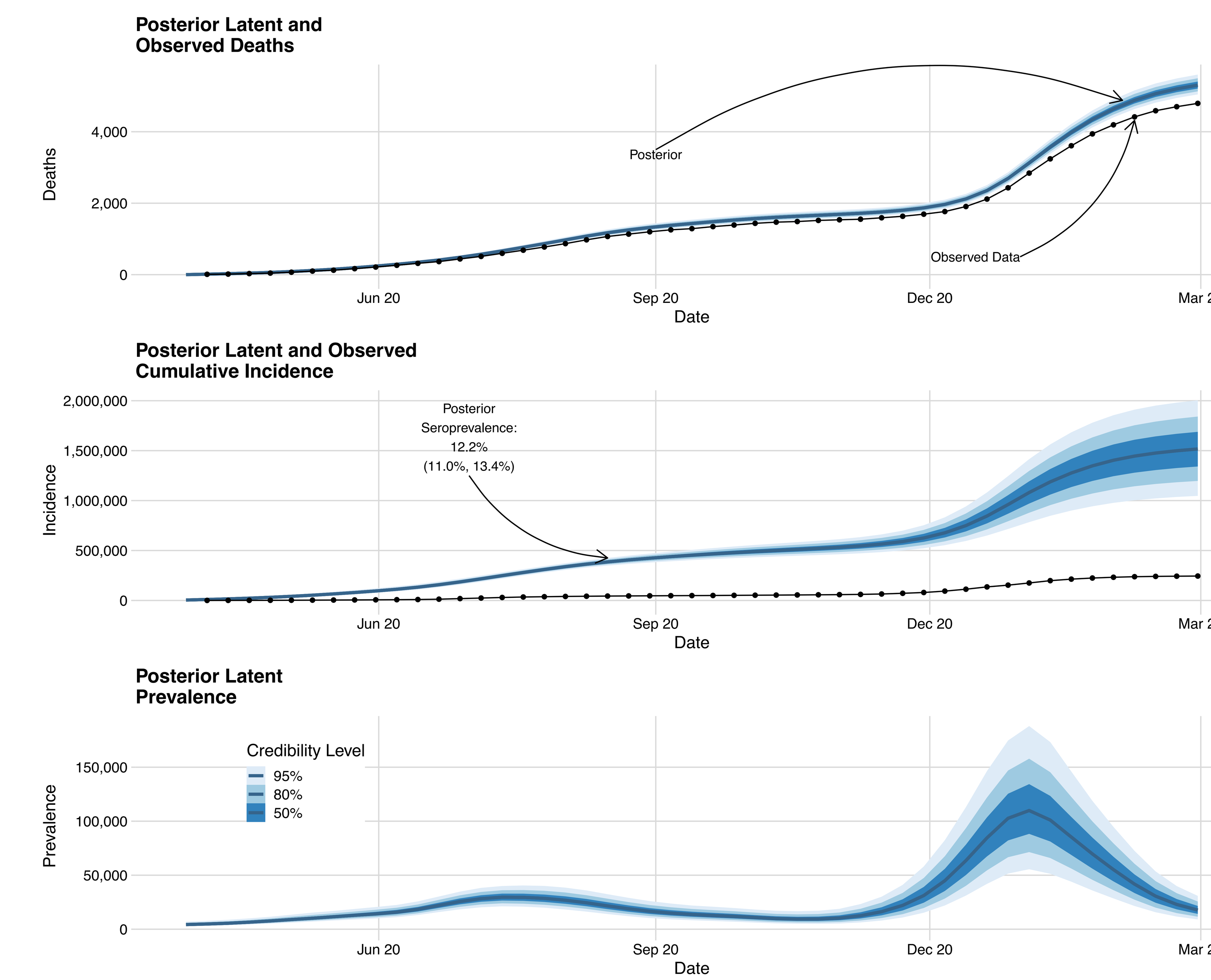
$$\Pr\left(\boldsymbol{\theta} \mid \boldsymbol{M},\boldsymbol{Y},\boldsymbol{Z}\right) \propto \Pr\left(\boldsymbol{M},\boldsymbol{Y},\boldsymbol{Z} \mid \boldsymbol{\theta}\right)\Pr(\boldsymbol{\theta})$$

- Fit using stemr R package https://fintzij.github.io/stemr/

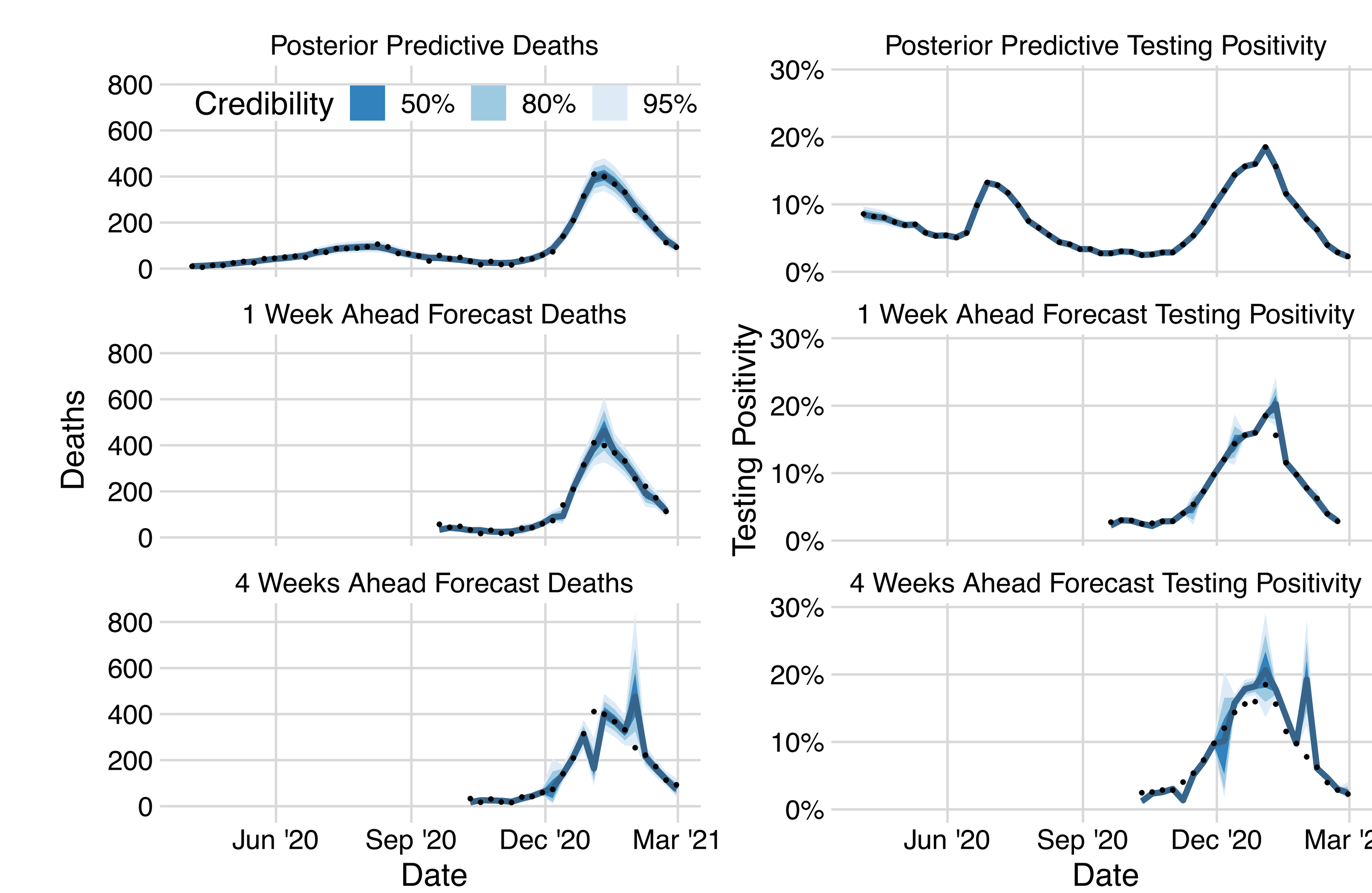
Results



Results



Posterior Predictive Distributions & Probabilistic Forecasts



- Sudden decrease in prevalence in January 2021 is attributable both to **high levels of accumulated immunity** and some **behavioral change**.
- Winter wave began in October 2021.
- 1/4-1/8 of cases are detected.
- By March 2021 1/3-2/3 of OC residents had been infected.
- Four weeks ahead forecasting is feasible when model parameters are stable.
- Preprint: https://arxiv.org/abs/2009.02654