

An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China

Song Lab, Department of Biostatistics
U-M School of Public Health

March 4, 2020

Objectives

- **AIM 1:** Utilize publicly available data from the China CDC to understand the trend of the coronavirus epidemic in Hubei Province and the other regions of China.
- **AIM 2:** Incorporate time-varying quarantine protocols in the modeling of infection dynamics using the classical SIR model (a system of ordinary differential equations) for infectious disease. **This dynamic infection system necessitates forecast the future trend and turning points of this epidemic.**
- **AIM 3:** Provide a health informatics toolbox via the R software to the public for their own analyses of disease spreading patterns using their own data.

Highlights of the Epidemiological Model and Software

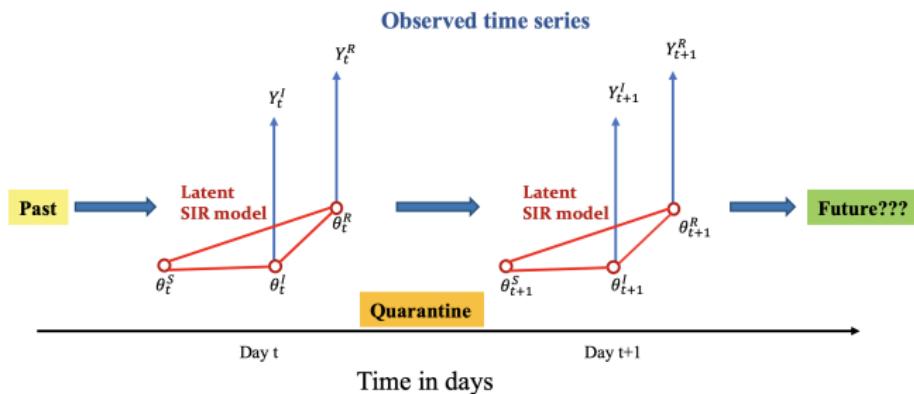
- Model time series of observed proportions of infected cases and removed cases (consisting of both recovered and death) emitted from the underlying Markov SIR infection dynamics.
- Implement estimation and prediction with quantification of uncertainty by Markov Chain Monte Carlo (MCMC) algorithm.
- Obtain posteriors of model parameters and predicted turning points with respective credible intervals.
- Create an R software package **eSIR**.

Epidemiological SIR Model with Time-varying Quarantines

Data: Observed the daily **proportions** of **infected cases** and **removed** (i.e. recovered and death) cases, respectively, Y_t^I and Y_t^R . We don't observe the proportion of **susceptible cases**.

Latent Markov Transition Processes: Assume the **probabilities** of susceptible, infected, and removed $\theta_t^S, \theta_t^I, \theta_t^R$ are Markovian such that $\theta_t^S + \theta_t^I + \theta_t^R = 1$.

Incorporate **time-varying quarantine protocols** in the SIR model



SIR Model: Latent Dynamic System of Infection



$$\frac{d\theta_t^S}{dt} = -\beta\theta_t^S\theta_t^I, \quad \frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I, \quad \frac{d\theta_t^R}{dt} = \gamma\theta_t^I,$$

- $\beta > 0$ is the disease transmission rate, $\gamma > 0$ is the removal rate
- $R_0 = \beta/\gamma$ is the reproduction number that indicates the expected number of cases generated by one infected case without intervention.

All People Stay in Three Rooms

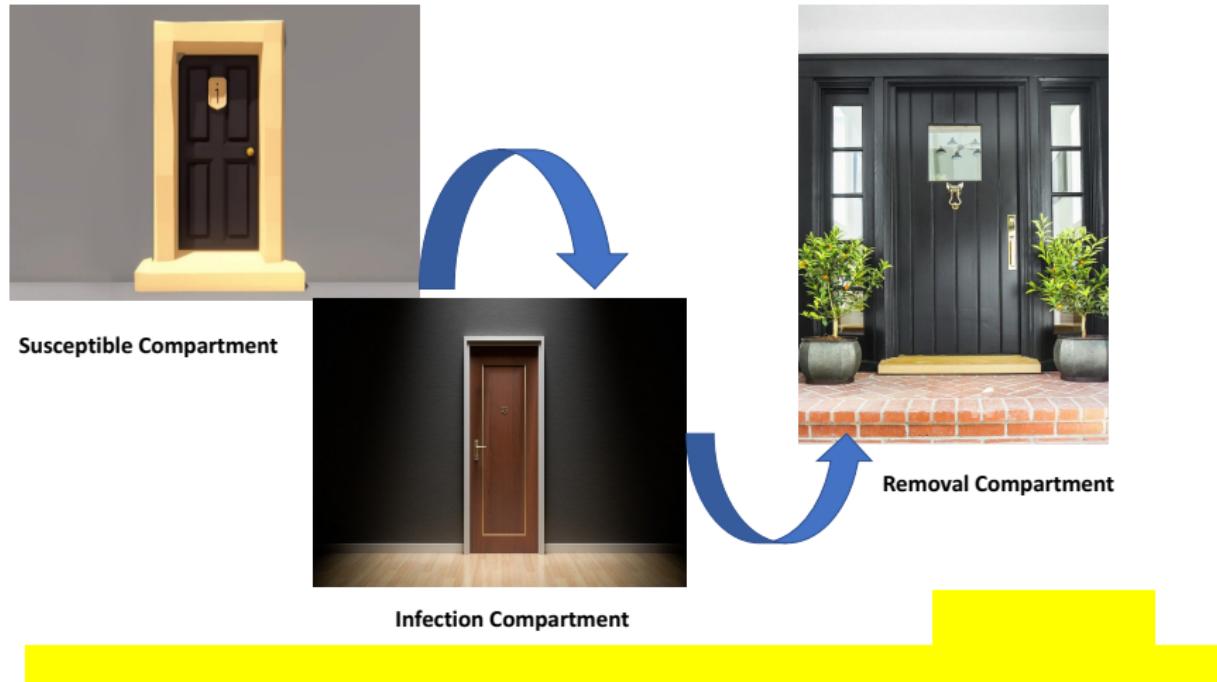


Figure: The proportions of people in the rooms are unknown.

Epidemiological SIR Model: Learning Infection Dynamics Using Data

Let Y_t^I and Y_t^R be the observed proportions of infection and removal state at time t . We assume Y_t^I and Y_t^R follows a **Beta-Dirichlet state-space model**, consisting of two observation processes:

$$Y_t^I | \boldsymbol{\theta}_t, \boldsymbol{\tau} \sim \text{Beta}(\lambda^I \theta_t^I, \lambda^I(1 - \theta_t^I)),$$
$$Y_t^R | \boldsymbol{\theta}_t, \boldsymbol{\tau} \sim \text{Beta}(\lambda^R \theta_t^R, \lambda^R(1 - \theta_t^R)),$$

where θ_t^I and θ_t^R are the respective **expected proportions**; the latent Markov SIR process:

$$\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}, \boldsymbol{\tau} \sim \text{Dirichlet}(\kappa f(\boldsymbol{\theta}_{t-1}, \beta, \gamma)),$$

where $\boldsymbol{\theta}_t = (\theta_t^S, \theta_t^I, \theta_t^R)^\top$ is the vector of the underlying population-level probabilities of susceptible, infectious and removed, with the **mean** as a function of $f(\boldsymbol{\theta}_{t-1}, \beta, \gamma)$, and $\boldsymbol{\tau} = (\beta, \gamma, \boldsymbol{\theta}_0^\top, \lambda, \kappa)^\top$ with λ^I , λ^R and κ being parameters controlling respective variability of the observation and latent processes.

Runge-Kutta Solution of SIR Model

$f(\cdot)$ is the mean transition probability determined by the SIR dynamic system, whose solution is given by:

$$\frac{d\theta_t^S}{dt} = -\beta\theta_t^S\theta_t^I, \quad \frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I, \quad \frac{d\theta_t^R}{dt} = \gamma\theta_t^I$$

By the fourth order Runge-Kutta(RK4) approximation:

$$f(\theta_{t-1}, \beta, \gamma) = \begin{pmatrix} \theta_{t-1}^S + 1/6[k_{t-1}^{S_1} + 2k_{t-1}^{S_2} + 2k_{t-1}^{S_3} + k_{t-1}^{S_4}] \\ \theta_{t-1}^I + 1/6[k_{t-1}^{I_1} + 2k_{t-1}^{I_2} + 2k_{t-1}^{I_3} + k_{t-1}^{I_4}] \\ \theta_{t-1}^R + 1/6[k_{t-1}^{R_1} + 2k_{t-1}^{R_2} + 2k_{t-1}^{R_3} + k_{t-1}^{R_4}] \end{pmatrix}.$$

Priors in the Epidemiological SIR Model

In the Markov Chain Monte Carlo implementation, we specified the following priors:

$$\theta_0^I \sim \text{Beta}(1, (Y_1^I)^{-1}), \quad \theta_0^R \sim \text{Beta}(1, (Y_1^R)^{-1}), \quad \theta_0^S = 1 - \theta_0^R - \theta_0^I.$$

The hyper-parameters used in the prior distributions of reproduction number, transmission rate and removal rate are obtained by the analysis of the Hong Kong SAS data.

$$R_0 \sim \text{LogN}(1.099, 0.096) \Rightarrow E(R_0) = 3.15, \text{SD}(R_0) = 1;$$

$$\gamma \sim \text{LogN}(-2.955, 0.910) \Rightarrow E(\gamma) = 0.0117, \text{SD}(\gamma) = 0.1;$$

$$\beta = R_0 \gamma.$$

The default sets large variances in both observation and latent processes, which may be adjusted over the course of epidemic with more data become available.

$$\kappa \sim \text{Gamma}(2, 0.0001), \lambda^I \sim \text{Gamma}(2, 0.0001), \lambda^R \sim \text{Gamma}(2, 0.0001).$$

Markov Chain Monte Carlo (MCMC) Algorithm

Let t_0 be the last date of data availability. Forecast spans over a period of T days over $[t_0 + 1, T]$.

From the posteriors, we first take M draws from $[\boldsymbol{\theta}_{1:t_0}, \tau | Y_{1:t_0}]$ to learn the SIR dynamics from day 1 to day t_0 .

Then for each solution path $m = 1, \dots, M$, we extrapolate the SIR trajectory in future T days over $[t_0 + 1, T]$ by the MCMC algorithm:

- (i) Draw $\boldsymbol{\theta}_t^{(m)}$ from $[\boldsymbol{\theta}_t | \boldsymbol{\theta}_{t-1}^{(m)}, \tau^{(m)}]$, $t = t_0 + 1, t_0 + 2, \dots, T$,
- (ii) Draw $Y_t^{(m)}$ from $[Y_t | \boldsymbol{\theta}_t^{(m)}, \tau^{(m)}]$, $t = t_0 + 1, t_0 + 2, \dots, T$;

Output: posterior draws from $[\boldsymbol{\theta}_{1:t_0:T} | Y_{1:t_0}]$ and predicted proportions $[Y_{(t_0+1):T} | Y_{1:t_0}]$

MCMC Algorithm: Start with Observed Data

Observed data

- (y_t^I, y_t^R)

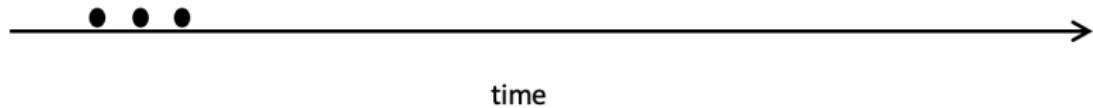


Figure: Data collection

MCMC Algorithm: Learning the SIR Dynamics

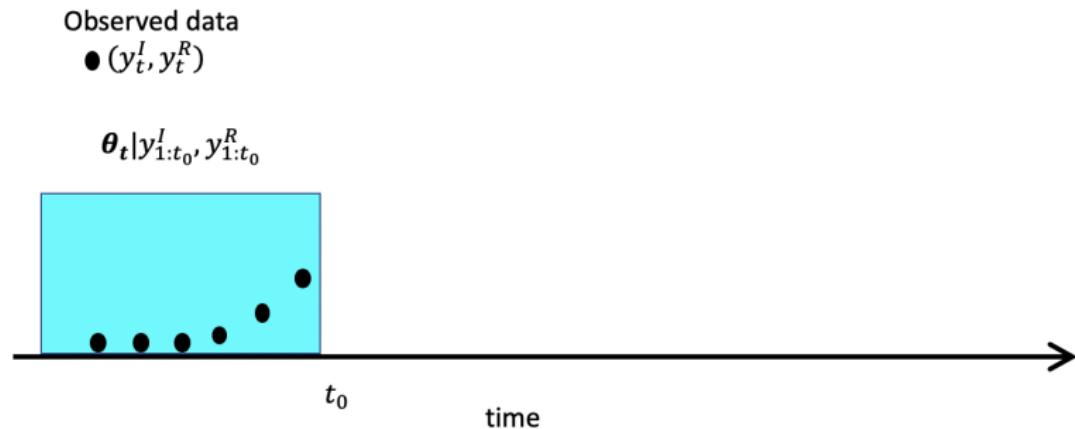


Figure: Posterior distributions of latent processes $\theta_{1:t_0}$ and parameters (β, γ)

MCMC Algorithm: Extrapolating the SIR Dynamics for forecast

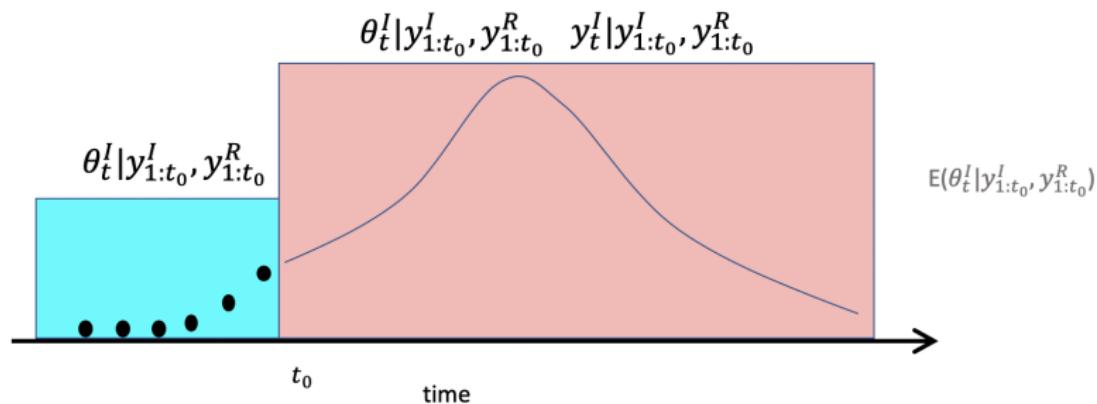


Figure: Forecast with the mean probability curve of prevalence (black line) that emits predicted proportion of infected cases (salmon color).

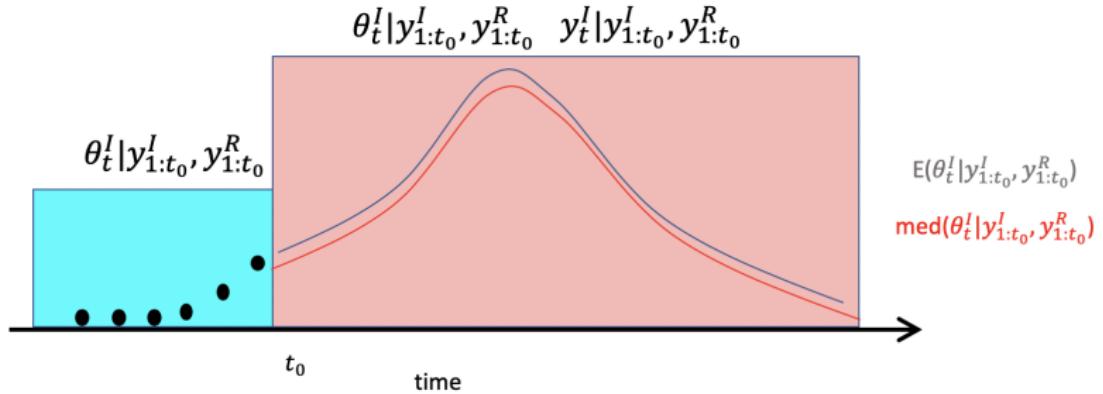
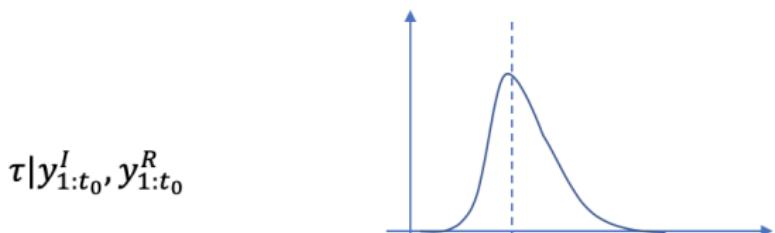
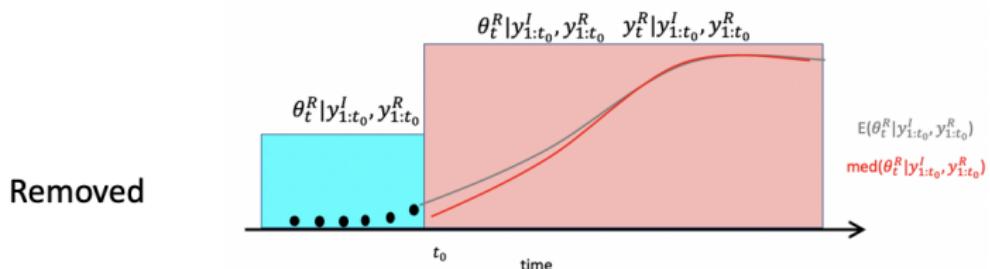
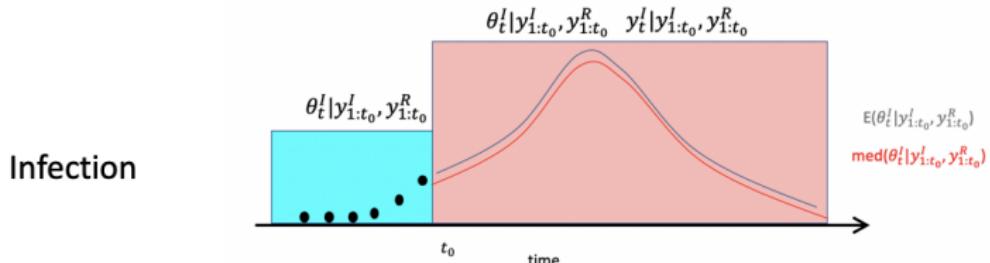


Figure: Forecast with the mean and median probability curves (black and red lines) that emits predicted proportions of infected cases (salmon color).

Forecasts of Both Infection and Removal Compartments



The Last Date of Data Availability t_0

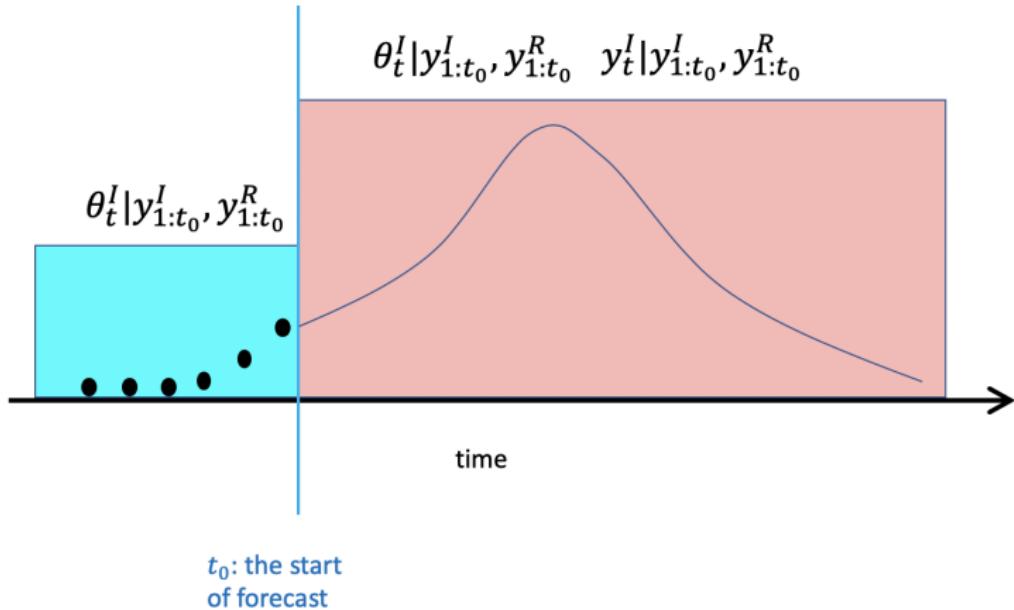
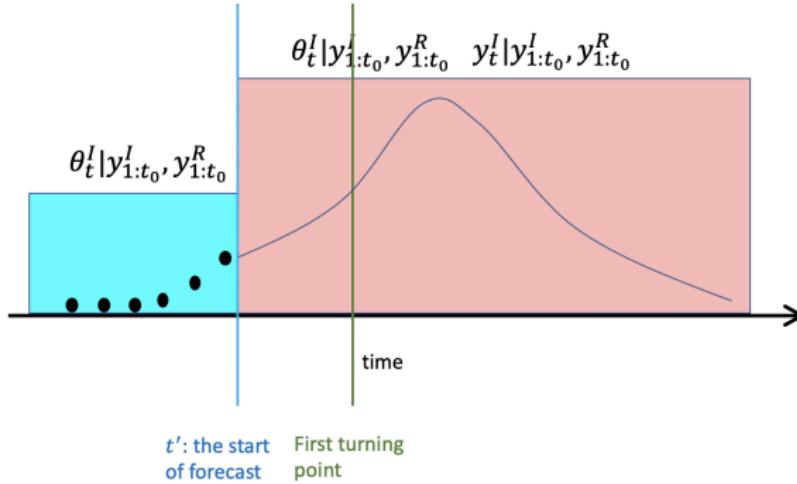


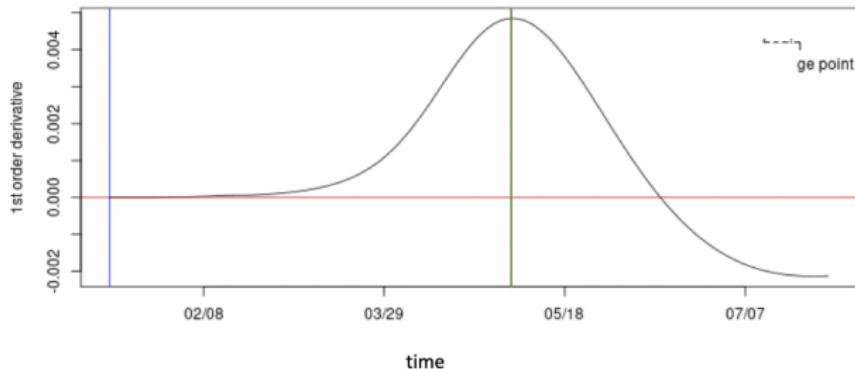
Figure: Key time points of interest

First Turning Point of Interest: Zero Acceleration



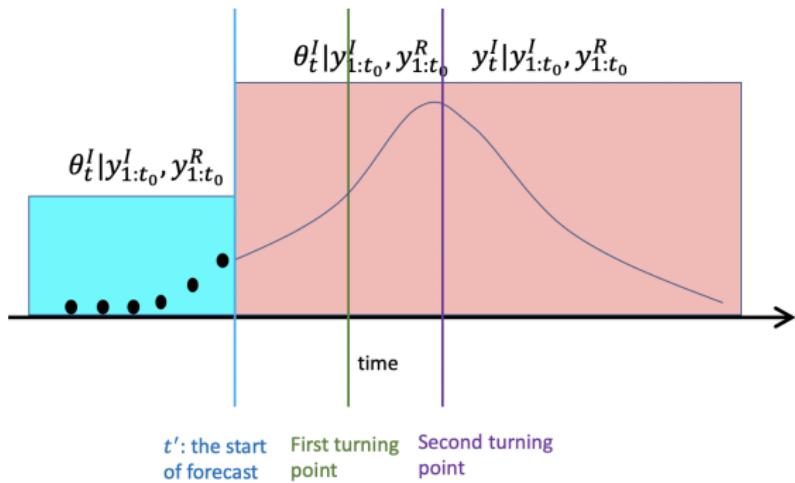
- The turning point at which the acceleration becomes zero, suggesting **LESS daily increase** in the infection prevalence

$$\frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I \text{ reaches maximum}$$



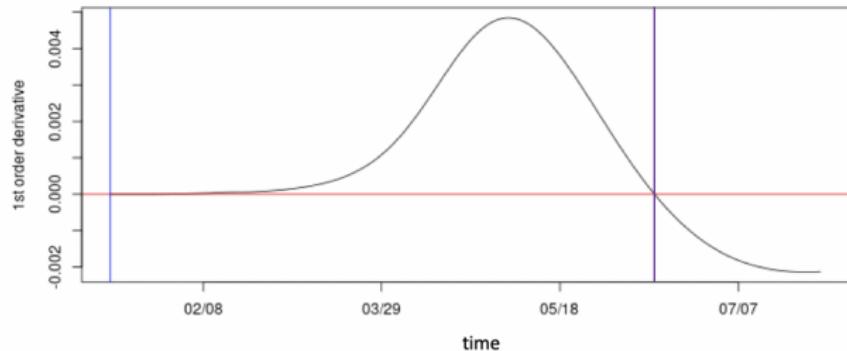
- The first turning point corresponds to the time when the first order derivation of infection prevalence reaches the maximum.

Second Turning Point of Interest: Zero Speed of Infection Prevalence



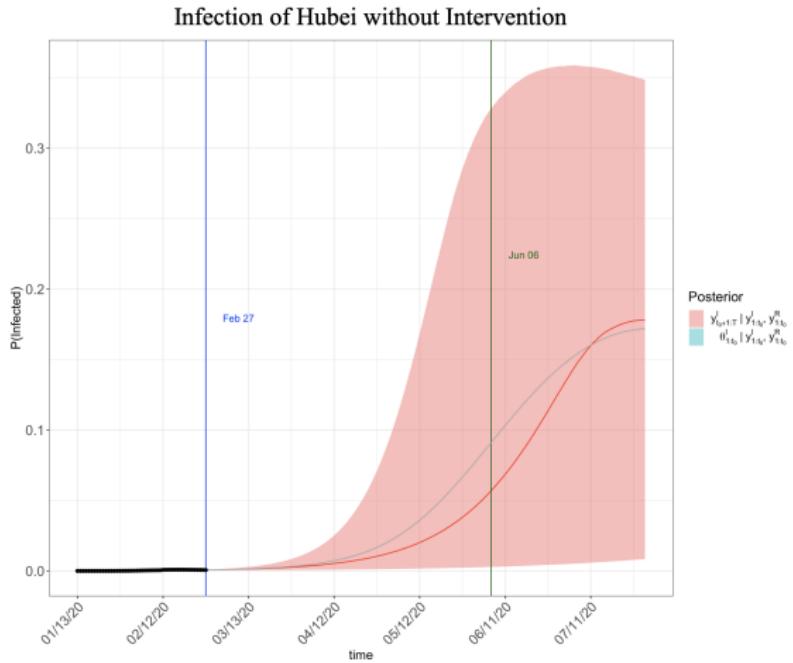
- Second turning point: the time from increase in infection prevalence to decrease, so we would see a STOP and turning back/down

$$\frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I = 0$$



- The second turning point corresponds to the time when the first order derivation of infection prevalence equals to zero.

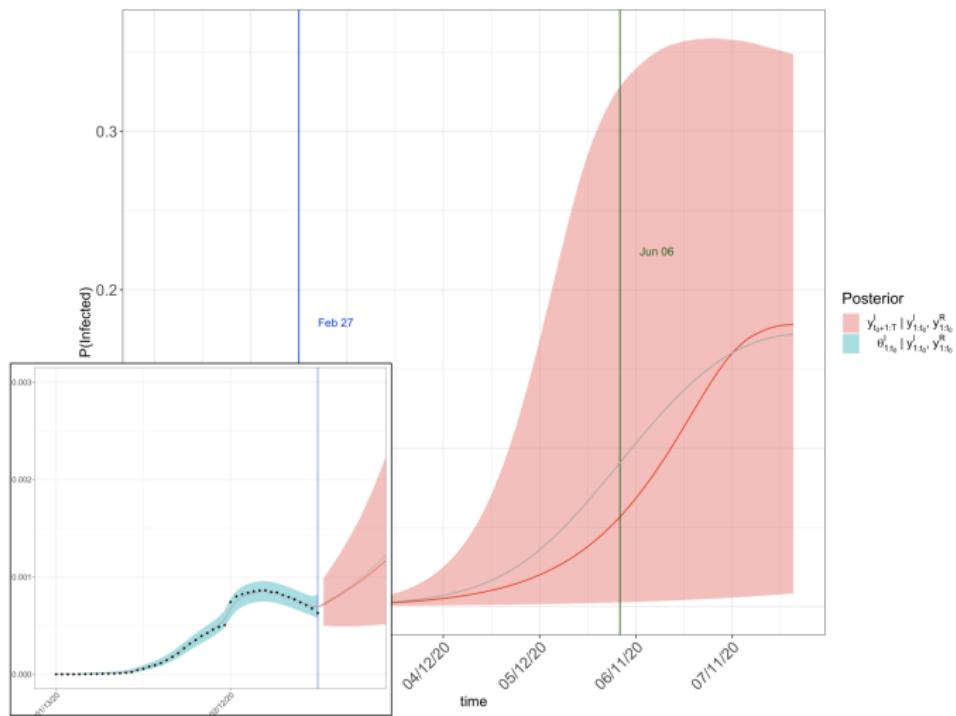
Forecast for Hubei WITHOUT Interventions



Should no interventions be imposed on the infection dynamics, the SIR indicated that there were no hope to see the end of this epidemic soon! The second turning point exceeds the plotting range, i.e. after July 31.

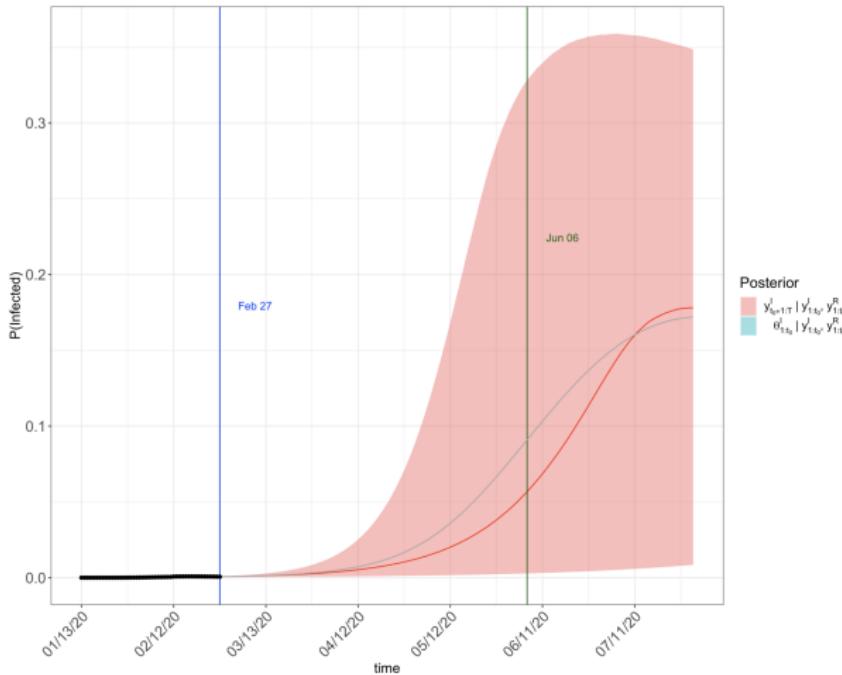
Quality of Learning Hubei Data

Infection of Hubei without Intervention



Estimates from Hubei Data

Infection of Hubei without Intervention



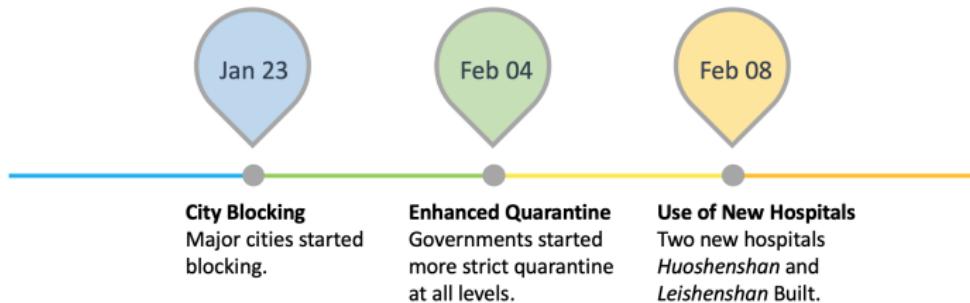
- $\hat{R}_0 = 2.61, \hat{\gamma} = 0.0306, \hat{\beta} = 0.0779$
- On average, with no interventions **about 25% Hubei population** would be infected by July 31.

Limitations of the SIR Model



- There are many forms of interventions used to alter the transmission rate β :
 - ① Personal protective measures like wearing masks, safety glasses, hygiene, etc.
 - ② Quarantine, hospitalization, city blockade, traffic control, limited activities, etc.
- Moreover, the virus can **mutate to evolve**
- In reality, the transmission rate β changes by time.

Various Isolation Measures Taken in Hubei Province



Incorporating Quarantine in the Model

Suppose at a given time, $q^S(t)$ and $q^I(t)$ are the chance of a person being either in-home quarantine or in-hospital isolation, respectively. The modified chance of being infected is

$$\beta\{1 - q^S(t)\}\theta_t^S\{1 - q^I(t)\}\theta_t^I := \beta\pi(t)\theta_t^S\theta_t^I$$

with $\pi(t) := \{1 - q^S(t)\}\{1 - q^I(t)\}$.

Thus, $\pi(t)$ modifies the chance of a susceptible person meeting with an infected person or *vice versa*, which is termed as a *transmission modifier* due to quarantine.

SIR Model with Transmission Rate Modifier



$$\frac{d\theta_t^S}{dt} = -\beta \pi(t) \theta_t^S \theta_t^I, \quad \frac{d\theta_t^I}{dt} = \beta \pi(t) \theta_t^S \theta_t^I - \gamma \theta_t^I, \quad \frac{d\theta_t^R}{dt} = \gamma \theta_t^I$$

The above model assumes primarily that the proportions of susceptible and infectious populations are not shrunken but the chance of a susceptible person meeting with an infected person is reduced by $\pi(t)$.

Two Types of Time-varying Transmission Rate Modifier

- $\pi(t) \in [0, 1]$
- Step function reflecting the government-initiated **macro** isolation measures. For example,

$$\pi(t) = \begin{cases} \pi_{01}, & \text{if } t \leq \text{Jan 23, no concrete quarantine protocols} \\ \pi_{02}, & \text{if } t \in (\text{Jan 23, Feb 4}], \text{ city blockade} \\ \pi_{03}, & \text{if } t \in (\text{Feb 4, Feb 8}], \text{ enhanced quarantine} \\ \pi_{04}, & \text{if } t > \text{Feb 8, opening of new hospitals} \end{cases}$$

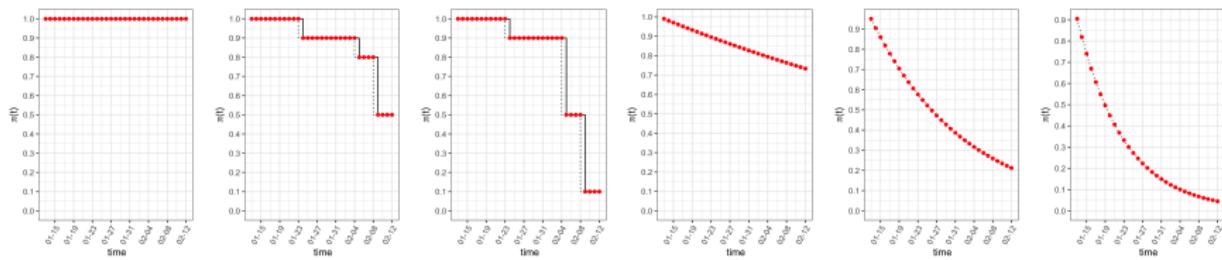
- Continuous function reflecting the gradually increased community-level awareness and responsibility of quarantine, or **micro** isolation measures. For example,

$$\pi(t) = \exp(-\lambda_0 t) \text{ or } \pi(t) = \exp\{-(\lambda_0 t)^\nu\}$$

Examples of Time-varying Transmission Rate Modifier

We choose $\exp(-\lambda_0 t)$ with $\lambda_0 = 0.01, 0.05, 0.1$ and step functions with:

- ① $\pi_{01} = 1, \pi_{02} = 1, \pi_{03} = 1, \pi_{04} = 1$
- ② $\pi_{01} = 1, \pi_{02} = 0.9, \pi_{03} = 0.8, \pi_{04} = 0.5$
- ③ $\pi_{01} = 1, \pi_{02} = 0.9, \pi_{03} = 0.5, \pi_{04} = 0.1$



Note, the step function with $(\pi_{01} = 1, \pi_{02} = 1, \pi_{03} = 1, \pi_{04} = 1)$ is equivalent to the continuous exponential one with $\lambda_0 = 0$

Constant $\pi(t)$: No Interventions in Hubei

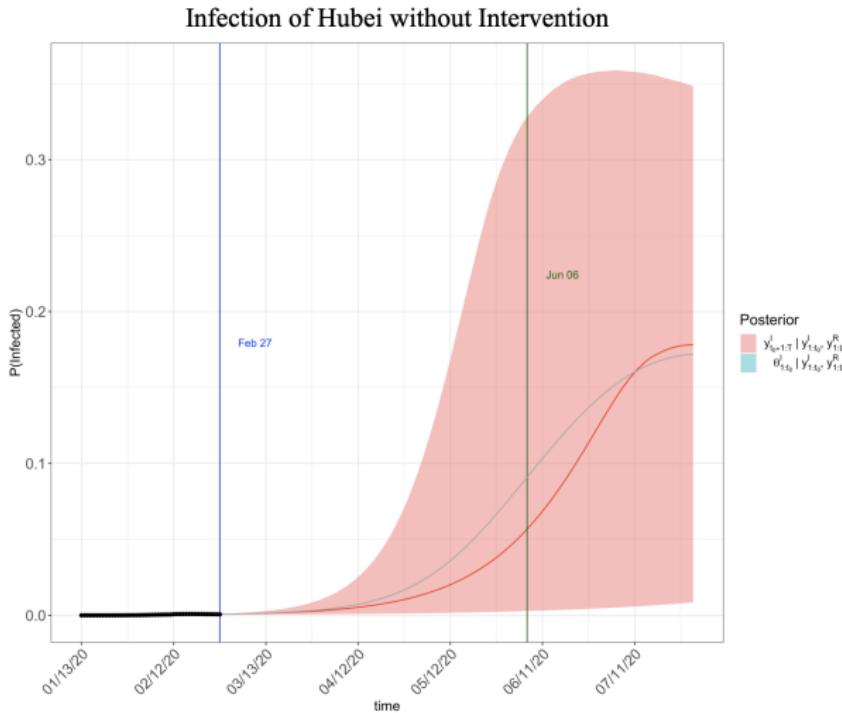


Figure: $\pi(t) \equiv 1$, 10.4 million population would be infected by July 31

Continuous $\pi(t)$: Micro Intervention in Hubei

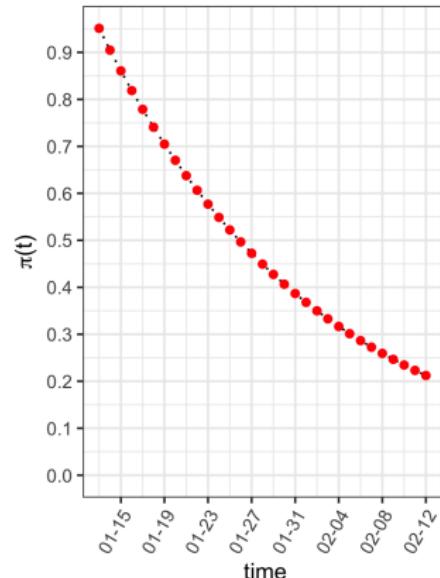
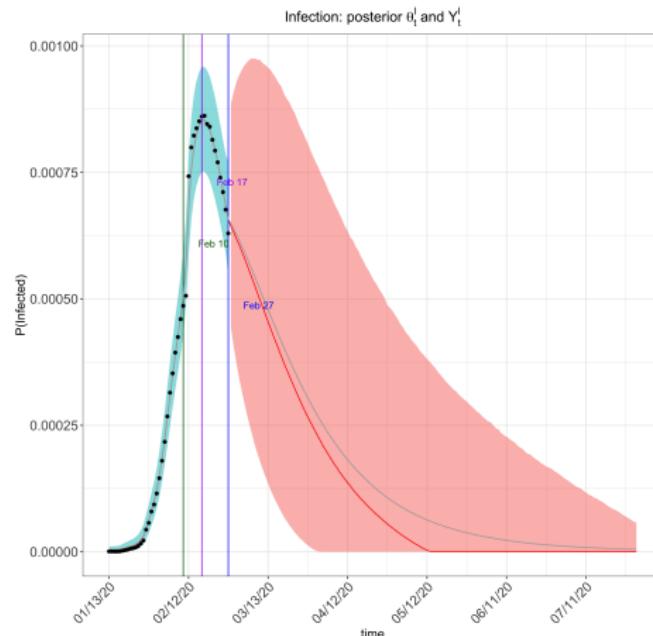


Figure: $\pi(t) = \exp(-0.05t)$; 49,921 infected cases on Feb 17

$$R_0 = 6.275, \gamma = 0.040, \beta = 0.253$$

Step Function $\pi(t)$: Macro Intervention in Hubei

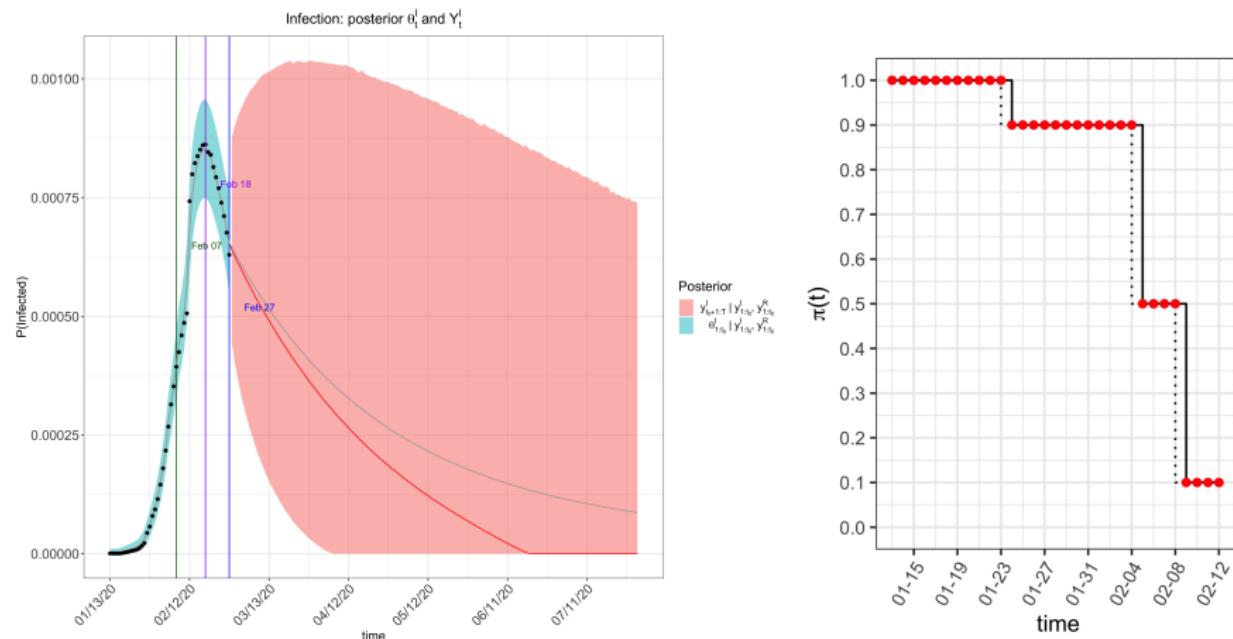
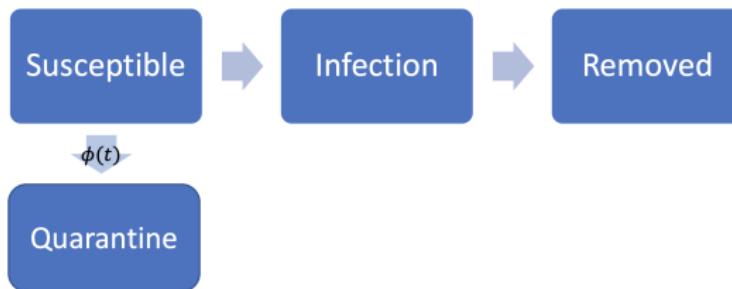


Figure: $\pi_{01} = 1, \pi_{02} = 0.9, \pi_{03} = 0.5, \pi_{04} = 0.1; 49,811$ infected cases on Feb 18

$$R_0 = 5.62, \gamma = 0.0371, \beta = 0.208$$

SIR with Time-varying Quarantine Compartment

To incorporate government-enforced stringent in-home isolation policy, we expanded the SIR model by adding a **quarantine compartment** with a time-varying rate of quarantine ϕ_t , the chance of a susceptible person being willing to take in-home isolation at time t . The extended SIR is



$$\begin{aligned}\frac{d\theta_t^Q}{dt} &= \phi(t)\theta_t^S, \quad \frac{d\theta_t^S}{dt} = -\beta\theta_t^S\theta_t^I - \phi(t)\theta_t^S, \\ \frac{d\theta_t^I}{dt} &= \beta\theta_t^S\theta_t^I - \gamma\theta_t^I, \quad \frac{d\theta_t^R}{dt} = \gamma\theta_t^I.\end{aligned}$$

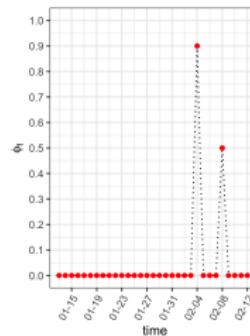
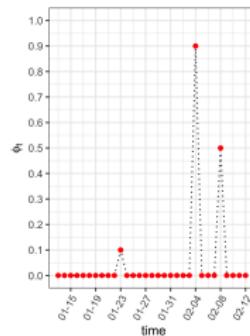
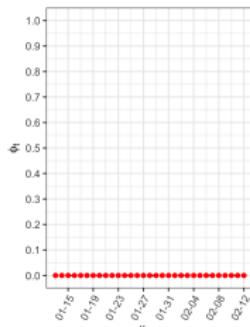
Examples of Time-varying Quarantine Rate: $\phi(t)$

A pre-specified Dirac delta function:

$$\phi(t) = \begin{cases} \phi_{01}, & \text{if } t = \text{Jan 23, city blockade} \\ \phi_{02}, & \text{if } t = \text{Feb 4, enhanced quarantine} \\ \phi_{03}, & \text{if } t = \text{Feb 8, opening of new hospitals} \\ 0, & \text{otherwise} \end{cases}$$

Some examples of quarantine rates we considered:

- $\phi_{01} = 0, \phi_{02} = 0, \phi_{03} = 0$
- $\phi_{01} = 0.1, \phi_{02} = 0.9, \phi_{03} = 0.5$
- $\phi_{01} = 0, \phi_{02} = 0.9, \phi_{03} = 0.50$



$\phi(t)$: Quarantine in Hubei

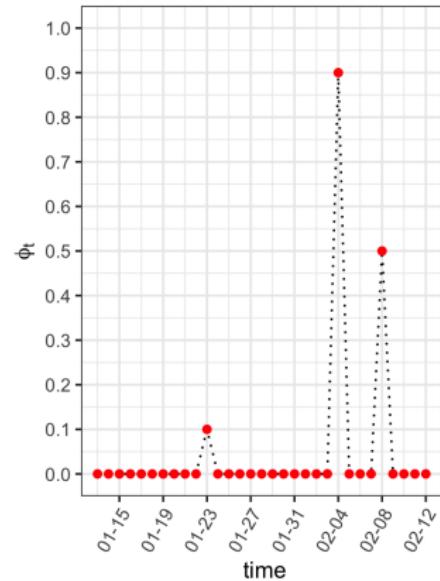
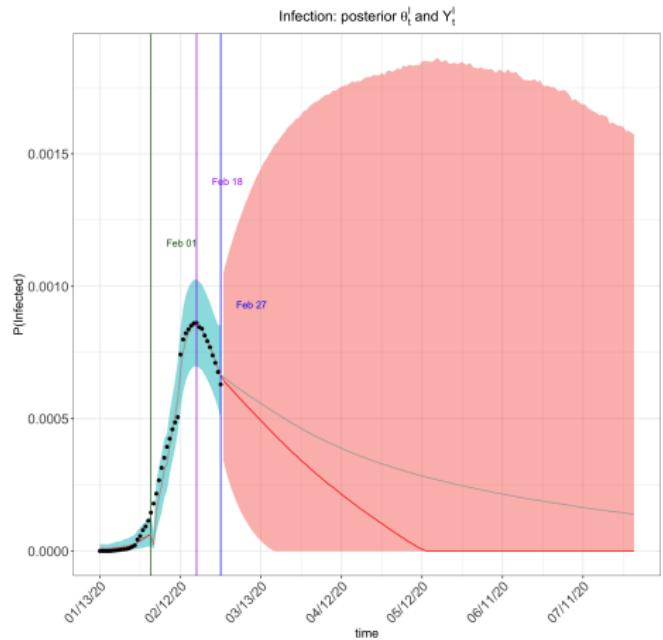


Figure: $\phi_{01} = 0.1, \phi_{02} = 0.9, \phi_{03} = 0.5$; 49,996 infected on Feb 18

Under-reporting Calibration

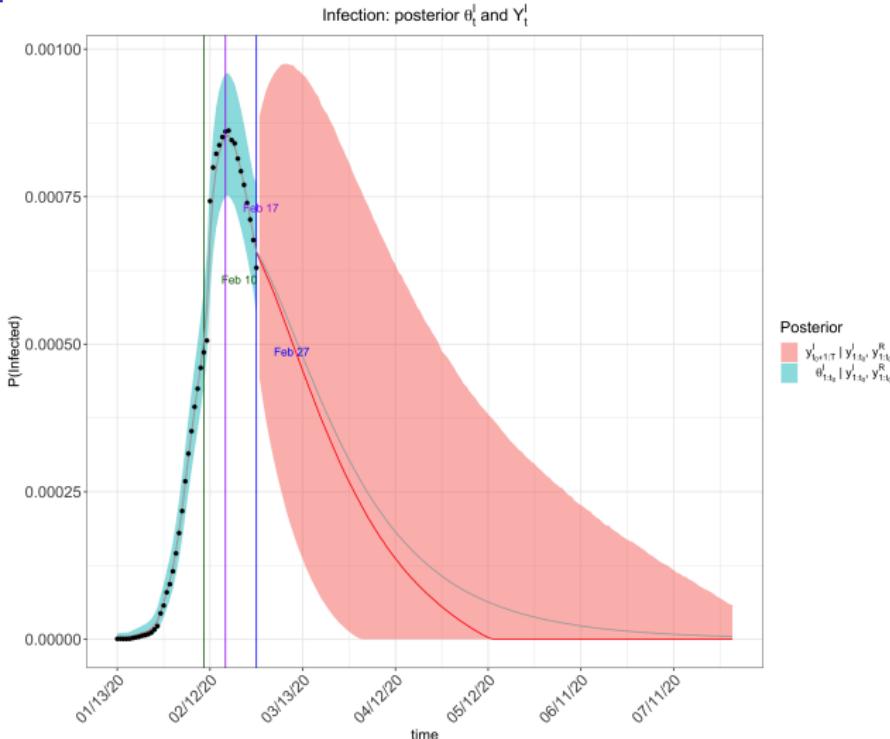


Figure: $\pi(t) = \exp(-0.05t)$

Under-reporting Calibration

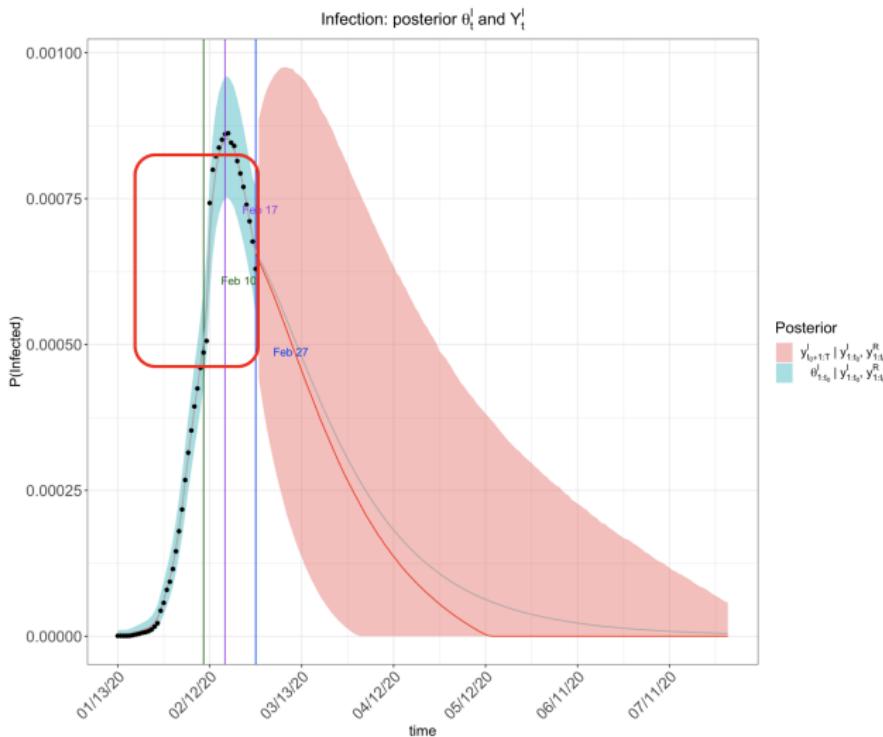


Figure: $\pi(t) = \exp(-0.05t)$

Under-reporting Calibration

The issue of under-reporting (till Feb 12) of the infected cases due to

- low sensitivity of the diagnosis test;
- lack of the test kit.

Calibration Procedure:

- Assume an exponential growth curve for the cumulative number of infected cases on Feb 12 or before as

$$y(t) = ae^{\lambda t} + b.$$

- Minimize the one-step ahead extrapolation error on Feb 13 ($t = 32$) by the means of Lagrange Multipliers with the constraints of suitable boundary conditions:

$$\begin{aligned} \min_{a, \lambda} \quad & \{y(32) - ae^{32\lambda} + a\}^2 \\ \text{s.t.} \quad & ae^{31\lambda} - a = y(31). \end{aligned}$$

Under-reporting Calibration

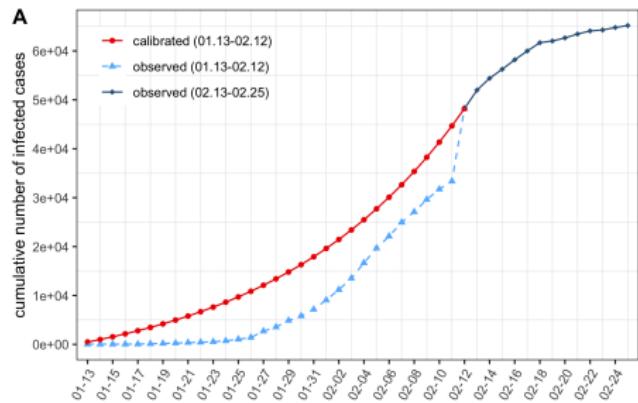
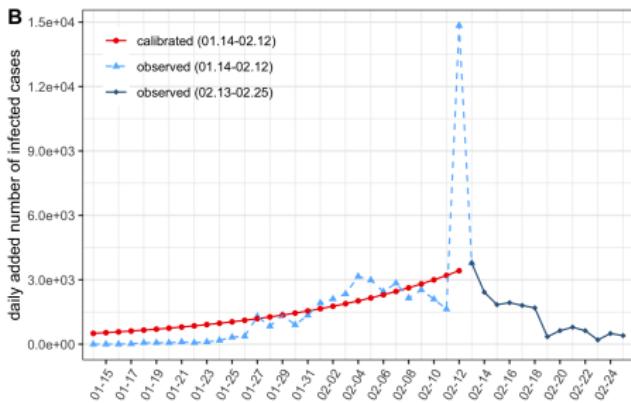
A**B**

Figure: Under-reporting calibration of the infected cases by the red curve.

Calibrated Data: Hubei with Transmission Rate Modifier $\pi(t)$

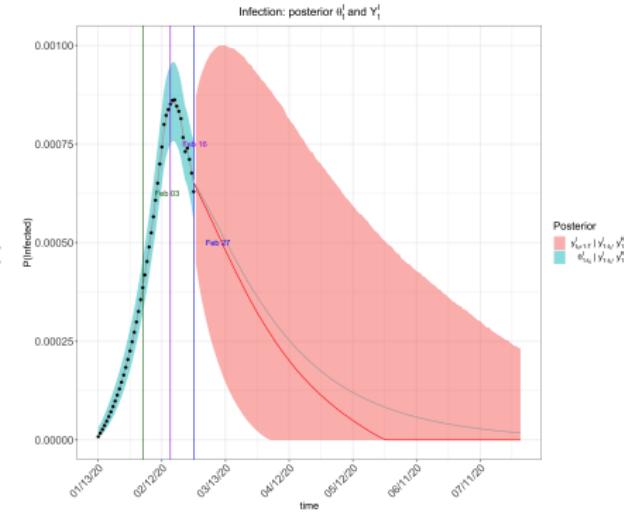
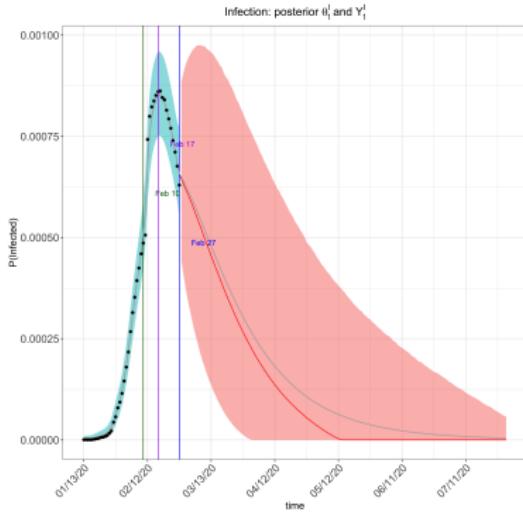


Figure: $\pi(t) = \exp(-0.05t)$

- without calibration (left): $R_0 = 6.275, \gamma = 0.040, \beta = 0.253$
- with calibration (right): $R_0 = 4.918, \gamma = 0.027, \beta = 0.134$

Calibrated Data: Hubei with Quarantine Rate $\phi(t)$

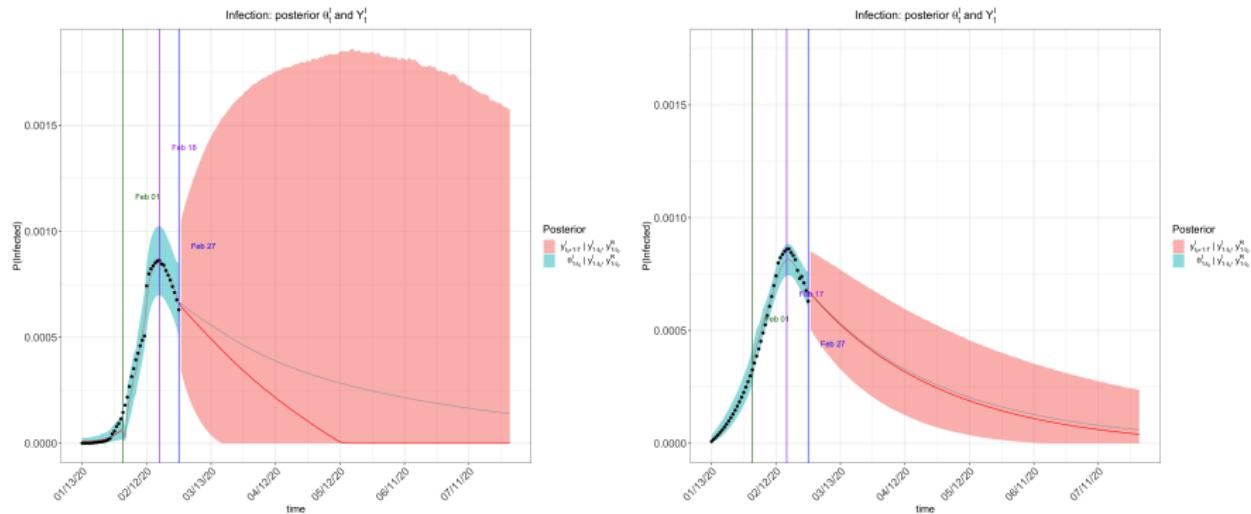


Figure: $\phi_{01} = 0.1, \phi_{02} = 0.9, \phi_{03} = 0.5$

- without calibration (left): $R_0 = 1.373, \gamma = 2.955, \beta = 4.059$
- with calibration (right): $R_0 = 4.756, \gamma = 0.021, \beta = 0.106$

A Simple Validation of SIR: With No Data Calibration

Use data up to 2/23 to train the model, and plot 5 data up to 2/28

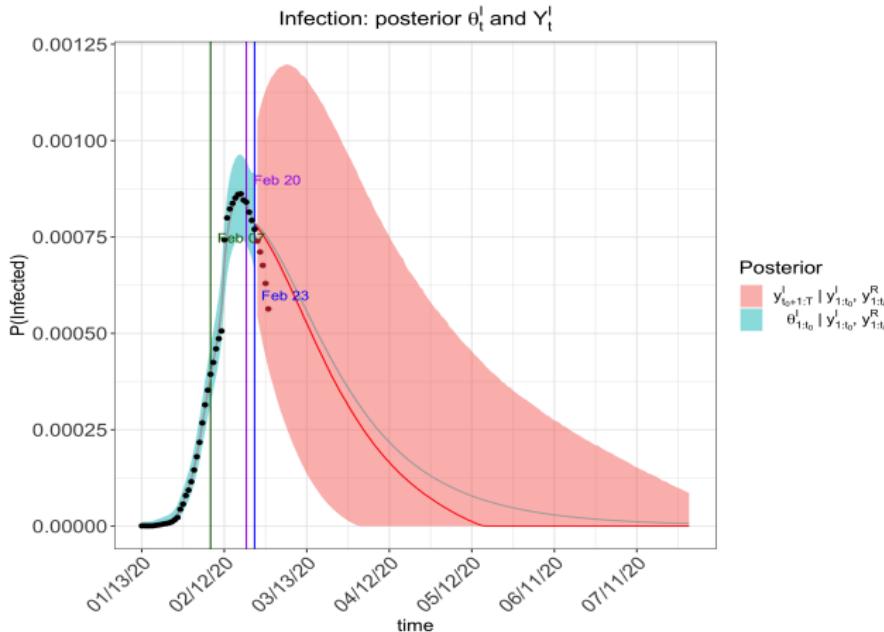


Figure: $\pi(t) = \exp(-0.05t)$. The credible bands contain all 5 data points. The SIR captured the underlying infection dynamics.

A Simple Validation: With Data Calibration

Use data up to 2/23, and plot 5 data points till 2/28

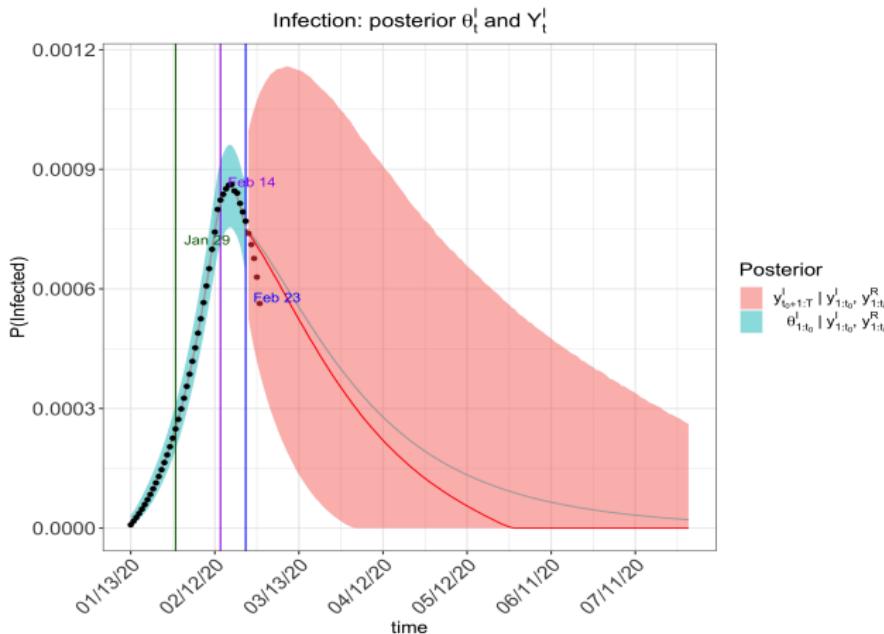


Figure: $\pi(t) = \exp(-0.05t)$. The credible bands contain all 5 data points. The SIR captured the underlying infection dynamics.

An R package: eSIR

<https://github.com/lilywang1988/eSIR>

R package eSIR: extended state-space SIR epidemiological models

Song Lab 2020-02-29

Purpose

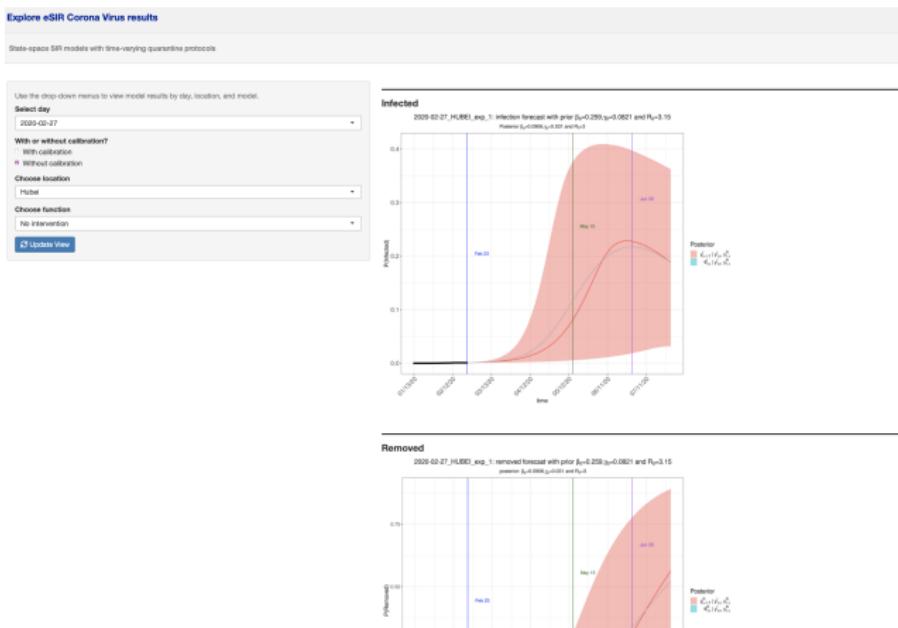
The outbreak of novel Corona Virus disease (a.k.a. COVID-19), originated in Wuhan, the capital of Hubei Province spreads quickly and affects many cities in China as well as many countries in the world. The Chinese government has enforced very stringent quarantine and inspection to prevent the worsening spread of COVID-19. Although various forms of forecast on the turning points of this epidemic within and outside Hubei Province have been published in the media, none of the prediction models has explicitly accounted for the time-varying quarantine protocols. We extended the classical SIR model for infectious disease by incorporating forms of medical isolation (in-home quarantine and hospitalization) in the underlying infectious disease dynamic system. Using the state-space model for both daily infected and hospitalized incidences and MCMC algorithms, we assess the effectiveness of quarantine protocols for confining COVID-19 spread in both Hubei Province and the other regions of China. Both predicted turning points and their credible bands may be obtained from the extended SIR under a given quarantine protocol. R software packages are also made publicly available for interested users.

The standard SIR model has three components: susceptible, infected, and removed (including the recovery and dead). In the following sections, we will introduce the other extended state-space SIR models and their implementation in the package. All the results provided are based on very short chains. Please set at least `M=5e5` and `nburnin=2e5` to obtain stable MCMC chains via `rjags`.



An R Shiny App is ready for use

<https://umich-biostatistics.shinyapps.io/eSIR/>



A Manuscript Is available on medRxiv



HOME | ABOUT | SUBMIT | ALERTS / RSS

Search



Advanced Search

[Comment on this paper](#)

An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China

Peter X Song, Lili Wang, Yiwang Zhou, Jie He, Bin Zhu, Fei Wang, Lu Tang, Marisa Eisenberg

doi: <https://doi.org/10.1101/2020.02.29.20029421>

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

Abstract

Info/History

Extension: Add Hospitalization Compartment

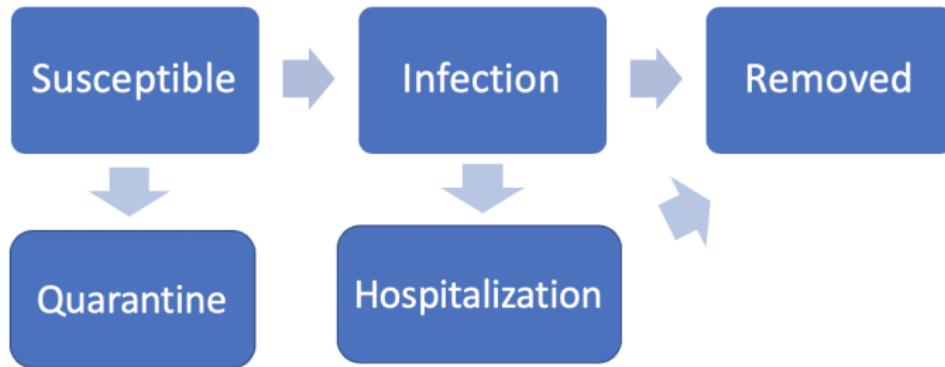


Figure: Include both quarantine and hospitalization compartments

Add Hospitalization Compartment

$$\begin{aligned}\frac{d\theta_t^Q}{dt} &= \phi_t \theta_t^S, \\ \frac{d\theta_t^S}{dt} &= -\beta \theta_t^S \theta_t^I - \phi_t \theta_t^S \\ \frac{d\theta_t^I}{dt} &= \beta \theta_t^S \theta_t^I - \gamma^R \theta_t^I - \gamma^H \theta_t^I, \\ \frac{d\theta_t^R}{dt} &= \gamma^R \theta_t^I + \gamma^{RH} \theta_t^I, \\ \frac{d\theta_t^H}{dt} &= \gamma^H \theta_t^I - \gamma^{RH} \theta_t^I\end{aligned}$$

Add Suspected Compartment

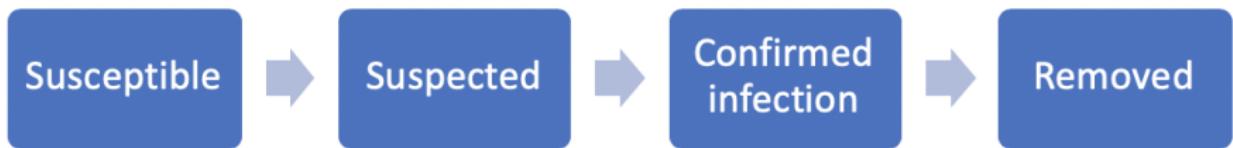


Figure: Include the suspected compartment

Add Suspected Compartment

$$\frac{d\theta^S}{dt} = -\beta\theta^S\theta^U$$

$$\frac{d\theta^U}{dt} = \beta\theta^S\theta^U - \gamma^C\theta^U,$$

$$\frac{d\theta^C}{dt} = \gamma^C\theta^U - \gamma^R\theta_t^C,$$

$$\frac{d\theta^H}{dt} = \gamma^R\theta^C$$

Comprehensive Dynamic Model

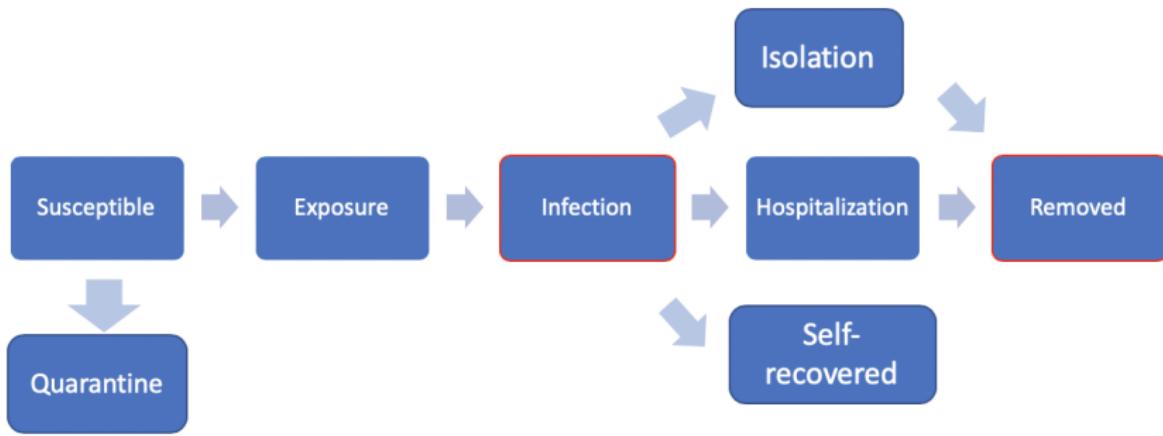


Figure: Include multiple compartments

Comprehensive Dynamic model

$$\begin{aligned}\frac{d\theta^Q}{dt} &= \phi_t \theta^S, \quad \frac{d\theta^S}{dt} = -\beta \theta^S \theta^I - \phi_t \theta^S, \\ \frac{d\theta^E}{dt} &= \beta \theta^S \theta^I - \alpha \theta^E, \\ \frac{d\theta^I}{dt} &= \alpha \theta^E - (1 - \rho_t - \pi_t) \gamma^H \theta^I - \rho_t \gamma^S \theta^I - \pi_t \gamma^C \theta^I, \\ \frac{d\theta^H}{dt} &= (1 - \rho_t - \pi_t) \gamma^H \theta^I - \eta^H \theta^H, \\ \frac{d\theta^S}{dt} &= \rho_t \gamma^S \theta^I - \eta^S \theta^S, \\ \frac{d\theta^C}{dt} &= \pi_t \gamma^C \theta^I, \\ \frac{d\theta^R}{dt} &= \eta^H \theta^H + \eta^S \theta^S,\end{aligned}$$

Thank you!

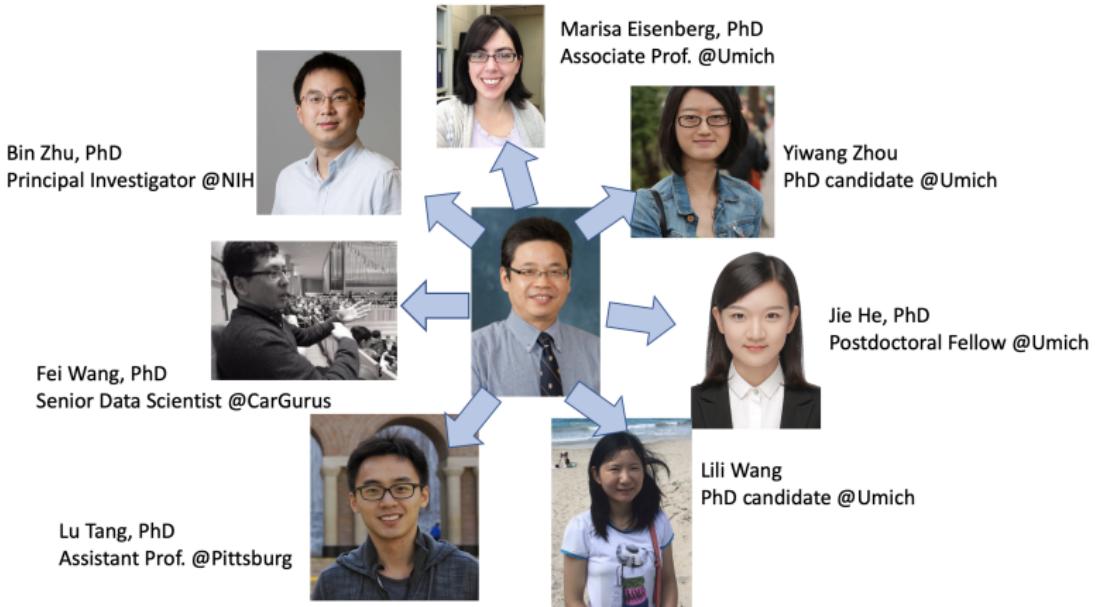


Figure: Song Lab: $R_0 = 7$