Epidemiological Modeling of the COVID-19 Pandemic: An Overview

Peter Song & Song Lab

Department of Biostatistics, School of Public Health, University of Michigan

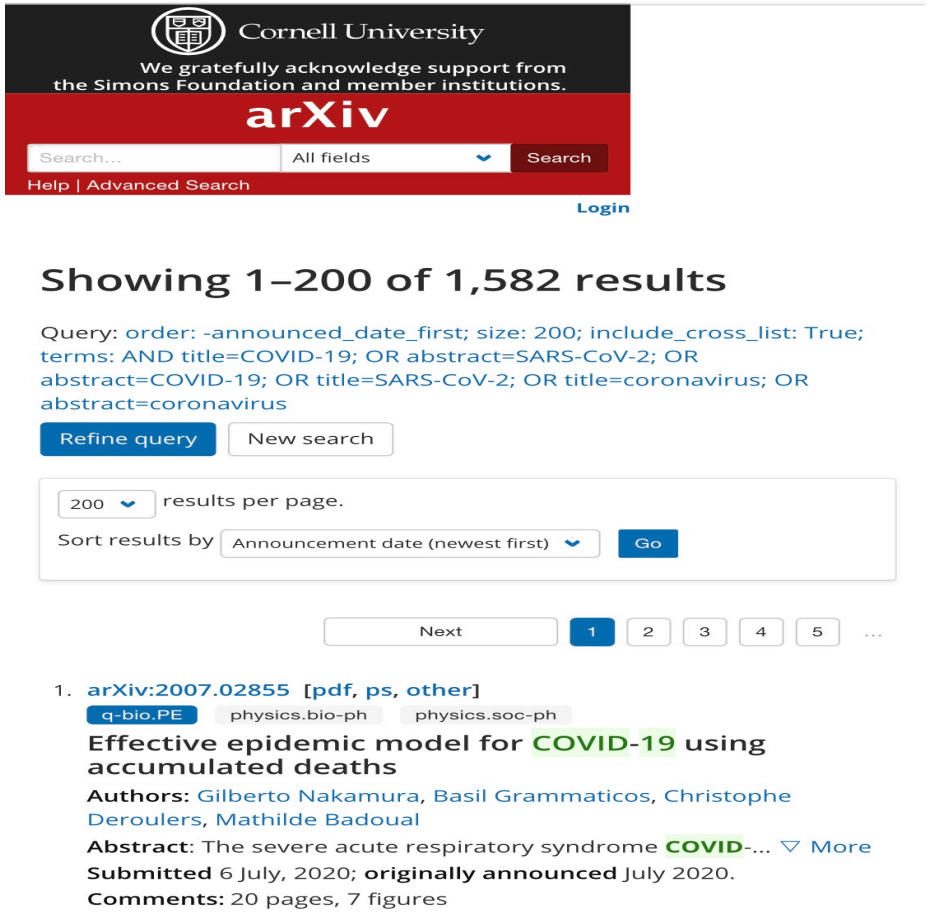
This research is partially supported by NSF DMS1811734 and NIH R01ES024732

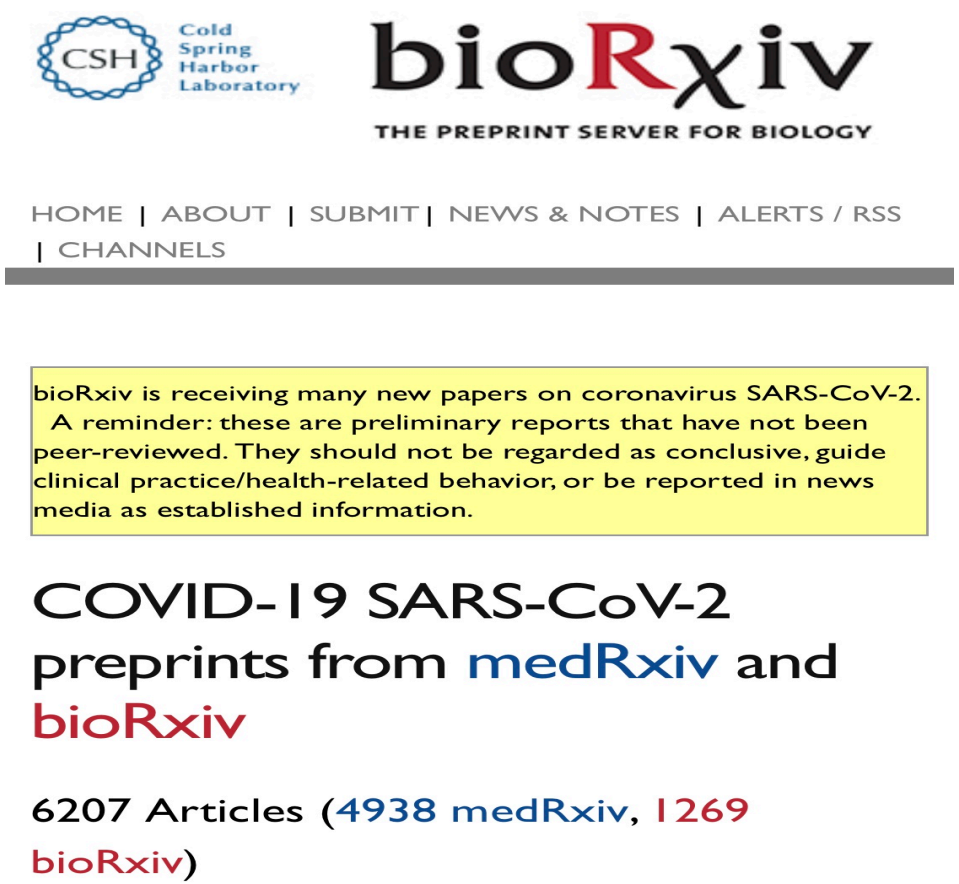
Acknowledgements

- Taikang Luojia (泰康珞珈）Institute
- China Data Institute
- Future Data Lab
- International Statistical Review

- Zhou et al. (2020). A Review of Multi-Compartment Models for Infectious Diseases. *International Statistical Review* (to appear) DOI: 10.1111/insr.12402.

A vast literature on modeling of COVID-19



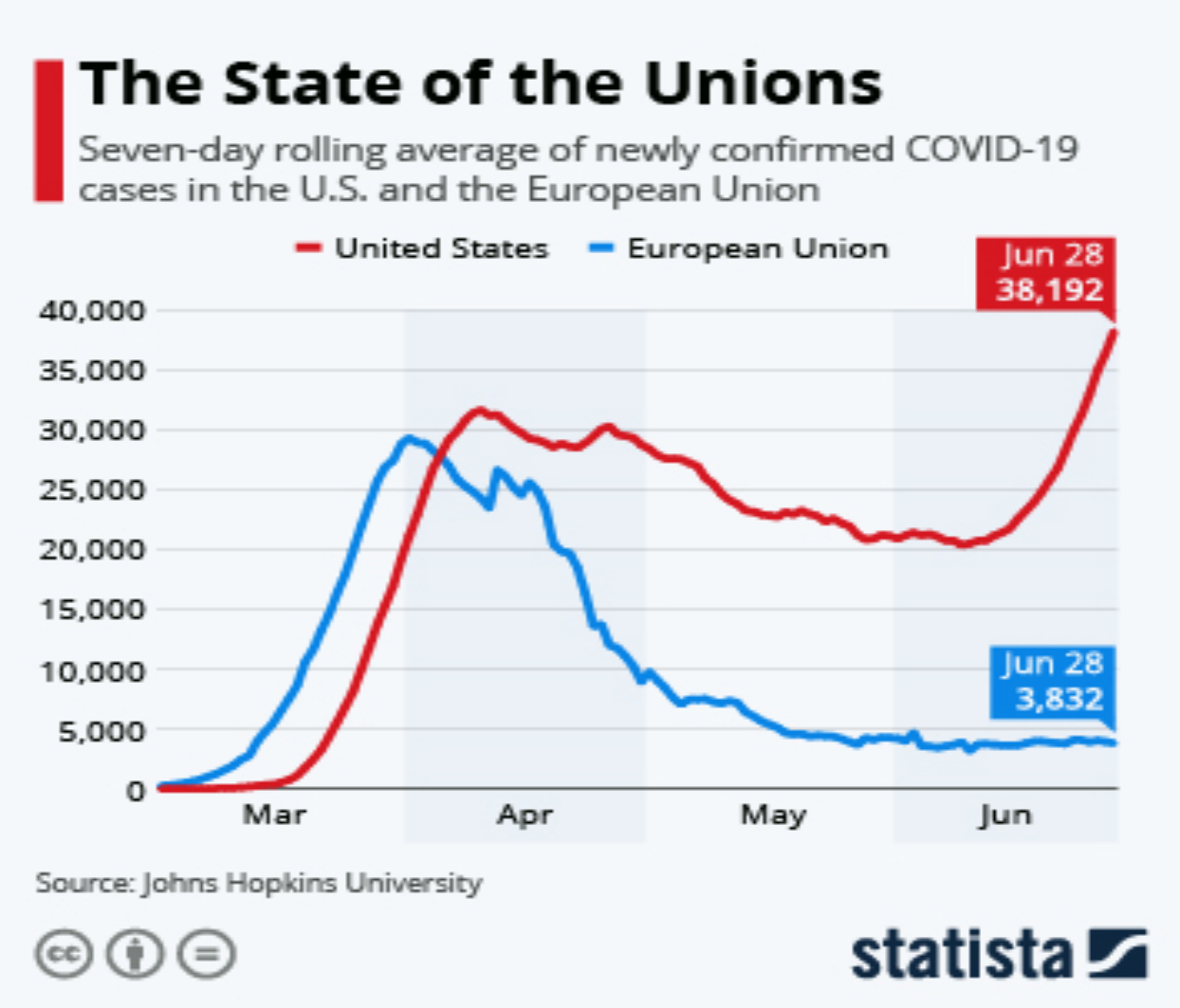


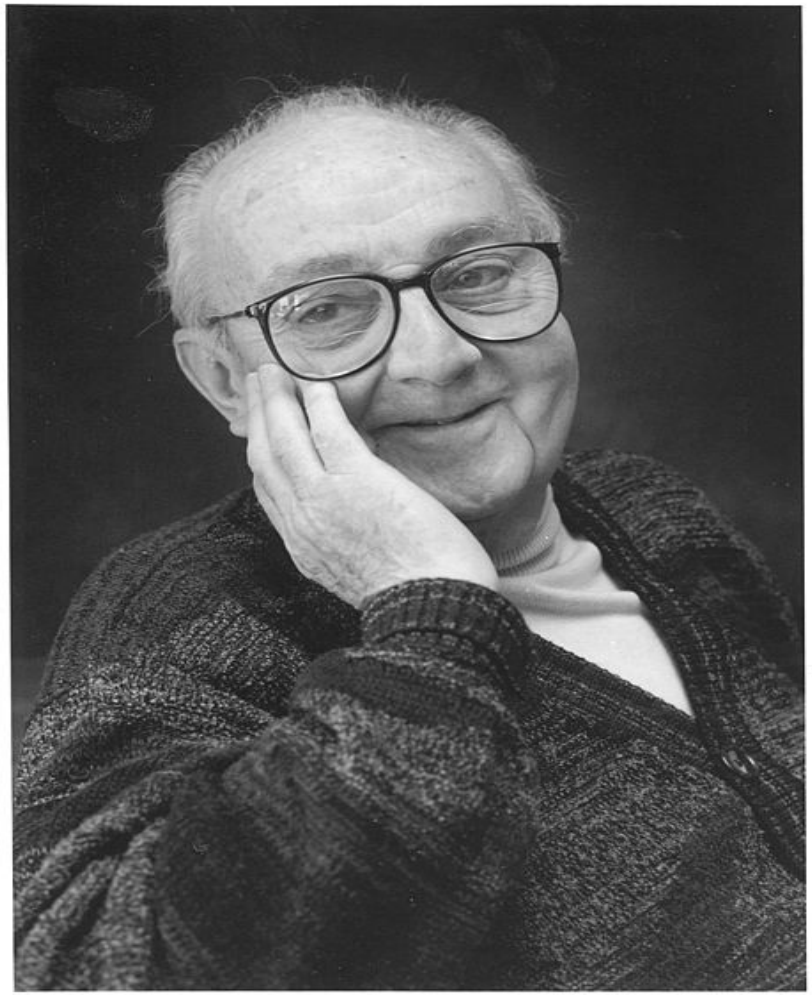
A snapshot of near 8K technical reports in three major archives but no model has emerged to be the one yet.

Many types of models for COVID-19

- **Epidemiological models (disease spread, prediction, contact tracing)**
- Diagnostic models (X-ray, blood test, viral test for symptoms)
- Medical treatment models (clinical trials, drug comparisons)
- Serological survey models (asymptomatic cases)
- Economic models (COVID-19 on economy)
- Policy models (assessing policies for social distancing, reopening, travel suspension)

Box: “All models are wrong but some are useful”





GEP Box: from the Wikipedia

Basics in epidemiological modeling of COVID-19 pandemic

Modeling is a comprehensive analytic and computational task

- Is highly relied on study objectives
- Needs support of good data
- Requires insights and experiences from similar occasions
- Demands power of computing facility and data analytics
- Requires validation and dissemination

Open question 1

- How would researchers assess both timing and strength of the second wave of the COVID-19 pandemic?
- Is the second wave worse than the first one?

- ✓ Answers to these questions need a relatively accurate long-term prediction of the infection dynamics. Among so many different statistical models being able to predict future spreading patterns, we need to identify few ones or their combinations that are particularly useful to make **long-term predictions**.
- ✓ Can you? If so, How?

Open question 2

- How would government monitor the likelihood of a recurring surge of COVID-19 caused by business reopenings?
- Does the social distancing measure help reduce a potentially rising risk?

✓ Answers to these questions require **adequate data** that may not be easily collected by routine public health surveillance approaches. We may work with practitioners to develop good sampling instruments and schemes for community risk surveillance.

Open question 3

- Is face mask protective?
- If so, how to assess the compliance of face mask wearing?

✓ Answers to these questions require the **causal effect** analysis of face mask wearing on disease progression are very challenging. This is because there is no randomization in the intervention allocation and many confounding factors are unobserved.

Open question 4

➢ Is there evidence that the contagion and fatality of the coronavirus decays over time and/or in some regions due to an increasing recovery rate of virus carriers and a decreasing rate of case fatality?

✓ Answers to this question pertain to multi-dimensional **competing risk** models and **hot-spot** analyses. We ought to work out some thoughtful and convincing methods and solutions to the public.

Challenge 1

We don’t have relevant, good quality data!

✓Public Surveillance Data

- ❖Many biases in surveillance data
- ❖Cell phone data/mobility data?
- ❖Viral testing data?
- ❖Serological testing data?
- ❖Symptom data?
- ❖Medical treatment data?
- ❖Age, race, other demographics

Challenge 2

We don’t have good models for infectious diseases!

- Multi-compartment models are so far most useful in prediction

❖No other good models are available to answer difficult questions, rather than merely for summary statistics and data visualization

Challenge 3

We don’t have much time to develop and test new ideas/methods/software!

- We have only 2-4 weeks to develop and test new ideas and methods in order to stay on competitive; otherwise ideas/methods may be out of date.
- We don’t have good preparations in terms of theory, methods, computing, and researchers in this field.
- We need software packages of new methods, which are lacking now.

Challenge 4

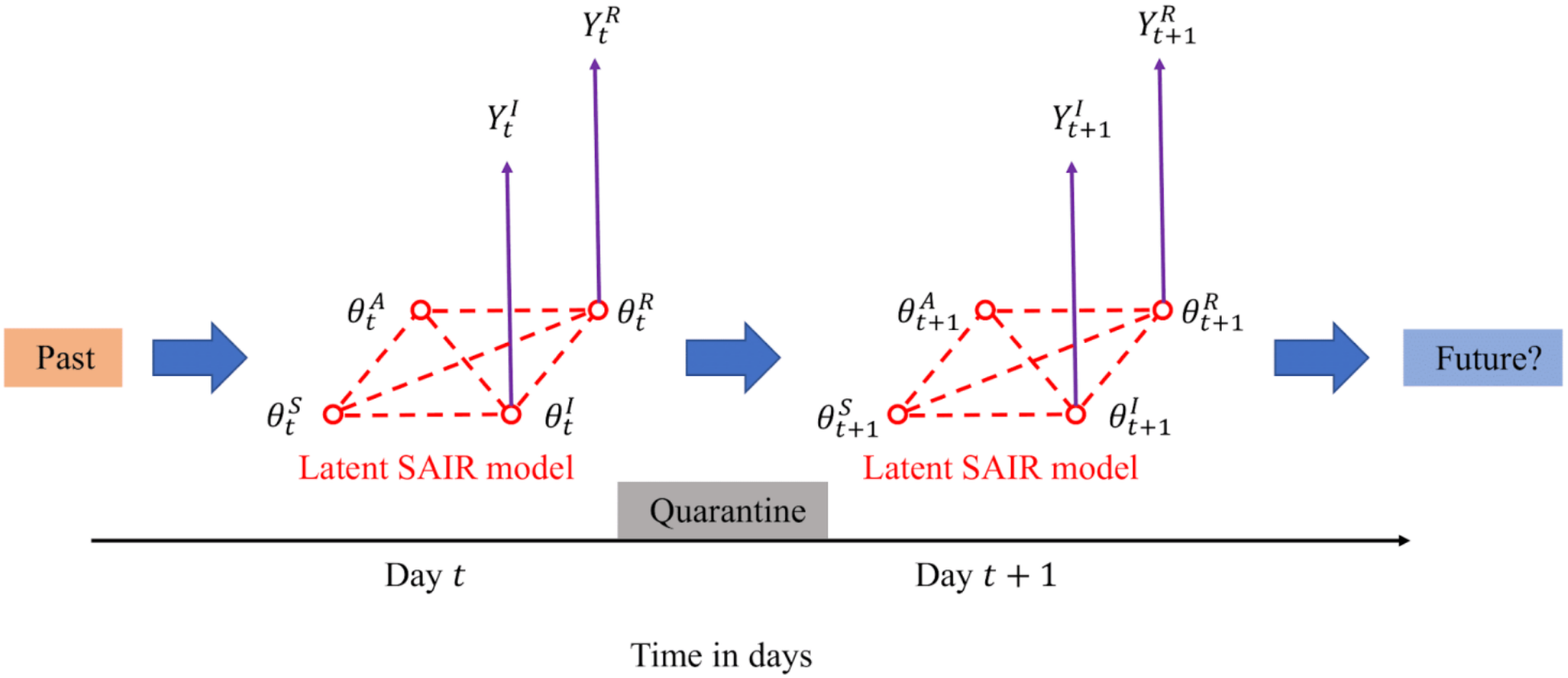
We don’t have adequate resources!

- Funding (NSF/NIH/Foundation)
- Data (e.g. Hospital Data)
- Research teams (Lack of adequate expertise)
- Computing (e.g. **software**)
- Publication

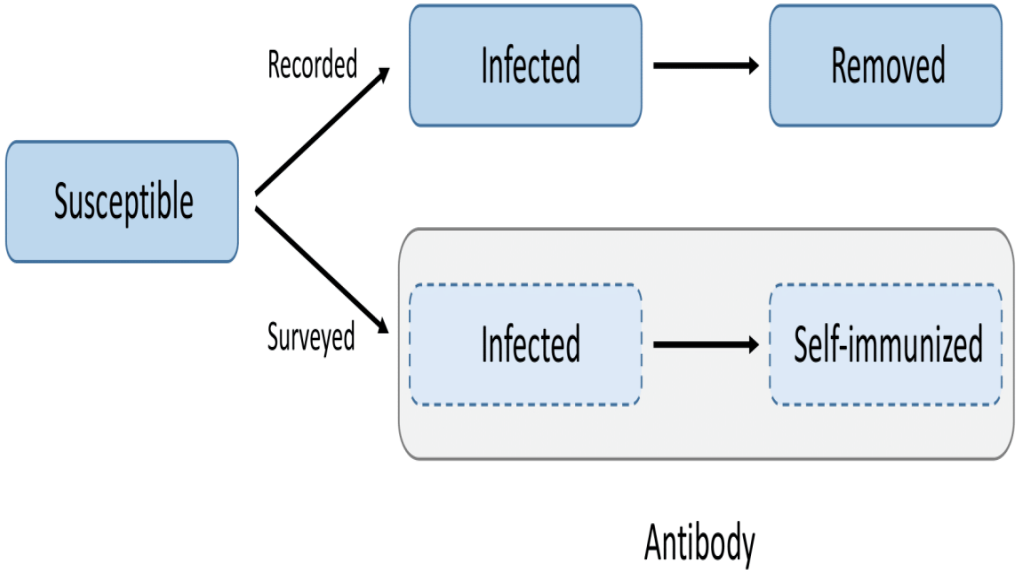
Model for risk prediction

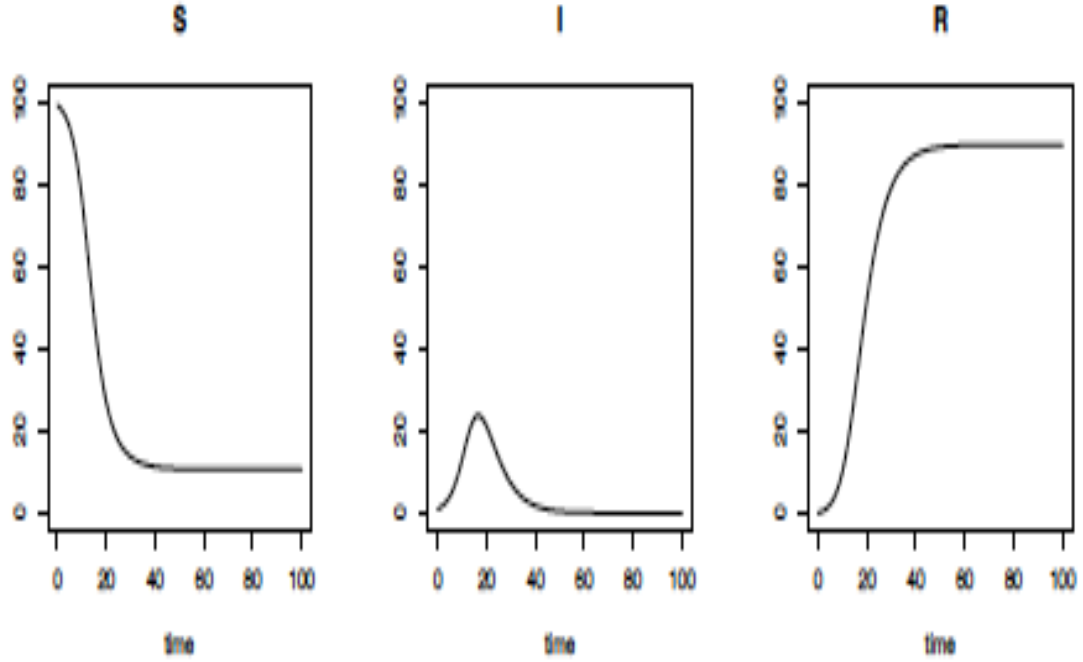
- A mechanistic model for infection dynamics
- A statistical framework of data collection with random errors
- A machine learning method to learn model parameters from data
- A statistical method to quantify prediction uncertainty
- A procedure to validate prediction accuracy

State-space temporal models for risk prediction

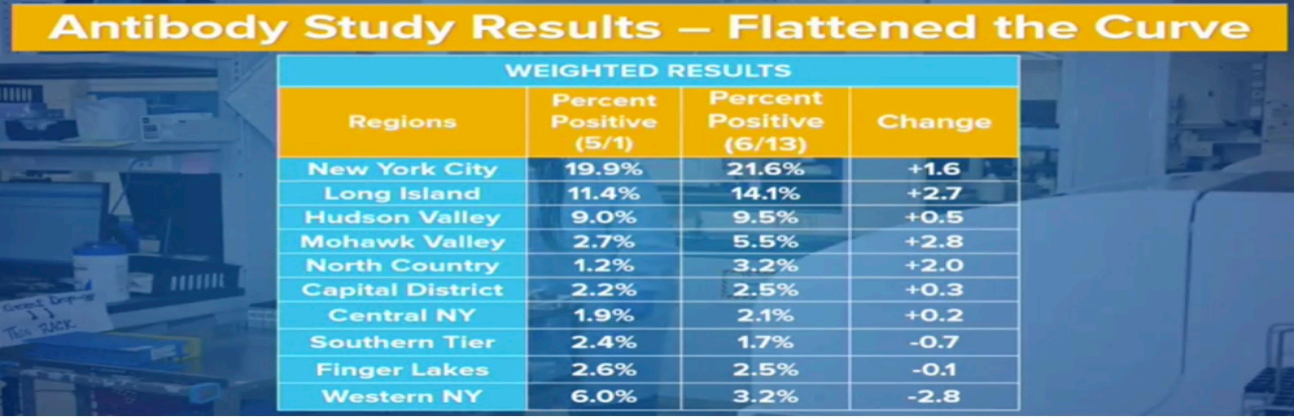


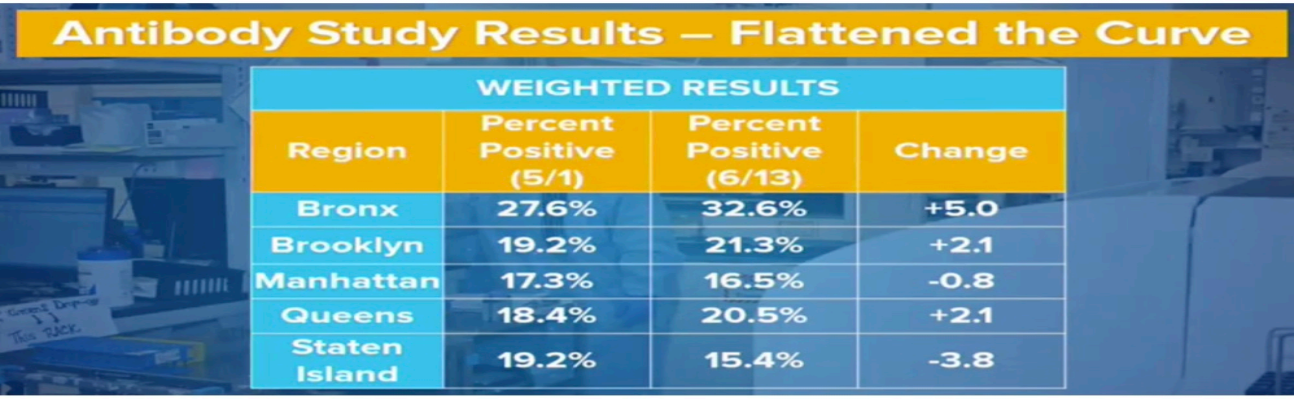
A mechanistic model

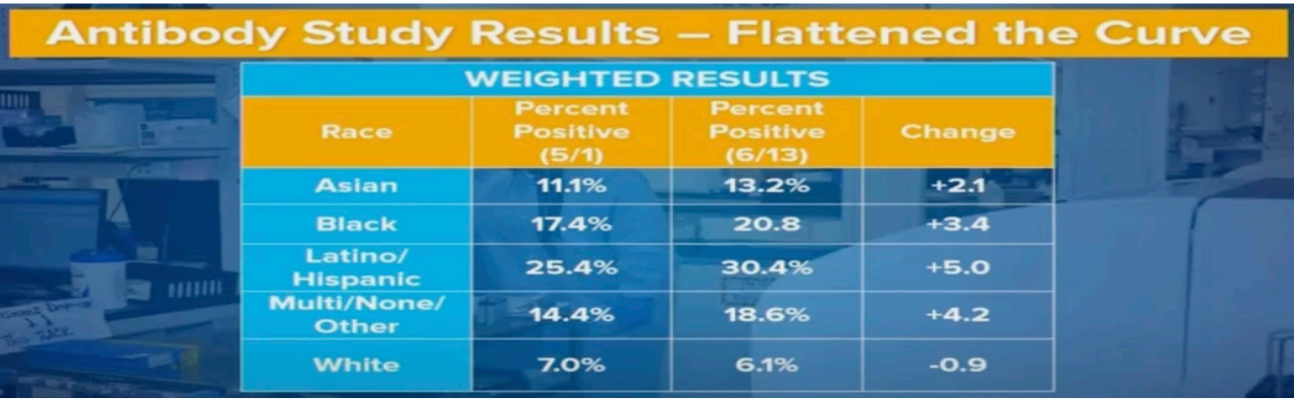




$$\frac{d\theta_t^A}{dt} = \alpha(t)\theta_t^S,\ \frac{d\theta_t^S}{dt} = -\alpha(t)\theta_t^S - \beta\pi(t)\theta_t^S\theta_t^I,\ \frac{d\theta_t^I}{dt} = \beta\pi(t)\theta_t^S\theta_t^I - \gamma\theta_t^I,\ \text{and}\ \frac{d\theta_t^R}{dt} = \gamma\theta_t^I,$$





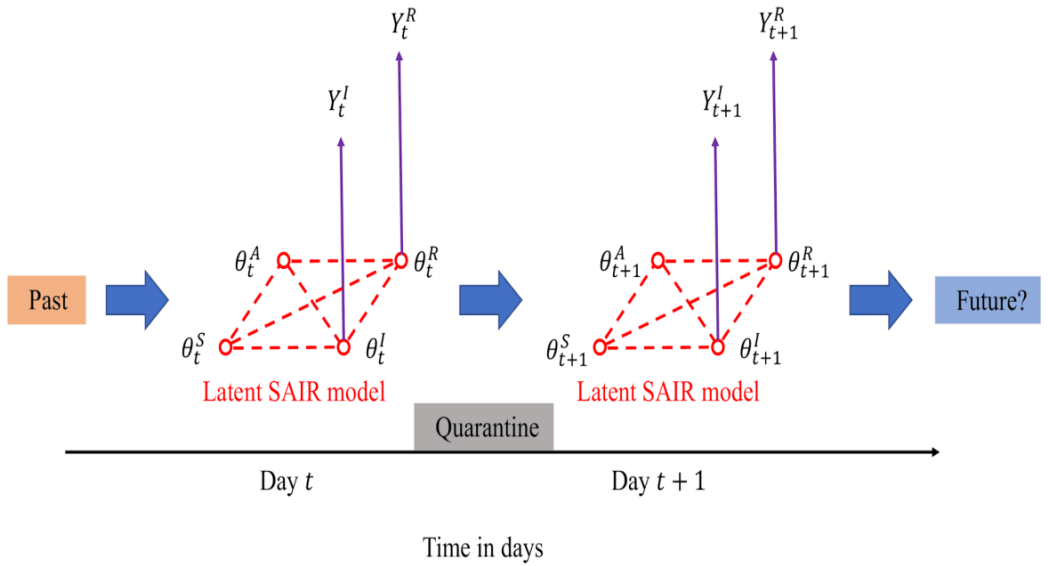


Powerpoint slides from Governor Cuomo's 6/16/20 press briefing

NY Governor Andres Cuomo released results of statewide antibody tests surveys; about 20% of NY residents had antibody tested positive. Thus, this self-immunization rate function is specified with a delta function with a jump of size 0.20 on May 1.

A statistical model with sampling schemes

$$Y_t^I|\theta_t, \lambda^I \sim \text{Beta}(\lambda^I\theta_t^I, \lambda^I(1-\theta_t^I)),$$
$$Y_t^R|\theta_t, \lambda^R \sim \text{Beta}(\lambda^R\theta_t^I, \lambda^R(1-\theta_t^I)).$$



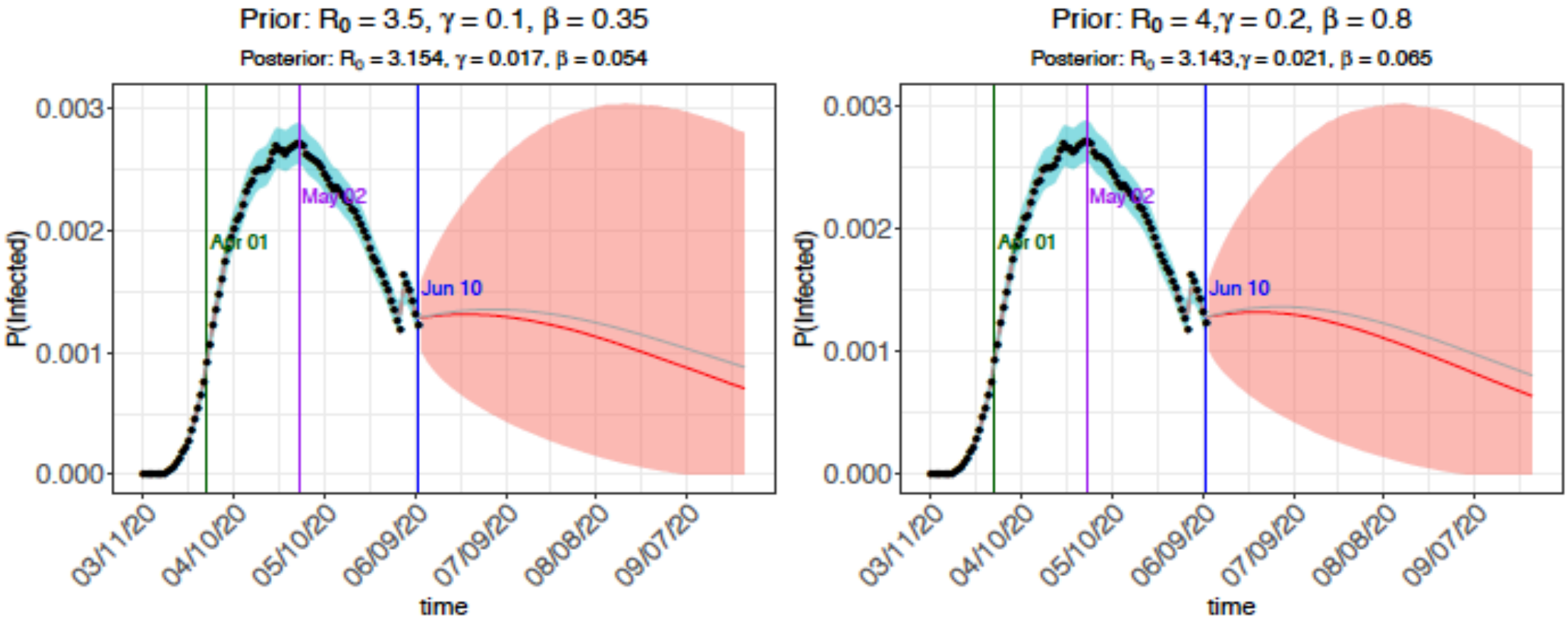
$$\theta_t|\theta_{t-1}, \tau \sim \text{Dirichlet}(\kappa f(\theta_{t-1}, \beta, \gamma))$$

Software

- Refer to our R package eSIR available on the Github: https://github.com/lilywang1988/eSIR/tree/master/install_pkg)

- Approximate the dynamic system of four ODEs by the 4^{th} order Runge-Kutta Method
- Derive posteriors and implement the MCMC using RJAGS in R software to draw MCMC samples from the posteriors
- Summarize the results from 200,000 MCMC draws after burn-in
- Provide initial values needed for county-level prediction

Analysis of Michigan COVID-19 Data

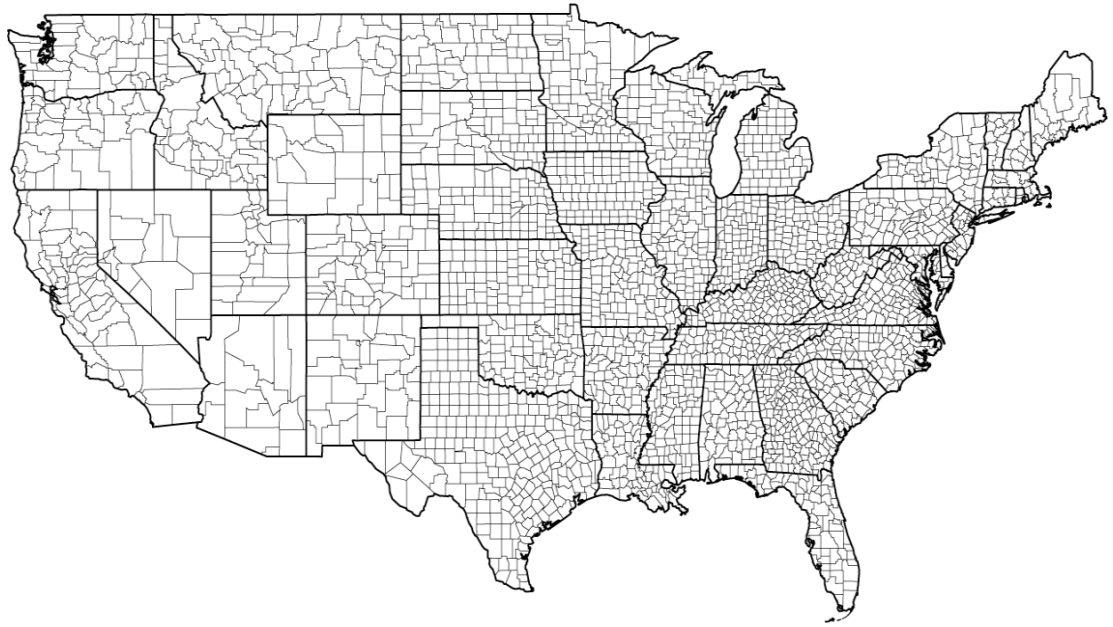


Prediction of the infection trend of COVID-19 in Michigan.

Models for community-level risk prediction

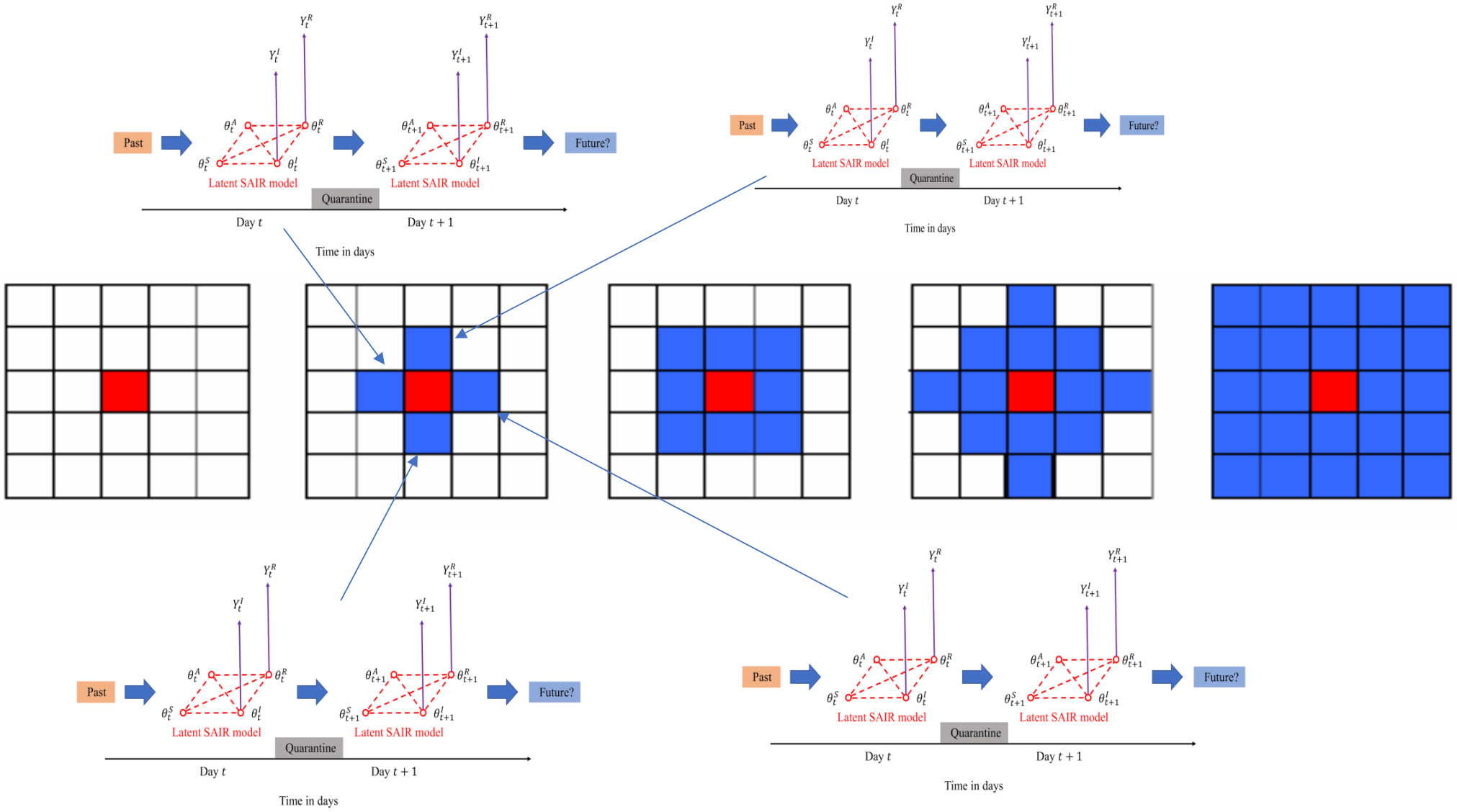
- To develop **a health information system** that provides **timely risk evaluation** and **prediction of the COVID-19 infection** in **communities.**

- Account for temporally varying **control measures** and **spatial heterogeneities** in the COVID-19 infection dynamics.

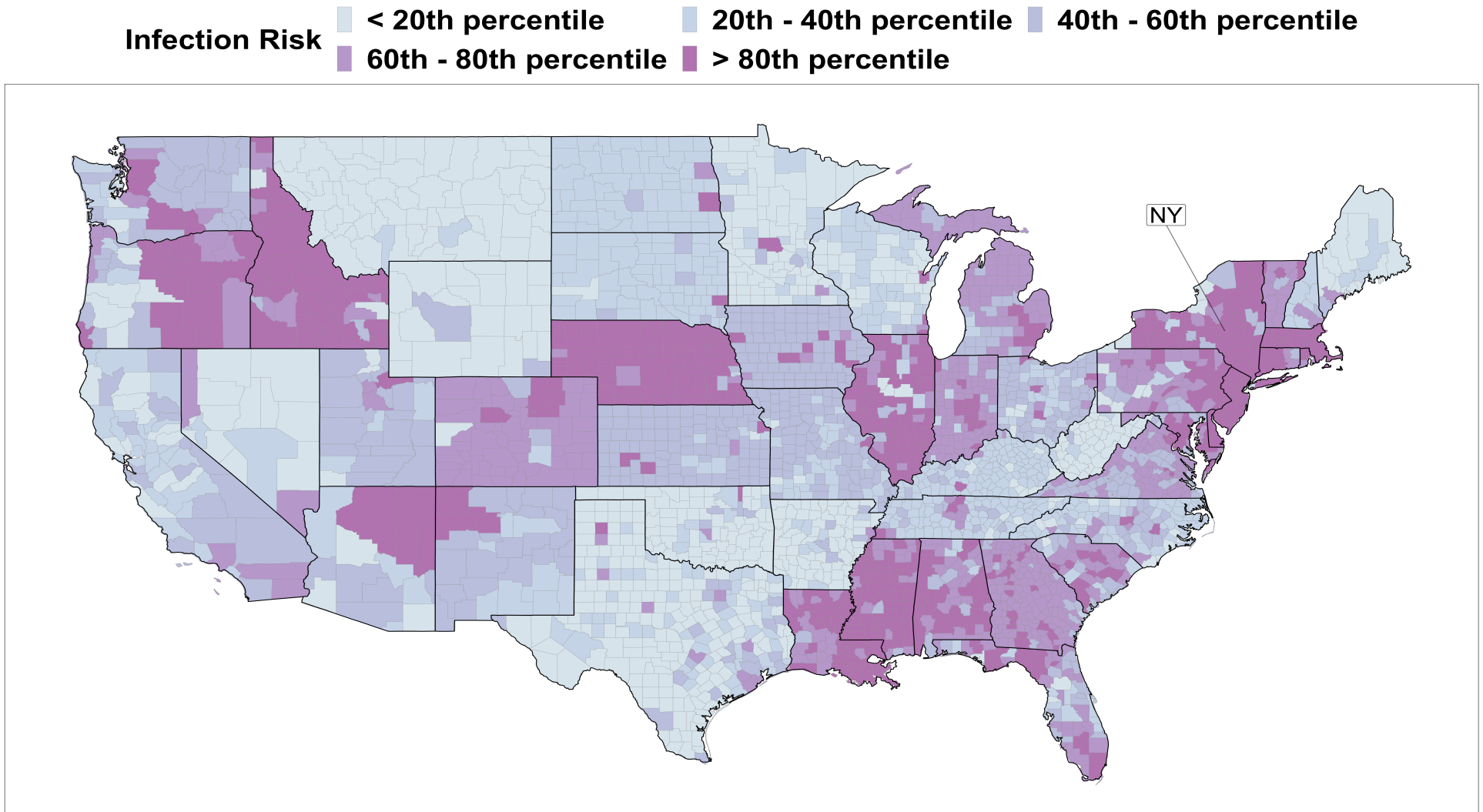


Zhou et al. (2020). A Spatiotemporal Epidemiological Prediction Model to Inform County-level COVID-19 Risk in the USA. *Harvard Data Science Review (to appear)*

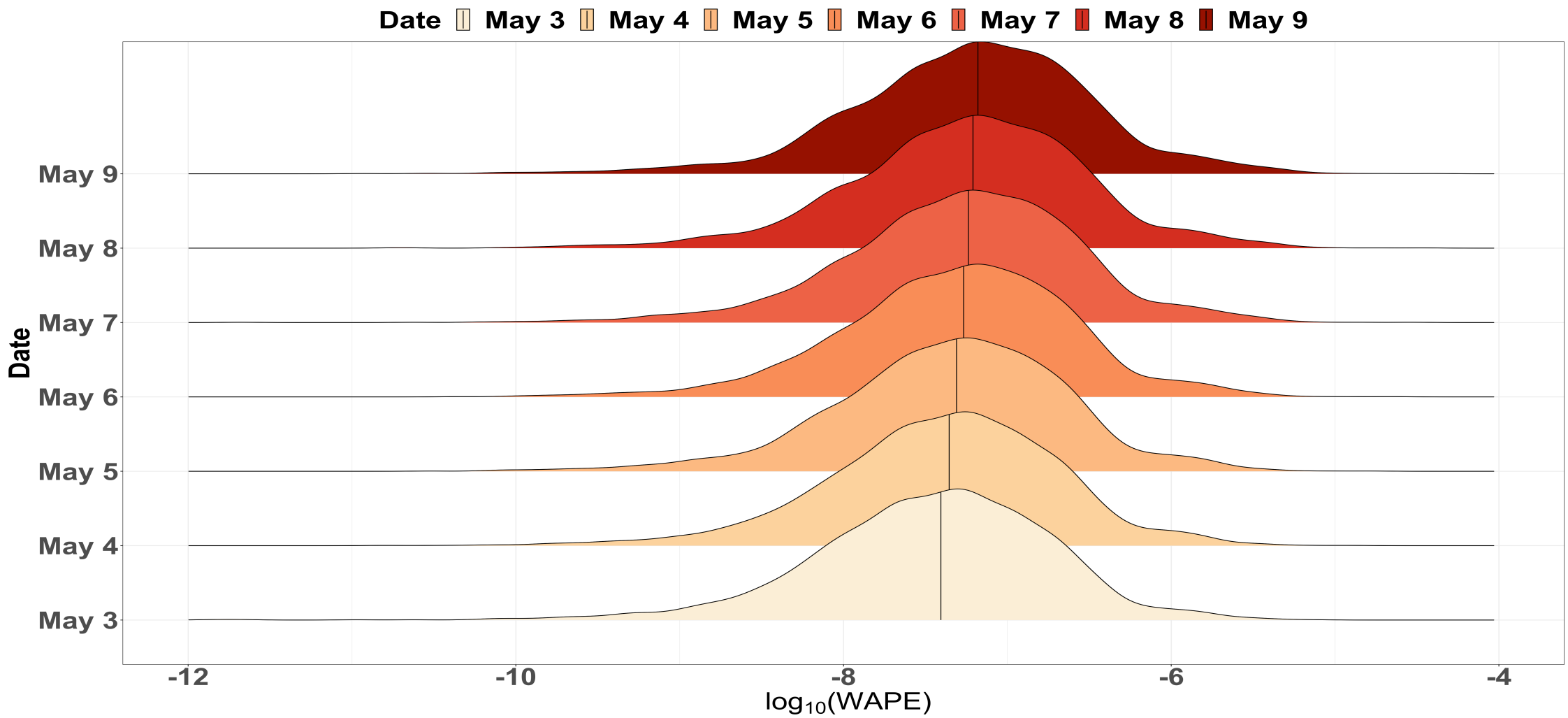
Spatiotemporal Cellular-Automata



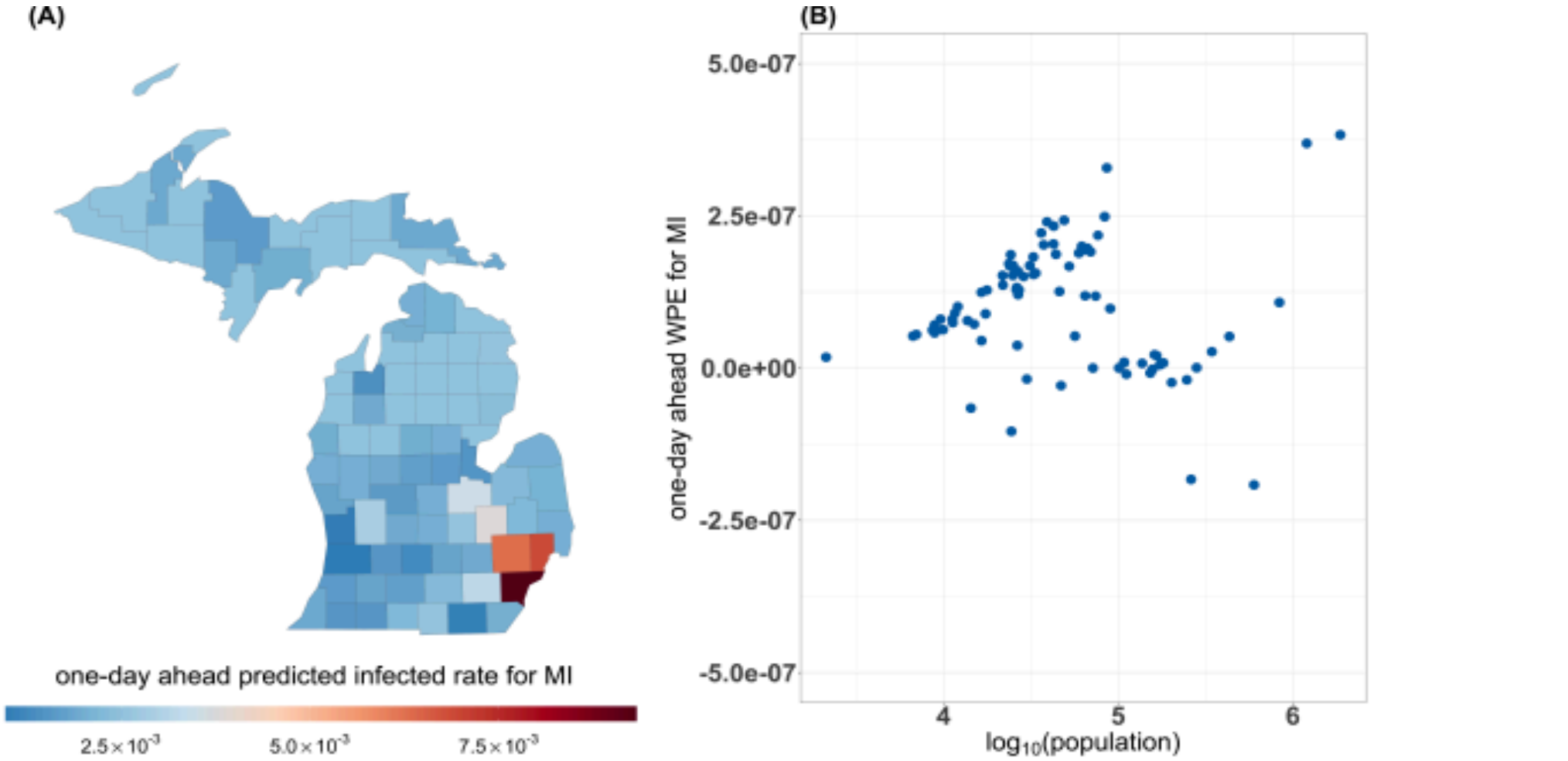
7-day ahead risk prediction on May 2 (3109 counties)



County-level weighted absolute prediction errors

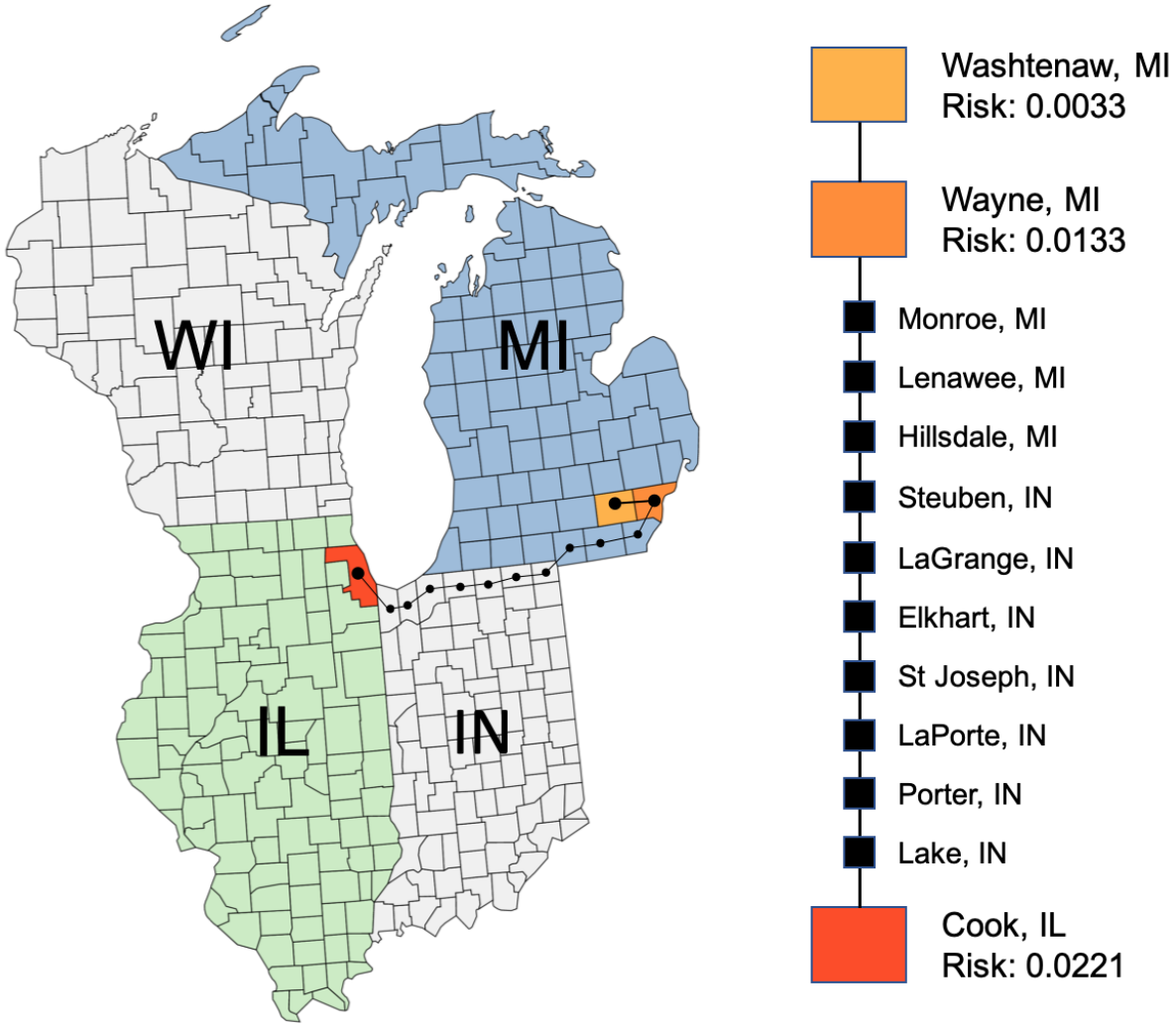


1-day ahead prediction of county-level risk in Michigan



(A) One-day ahead county-level predicted infectious prevalences in MI on May 2.
(B) One-day ahead weighted prediction error for all the counties of MI on May 2.

Risk prediction of a trip



- The risk score for this travel route is calculated to be 3.84×10^{-4}.
- Considering this, the traveler may revise their travel plan if this projected risk is too high.
- One situation may arise where a traveler takes a longer route (in terms of time/distance) in order to lower the travel risk.

Concluding Remarks:
Data source, quality and sharing

- **Data Quality:** Reliable and well-validated benchmark data

- **Serological survey:** Capture asymptomatic subjects, which is critical to overcome the issue of underreporting for infected cases.

- **Mobile tracking data:** Evaluate individual mobility and movement trajectory to derive personal risk of infection

- Multi-modality data integration and conquer data privacy issues.

Concluding Remarks: Future Directions

- **Transparency and reproducibility:** Open-source software (the IHME model and Gu's machine learning methods)
- **Bias correction:** Methods to deal with various types of biases in surveillance data.
- **Model diagnostics:** Methods to confirm model assumptions made in the data analysis.
- **Subgroup Analysis:** Age, race, other covariates to perform strata-specific analyses.
- **Meta analysis:** Combine results from different models
- **Online Learning:** Fast algorithms to update models and analysis results.

References

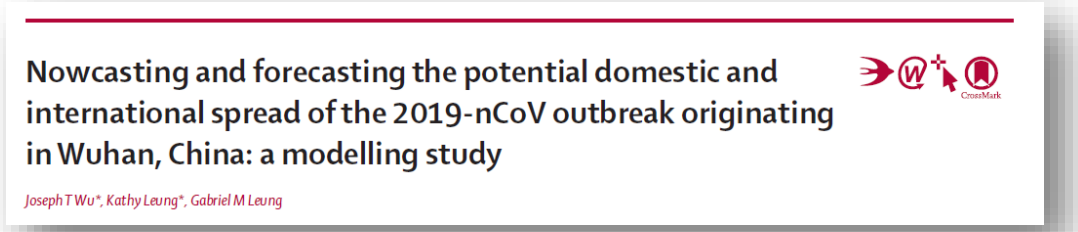
1. 1point3acres covid-19 us data. https://coronavirus.1point3acres.com/zh, 2020.
2. **China data lab 2020-02-202020 draft version covid-19, us, map harvard dataverse. https://doi.org/10.7910/DVN/HIDLTKdoi/10.7910/DVN/HIDLTK, 2020.**
3. List of airports in the united states. https://en.wikipedia.org/wiki/List_of_airports_in_the_United_States, 2016.
4. Social distancing scoreboard. https://www.unacast.com/covid19/social-distancing-scoreboard?view=county&fips=08097, 2020.
5. University of Maryland covid-19 impact analysis platform. https://data.covid.umd.edu/, 2020.
6. John Von Neumann, Arthur W Burks, et al. Theory of self-reproducing automata. IEEE Transactions on Neural Networks, 5(1):3-14, 1966.
7. Lili Wang, Yiwang Zhou, Jie He, Bin Zhu, Fei Wang, Lu Tang, Michael Kleinsasser, Daniel Barker, Marisa Eisenberg, and Peter X Song. An epidemiological forecast model and software assessing interventions on covid-19 epidemic in china (with discussion). Journal of Data Sciences, 2020.
8. Zhou, Y., Wang, L., Zhang, L., Shi, L., Yang, K., He, J., Zhao, B., Overton, W., Purkayastha,S., and Song, P.X.K. A spatiotemporal epidemiological prediction model to inform county-level covid-19 risk in the USA. Harvard Data Science Review, Special Issue 1 - COVID-19, 2020.
9. Lu Tang, Yiwang Zhou, Lili Wang, Soumik Purkayastha, Leyao Zhang, Jie He, Fei Wang, and Peter X.K. Song. A review of multi-compartment models for infectious diseases. International statistical review, 2020.

讨论：疫情模型回顾

Qiushi Chen 陈秋实
q.chen@psu.edu
Department of Industrial and Manufacturing Engineering
The Pennsylvania State University

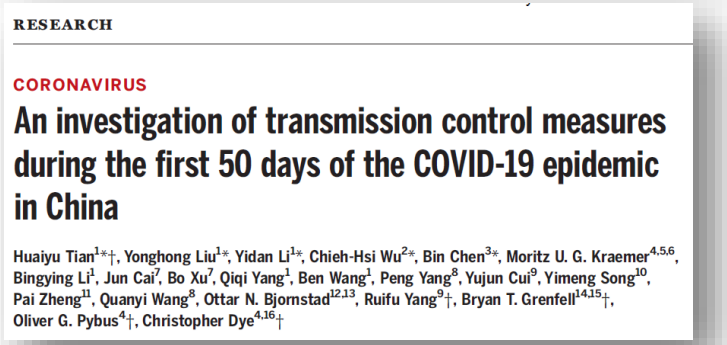
Modeling relies on study objectives

- **Key parameter inferences**, to **explain**
  - Wu 2020 Lancet: Estimate R_0
  - Tian 2020 Science: Estimate the reduction in transmission rate by the implemented control measures
- **Prediction and projection** of epidemic trend, to **evaluate** potential intervention alternatives
  - Ferguson 2020: Isolation, quarantine, school closure, etc.
  - Kissler 2020 Science: Social distancing: Duration? Intermittent strategy?



Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study

Joseph T Wu*, Kathy Leung*, Gabriel M Leung

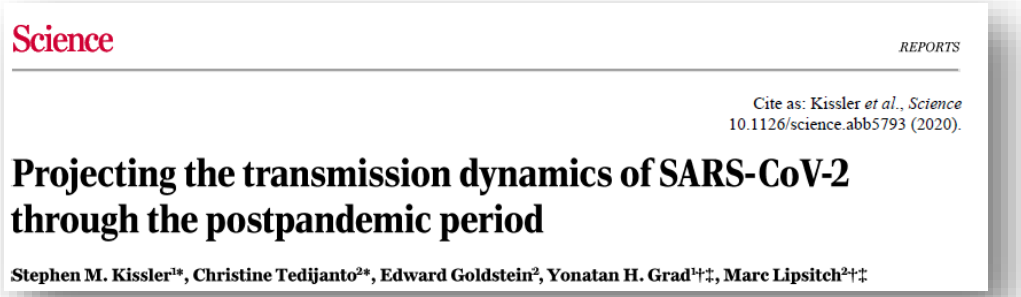


RESEARCH

CORONAVIRUS

An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China

Huaiyu Tian[1*†], Yonghong Liu[1*], Yidan Li[1*], Chieh-Hsi Wu[2*], Bin Chen[3*], Moritz U. G. Kraemer[4,5,6], Bingying Li[1], Jun Cai[7], Bo Xu[7], Qiqi Yang[1], Ben Wang[1], Peng Yang[8], Yujun Cui[9], Yimeng Song[10], Pai Zheng[11], Quanyi Wang[8], Ottar N. Bjornstad[12,13], Ruifu Yang[9†], Bryan T. Grenfell[14,15†], Oliver G. Pybus[4†], Christopher Dye[4,16†]

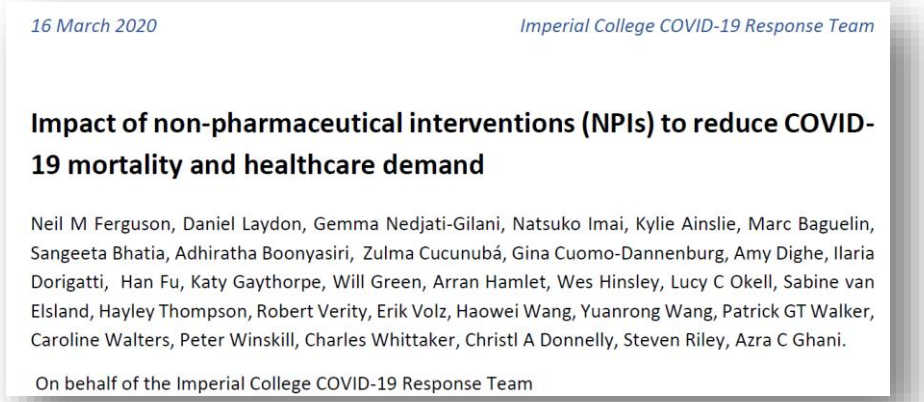


Science REPORTS

Cite as: Kissler *et al.*, *Science* 10.1126/science.abb5793 (2020).

Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period

Stephen M. Kissler[1*], Christine Tedijanto[2*], Edward Goldstein[2], Yonatan H. Grad[1†‡], Marc Lipsitch[2†‡]



16 March 2020 Imperial College COVID-19 Response Team

Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand

Neil M Ferguson, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg, Amy Dighe, Ilaria Dorigatti, Han Fu, Katy Gaythorpe, Will Green, Arran Hamlet, Wes Hinsley, Lucy C Okell, Sabine van Elsland, Hayley Thompson, Robert Verity, Erik Volz, Haowei Wang, Yuanrong Wang, Patrick GT Walker, Caroline Walters, Peter Winskill, Charles Whittaker, Christl A Donnelly, Steven Riley, Azra C Ghani.

On behalf of the Imperial College COVID-19 Response Team

Which type of models should you use?

- "Model" is a broad term!
  - Regression model, econometric model, time-series model, machine learning models…
  - "Top-down" approach

- **Epidemic model**: "Bottom-up" approach
  - Compartment model (deterministic, stochastic)
  - Branching process

Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts

Joel Hellewell, Sam Abbott*, Amy Gimma*, Nikos I Bosse, Christopher I Jarvis, Timothy W Russell, James D Munday, Adam J Kucharski, W John Edmunds, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Sebastian Funk†, Rosalind M Eggo†

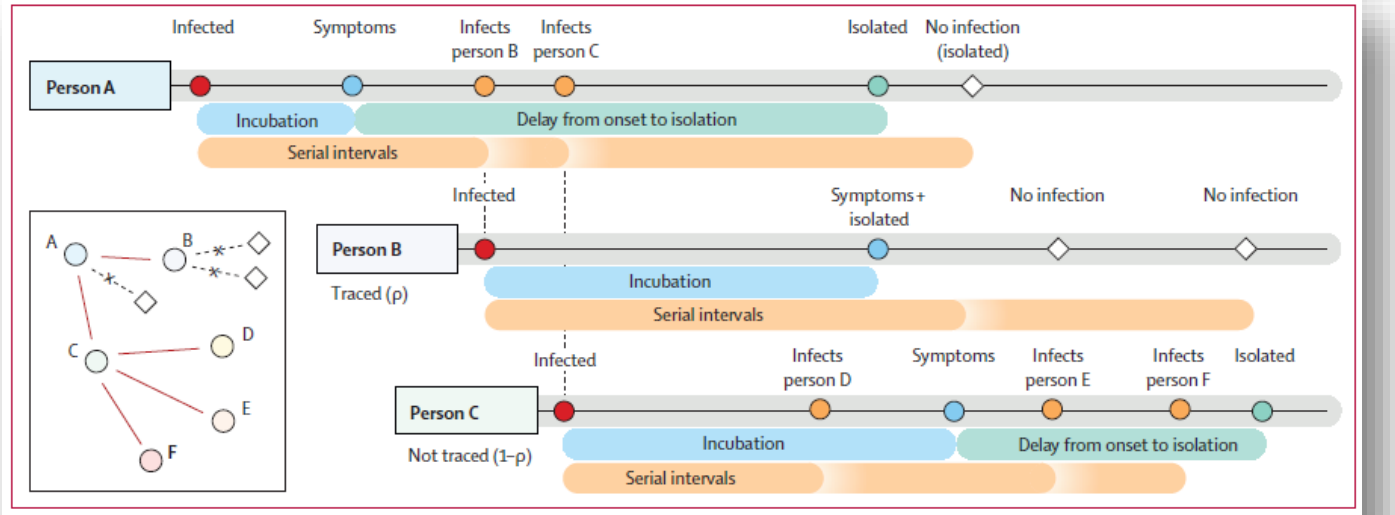
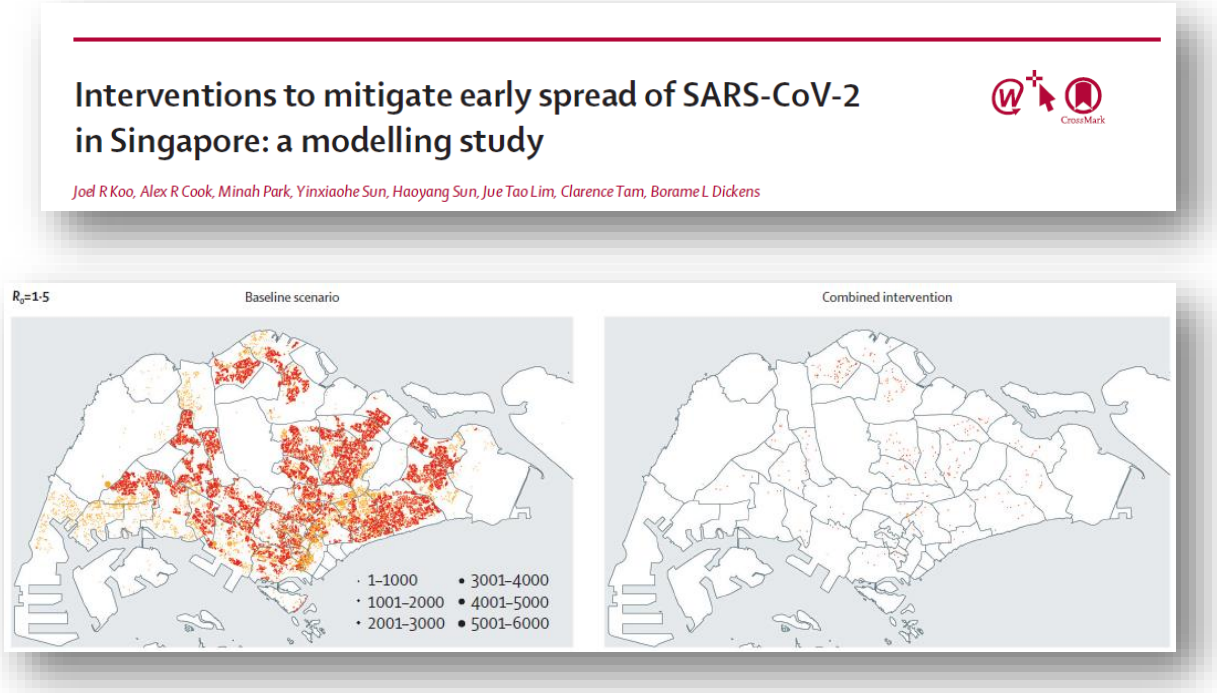


Figure 1: Example of the simulated process that starts with person A being infected

Which type of models should you use?

- "Model" is a broad term!
  - Regression model, econometric model, time-series model, machine learning models…
  - "Top-down" approach

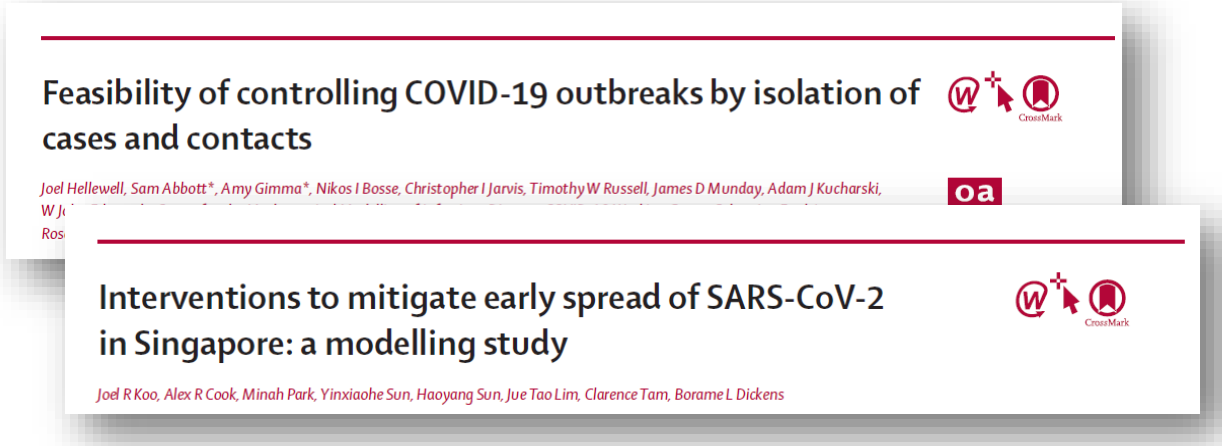
- **Epidemic model**: "Bottom-up" approach
  - Compartment model (deterministic, stochastic)
  - Branching process
  - Network model
  - Agent-based simulation model
    - Individual features, home/workplace/school, geospatial factors, etc.



Which type of models should you use?

- "Model" is a broad term!
  - Regression model, econometric model, time-series model, machine learning models…
  - "Top-down" approach

- **Epidemic model**: "Bottom-up" approach
  - Compartment model (deterministic, stochastic)
  - Branching process
  - Network model
  - Agent-based simulation model
    - Individual features, home/workplace/school, geospatial factors, etc.



- Challenges
  - Data availability, parameter complexity, computational efforts

Lessons learned and additional remarks

- How much details are enough? How realistic the model needs to be?
  - Consider computational complexity in parameter estimation/calibration
    - E.g., MCMC or high-dimensional optimization
  - Capture major driving factors
  - Sensitivity analysis, scenario analysis

- Develop the model for a specific country setting? For generic policy implications?

- A rich body of literature in epidemic modeling **before** COVID-19
  - Influenza, Ebola, HIV, TB, Hep C …
  - SMDM COVID-19 modeling repository (60+ models): https://tinyurl.com/SMDMcovidModeling