CONFRONTING MODELS WITH DATA

Micaela E. Martinez

Emory University

SPECTRUM OF MODELS

There is a spectrum of models ranging from purely theoretical models to fully data driven models. They differ by the amount of data integration into the model itself and the level to which the model is fit to data

theoretical
models

data-driven
models

Daniel Bernoulli's epidemiological model revisited

Klaus Dietz [a,*], J.A.P. Heesterbeek [b]

[a] Department of Medical Biometry, University of Tübingen, Westbahnhofstr. 55, 72070 Tübingen, Germany
[b] Faculty of Veterinary Medicine, University of Utrecht, Yalelaan 7, 3584 CL Utrecht, The Netherlands

Let $u(a)$ denote the probability for a newborn individual to be alive and susceptible at age a. Then $u(a)$ satisfies the differential equation

$$\frac{du}{da} = -[\lambda(a) + \mu(a)]u, \tag{1}$$

with the initial condition $u(0) = 1$.

The probability $w(a)$ to be immune and alive is given by

$$\frac{dw}{da} = [1 - c(a)]\lambda(a)u(a) - \mu(a)w, \tag{2}$$

with the initial condition $w(0) = 0$.

The solutions of these equations are

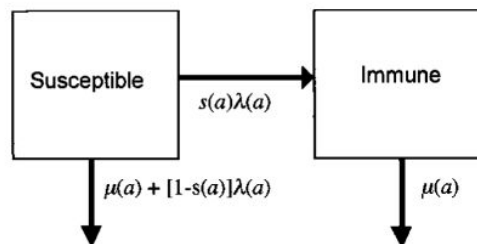$$u(a) = \exp\{-[\Lambda(a) + M(a)]\}, \tag{3}$$

Fig. 2. States and transitions of Bernoulli's epidemiological model for an immunizing infection in a cohort which is in equilibrium with respect to time. $s(a)$ = probability of surviving the infection. $\lambda(a)$ = force of infection; $\mu(a)$ = death-rate due to other diseases.

Fig. 1. Daniel Bernoulli (1700–1782). (Section from a painting by Nicolaus Grooth in 1760.)

In the 1700s Bernoulli created an age structured Susceptible-Immune model to estimate the change in life expectancy if smallpox were eradicated

THEORETICAL MODELS

- Theoretical models, such as Bernoulli's smallpox model have used to explore the behavior and impact of epidemics in theoretical populations, where simulations are run without one or more set of parameters that are not tied to a particular real-world population

FULLY THEORETICAL

(1)
$$N = S + I + R$$

(2)
$$\frac{dS}{dt} = \mu N - \beta IS - \delta S$$

(3)
$$\frac{dI}{dt} = \beta IS - \gamma I - \alpha I - \delta I$$

(4)
$$\frac{dR}{dt} = \gamma I - \delta R$$

Here S, I, and R are fractions of the population and therefore can represent any population

MODEL WITH REAL-WORLD POPULATION

(1)
$$N = S + I + R$$

Set N equal to the population under study.

(2)
$$\frac{dS}{dt} = \mu N - \beta IS - \delta S$$

S, I, and R initial conditions can be fit to data and/or we can base them on real world data

(3)
$$\frac{dI}{dt} = \beta IS - \gamma I - \alpha I - \delta I$$

(4)
$$\frac{dR}{dt} = \gamma I - \delta R$$

ADDING DEMOGRAPHY

Demography may be added to models as covariates, fixed or time-varying parameters, and/or constraints on initial conditions:

- population size
- births
- natural death (i.e., deaths from causes other than for the disease in study
- immigration
- emigration

ADDING REAL-WORLD POPULATION SIZE

$$(1) \quad N = S + I + R$$

$$(2) \quad \frac{dS}{dt} = \mu N - \beta IS - \delta S$$

$$(3) \quad \frac{dI}{dt} = \beta IS - \gamma I - \alpha I - \delta I$$

$$(4) \quad \frac{dR}{dt} = \gamma I - \delta R$$

Consider COVID-19 in NYC in March 2020. N = 8,461,961

We can assume that everyone was susceptible at the time SARS-CoV-2 came to NYC and there was initially one infected

```
> dat[1:10,c('date',"borough","new.cases.corrected","pop")]
      date borough new.cases.corrected      pop
6    3/2/20   total                   0  8461961
12   3/3/20   total                   0  8461961
18   3/4/20   total                   2  8461961
24   3/5/20   total                   2  8461961
30   3/6/20   total                   7  8461961
36   3/7/20   total                   0  8461961
42   3/8/20   total                   4  8461961
48   3/9/20   total                  12  8461961
54  3/10/20   total                  24  8461961
60  3/11/20   total                  13  8461961
>
```

ADDING REAL-WORLD BIRTHS & DEATHS

(1)
$$N = S + I + R$$

(2)
$$\frac{dS}{dt} = \mu N - \beta IS - \delta S$$

(3)
$$\frac{dI}{dt} = \beta IS - \gamma I - \alpha I - \delta I$$

(4)
$$\frac{dR}{dt} = \gamma I - \delta R$$

- Now consider the birth rate and the natural death rate in a population.

- Births can be added in as a rate or as a covariate time series.

- Non-specific mortality can be entered as a fixed or time varying rate.

- Incorporating data on time varying births and deaths is particularly important over long time horizons when birth and death rate vary and the population size fluctuates

Rise and fall of polio in the 20ᵗʰ century

ME Martinez-Bakker, King, & Rohani 2015. PLoS Biology
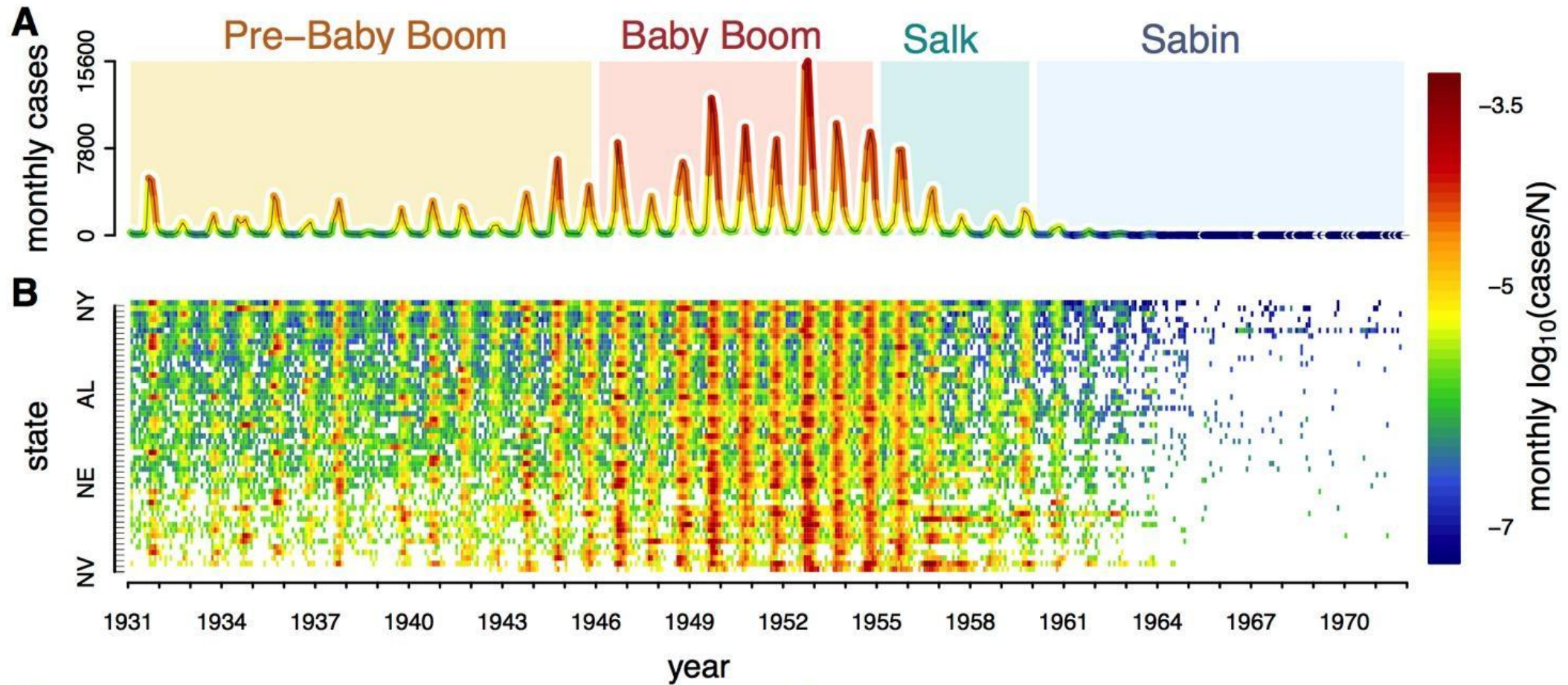
Incorporating data on time-varying births is important for studying childhood disease dynamics during time periods when birth rates varied because births are the source of susceptible recruitment

Rise and fall of polio in the 20ᵗʰ century
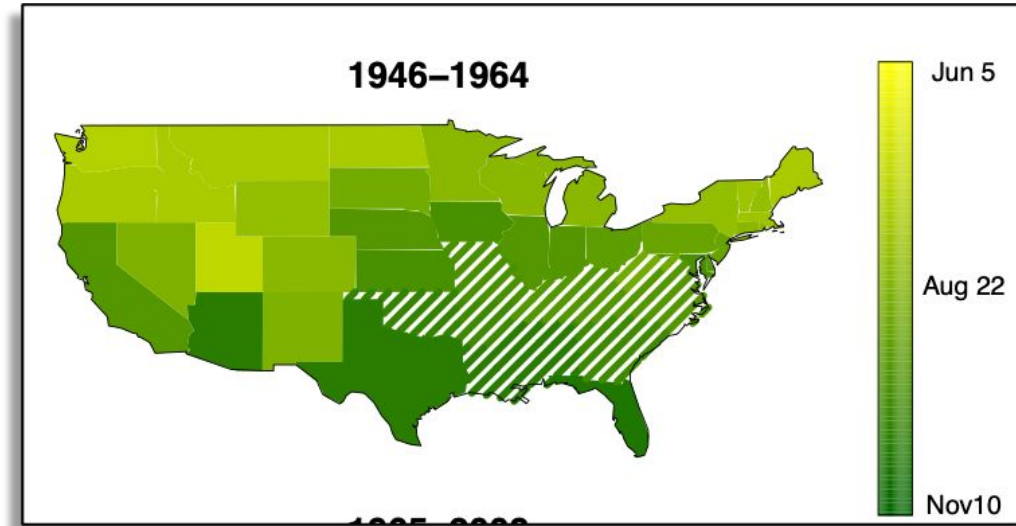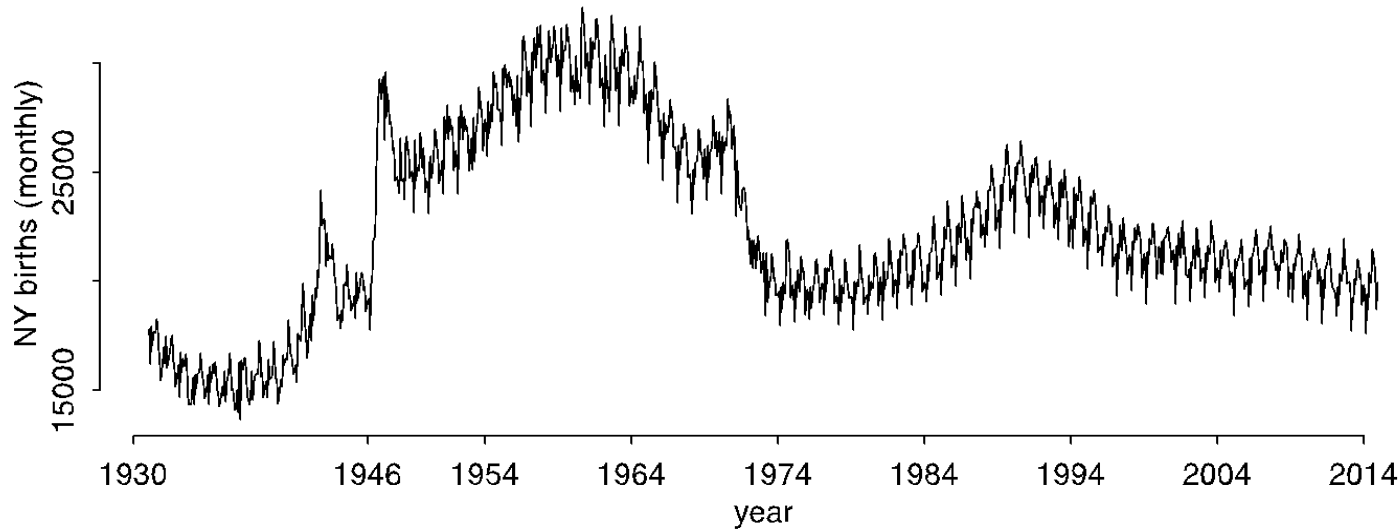
Incorporating data on births is important for capturing time-varying birth rates
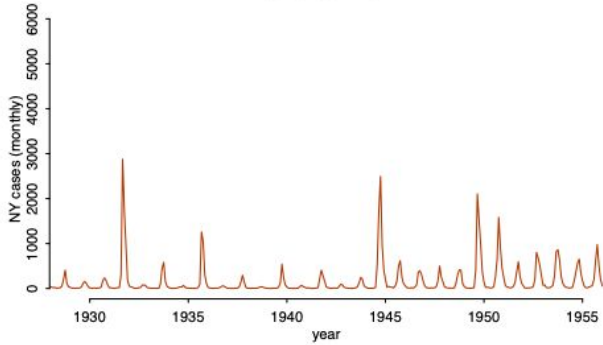
Birth data is important for studying multiple populations where birth have spatial variation

ME Martinez-Bakker, Bakker, King & Rohani (2014)
Proc B
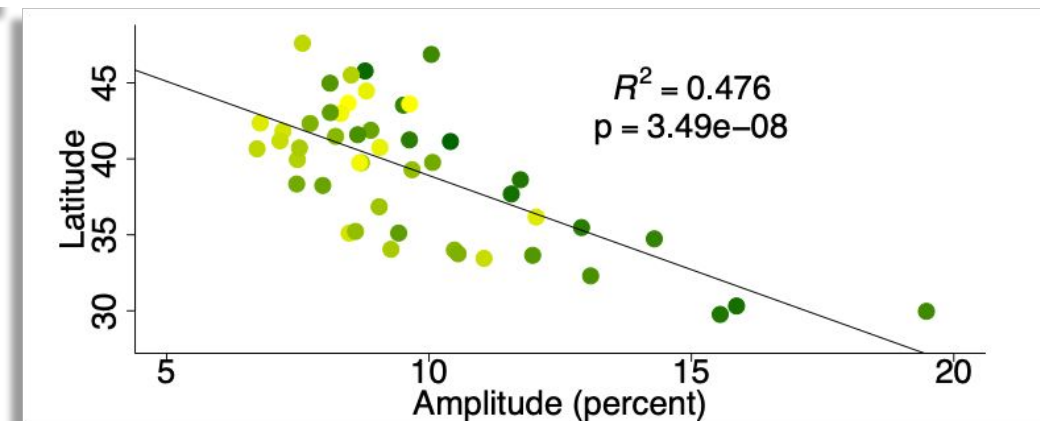
Infectious Diseases

measles: spring
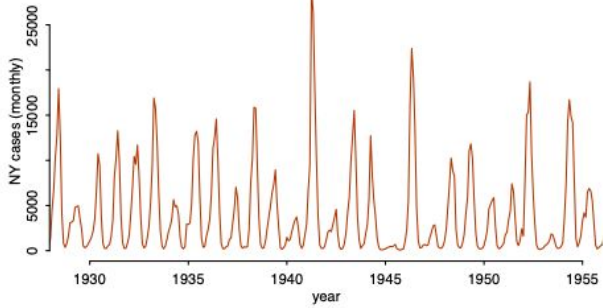
polio: summer

flu: winter

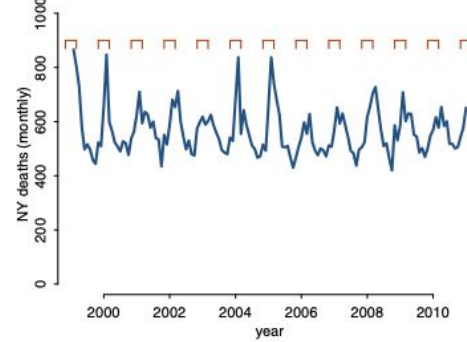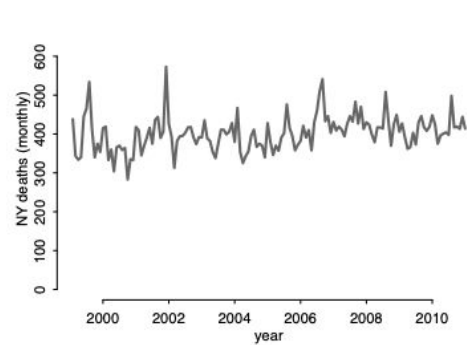**ME Martinez** (unpublished)
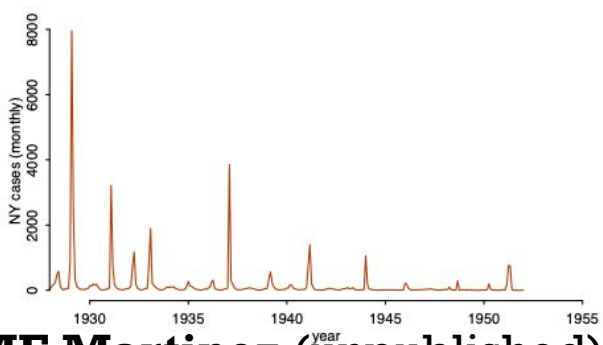
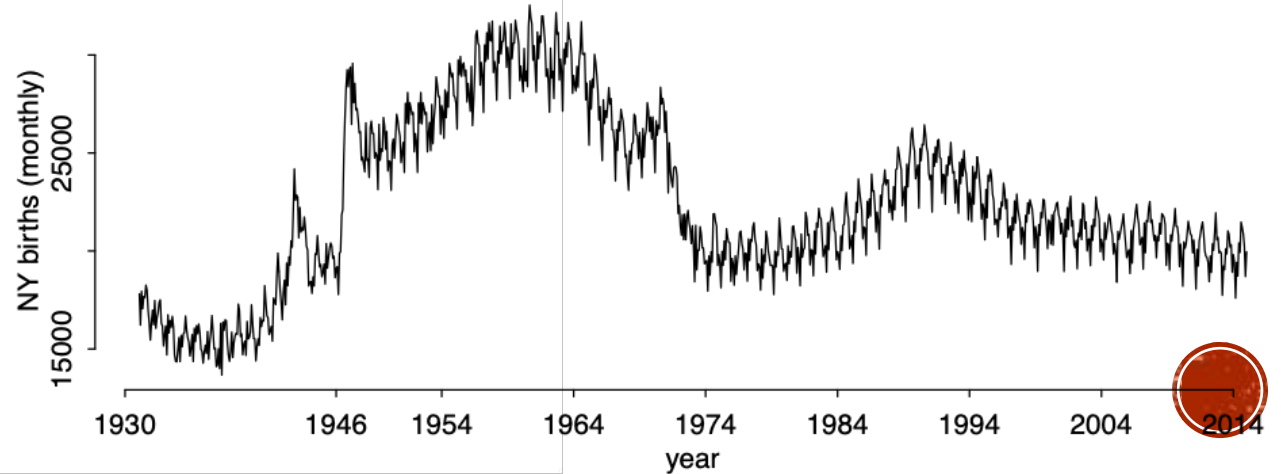# Non-Infectious Disease Mortality

Heart Attack

Oct–Feb
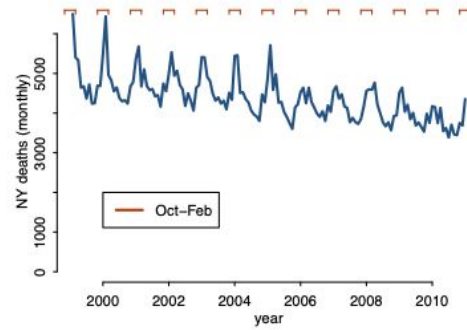
Stroke
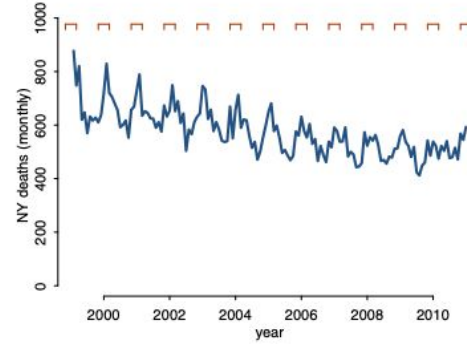
Alzheimer's Disease
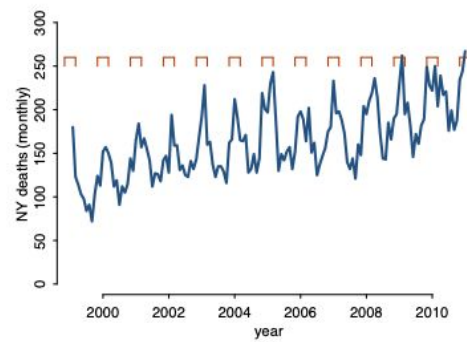
COPD

Accidents

Including variation in births, mortality, and/or population size can be important when there is co-variation over. Births and non-specific deaths can impact transmission through their impact on the size of the susceptible pool.

RESEARCH ARTICLE

Long-term dynamics of measles in London: Titrating the impact of wars, the 1918 pandemic, and vaccination

Alexander D. Becker[1]*, Amy Wesolowski[2], Ottar N. Bjørnstad[3,4], Bryan T. Grenfell[1,4,5]

1 Department of Ecology and Evolutionary Biology, Princeton, New Jersey, United States of America,
2 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, 3 Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, Pennsylvania, United States of America, 4 Fogarty International Center, National Institutes of Health, Bethesda, Maryland, United States of America, 5 Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, New Jersey, United States of America
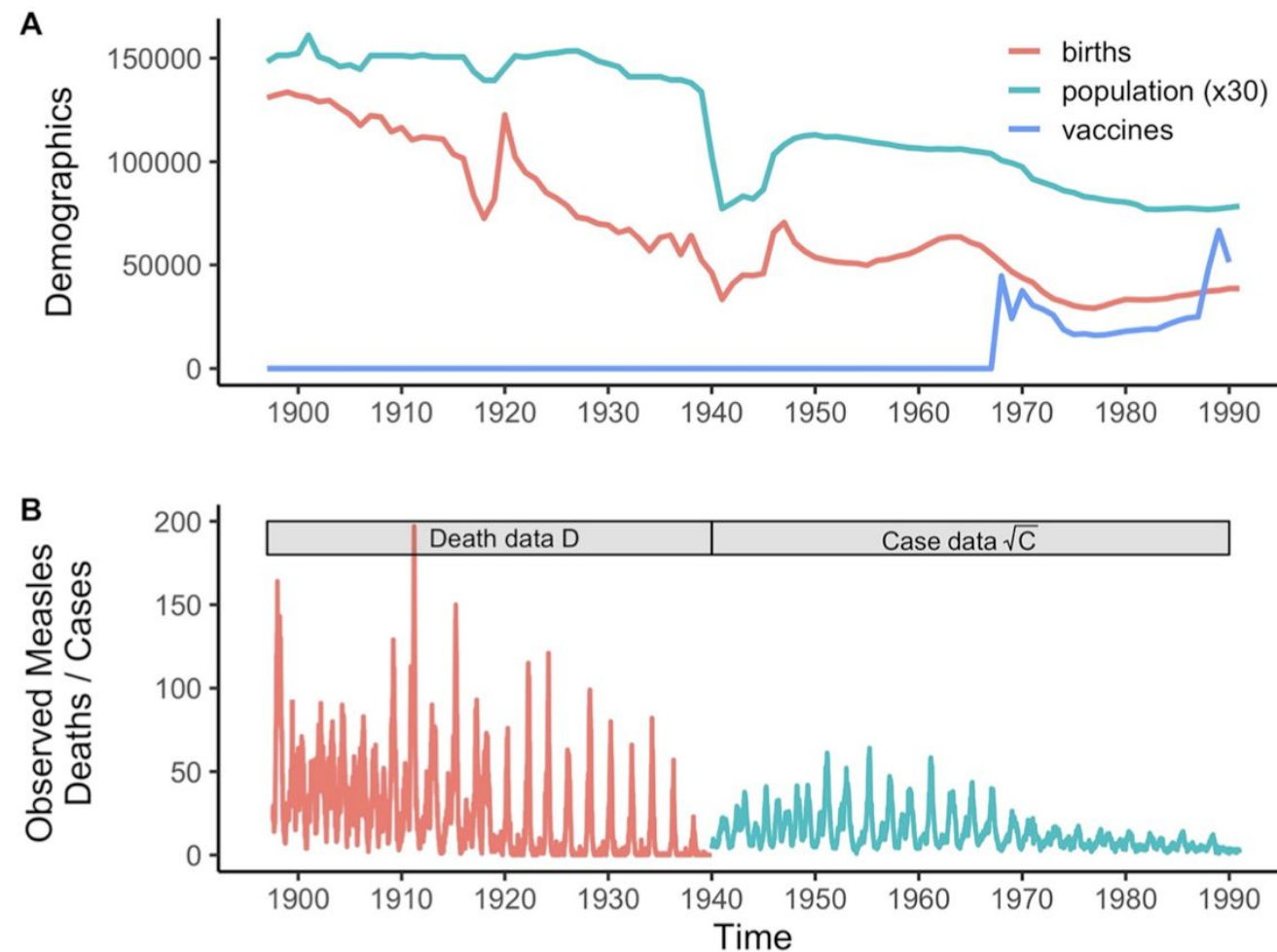
* adbecker@princeton.edu

Fig 1. **The demographic, vaccination, and measles data analyzed.** A) The observed population dynamics shown on a yearly scale. The major demographic fluctuations to births (red) and population counts (green) caused by WWII can be seen starting in 1940. B) Measles dynamics for London 1897–1991, shown on a weekly time scale with mortality (red) until 1940, and incidence (blue) through 1990. Note the case data are shown on a square root scale. Unscaled data are shown (inverted) in Fig 2.

ADDING REAL-WORLD IMMIGRATION & EMIGRATION

- immigration and emigration are often used in coupled SIR models where multiple populations are coupled via migration and SIR. Since the migration of infected individuals will have the most impact on the system, we can explore an example of only tracking infection migration

$$\frac{dS}{dt} = \mu N - \beta IS - \delta S$$

$$\frac{dI}{dt} = \beta IS - \gamma I - \alpha I - \delta I + e_2 I_2 - e_1 I_1$$

$$\frac{dR}{dt} = \gamma I - \delta R$$

$$\frac{dS}{dt} = \mu N - \beta IS - \delta S$$

$$\frac{dI}{dt} = \beta IS - \gamma I - \alpha I - \delta I + e_1 I_1 - e_2 I_2$$

$$\frac{dR}{dt} = \gamma I - \delta R$$

ADDING COVARIATES

- Covariates are time series data or time-varying functions put into the model that covary with the infection process

- Seasonal covariates are common in models of recurrent epidemics

- covariates may represent behavioral changes, economic changes, interventions (essentially any time varying elements important that impact disease dynamics other than demography)

SEASONAL FORCING OF TRANSMISSION

- Transmission may be <u>seasonally forced</u> via environmental conditions, host behavior, and/or host physiology

- Common seasonal forcing include <u>term-time forcing</u> based on school terms in childhood disease models, as well as climate and weather forcing for infections whose transmission is sensitive to environmental conditions

TEMPERATURE AND HUMIDITY IN MODELS

Specific humidity impacts flu survival and transmission and may be added into models as a covariate into Beta(t) to capture the seasonal transmission of flu
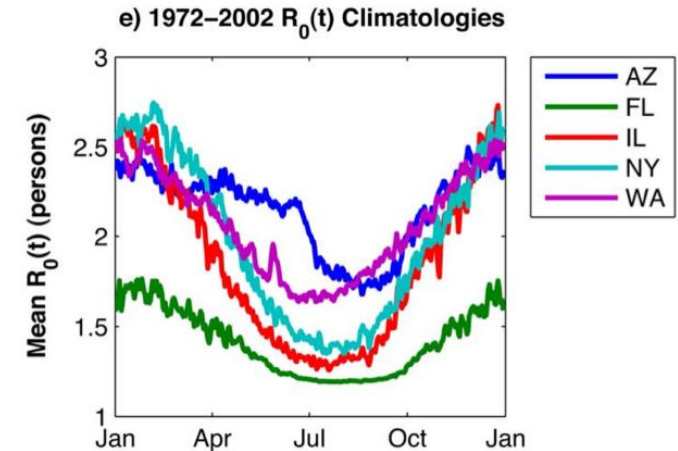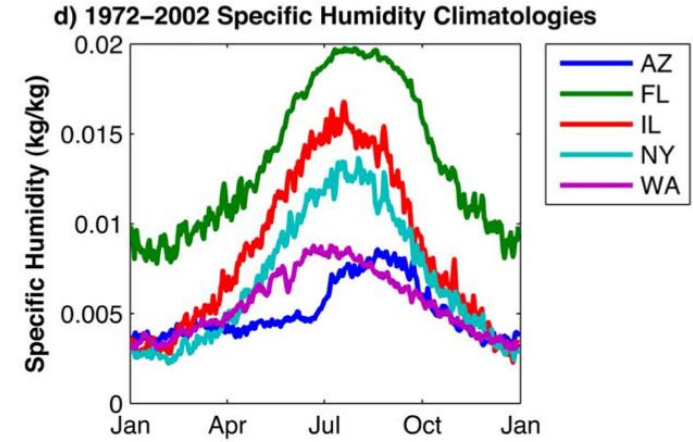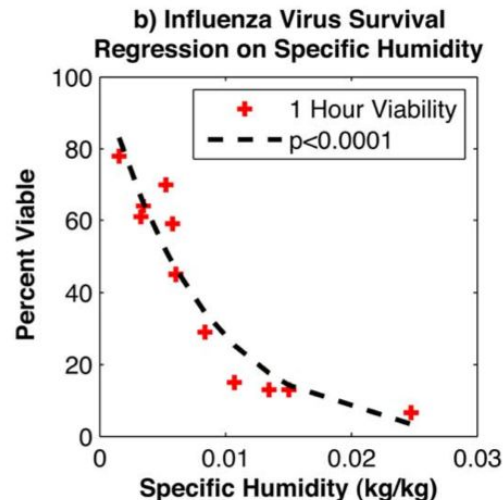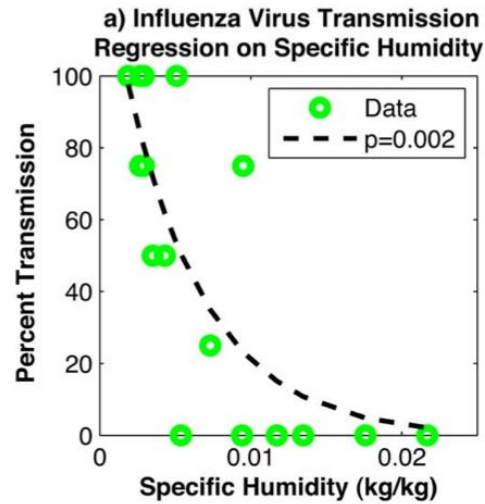
## Absolute Humidity and the Seasonal Onset of Influenza in the Continental United States

Jeffrey Shaman[1]*, Virginia E. Pitzer[2,3,4], Cécile Viboud[2], Bryan T. Grenfell[2,4,5], Marc Lipsitch[6,7,8]

SCHOOL TERMS

School terms can be added into models of childhood disease transmission to modulate Beta(t) seasonally.

B-splines can also be uses to estimate the seasonal transmission rate as shown here for chickenpox in Thailand

Bakker et al. Amer J Epi 2021
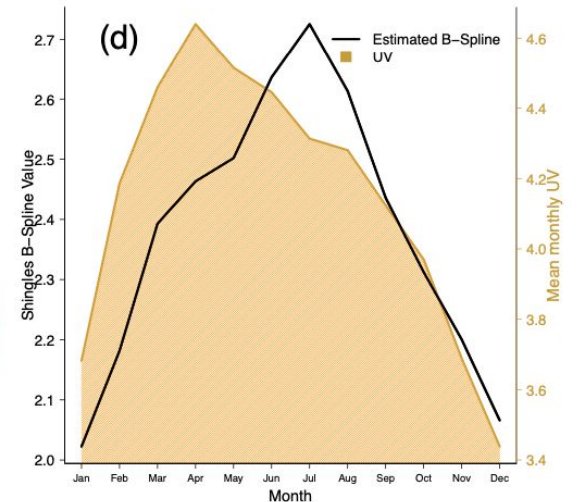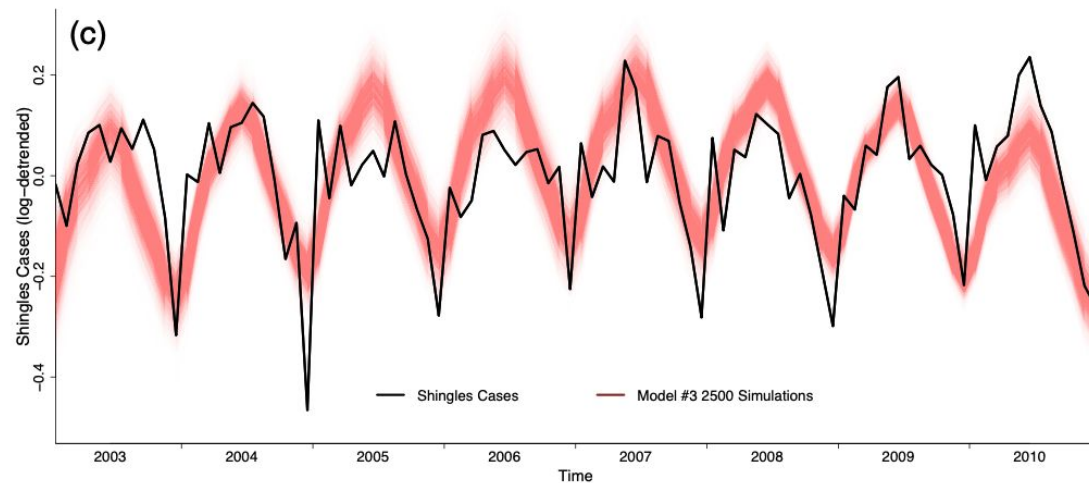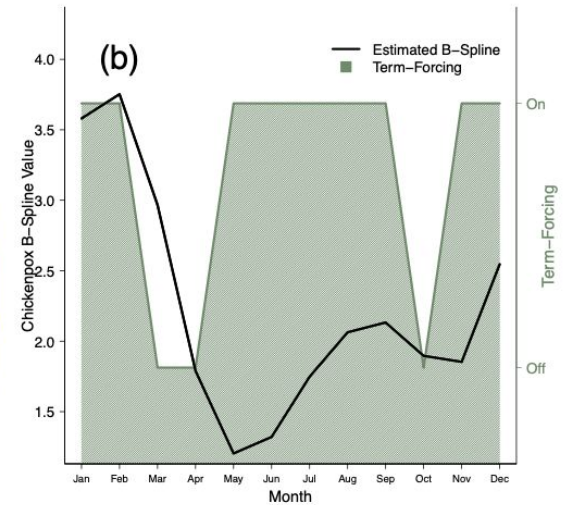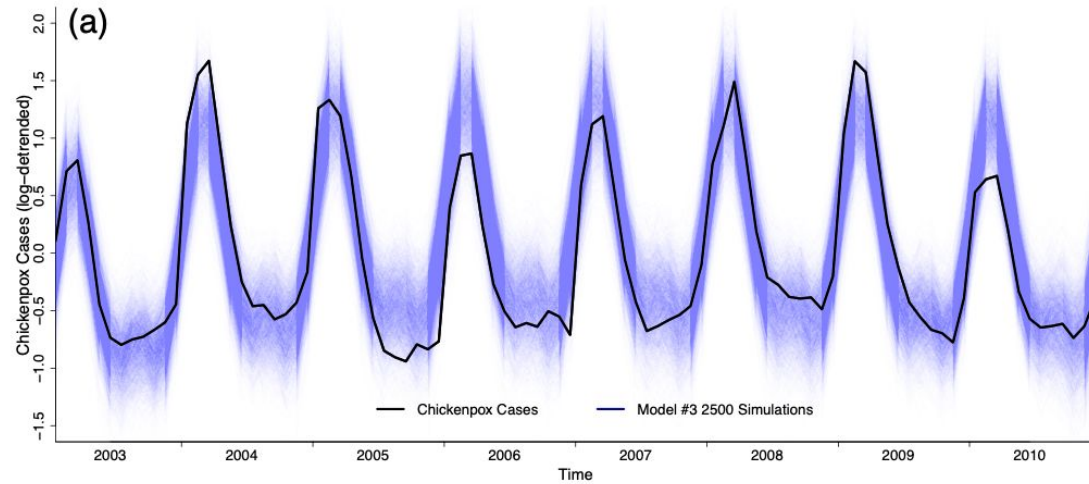
MOBILITY
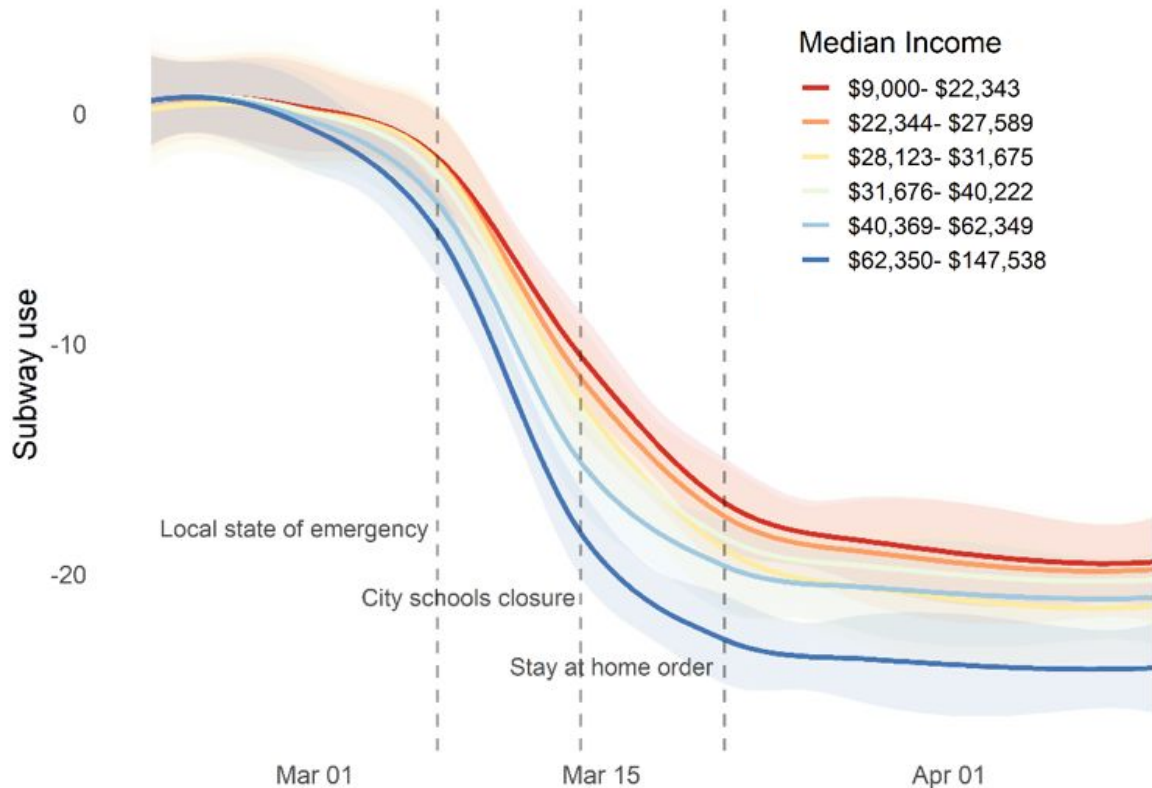
Socioeconomic Disparities in Subway Use and COVID-19 Outcomes in New York City

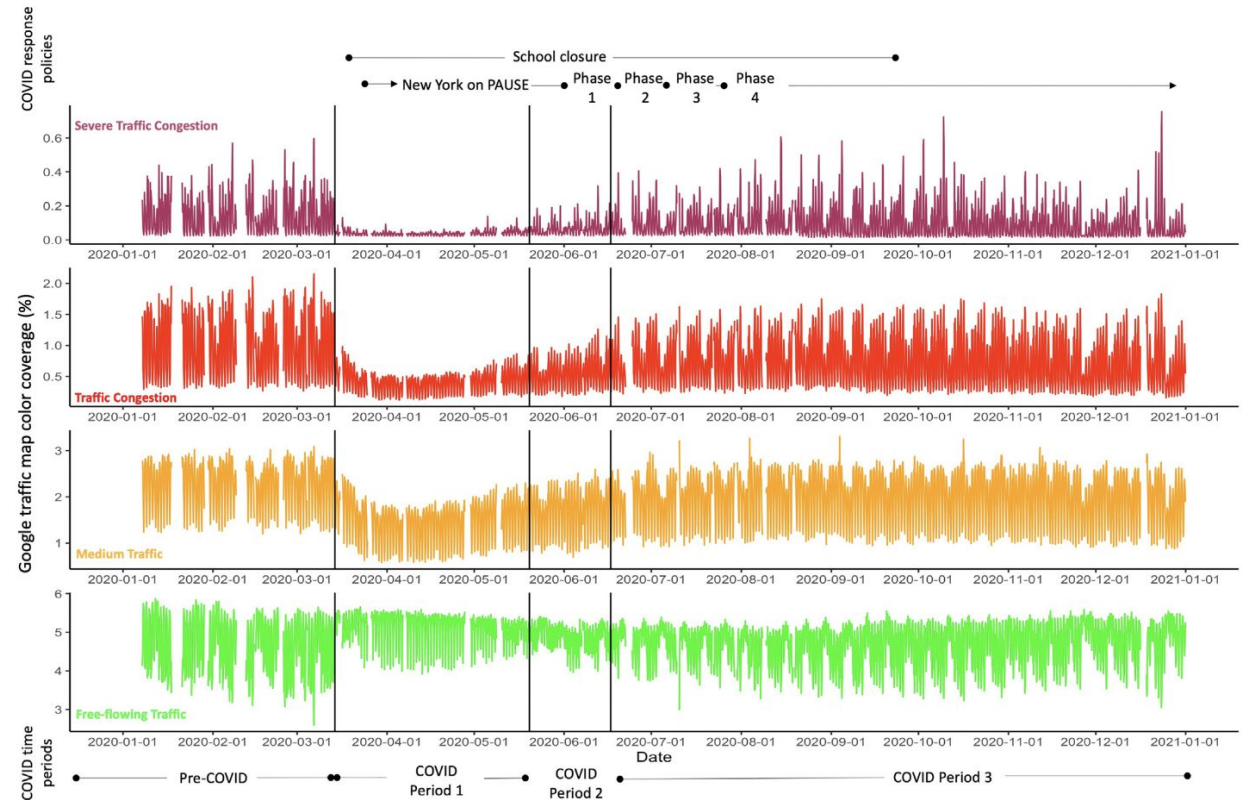Karla Therese L. Sy*, Micaela E. Martinez*, Benjamin Rader, and Laura F. White

Social-distancing fatigue: Evidence from real-time crowd-sourced traffic data

Jenni A. Shearston *, Micaela E. Martinez, Yanelli Nunez, Markus Hilpert

Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, 722 West 168th St., New York, NY 10032, USA
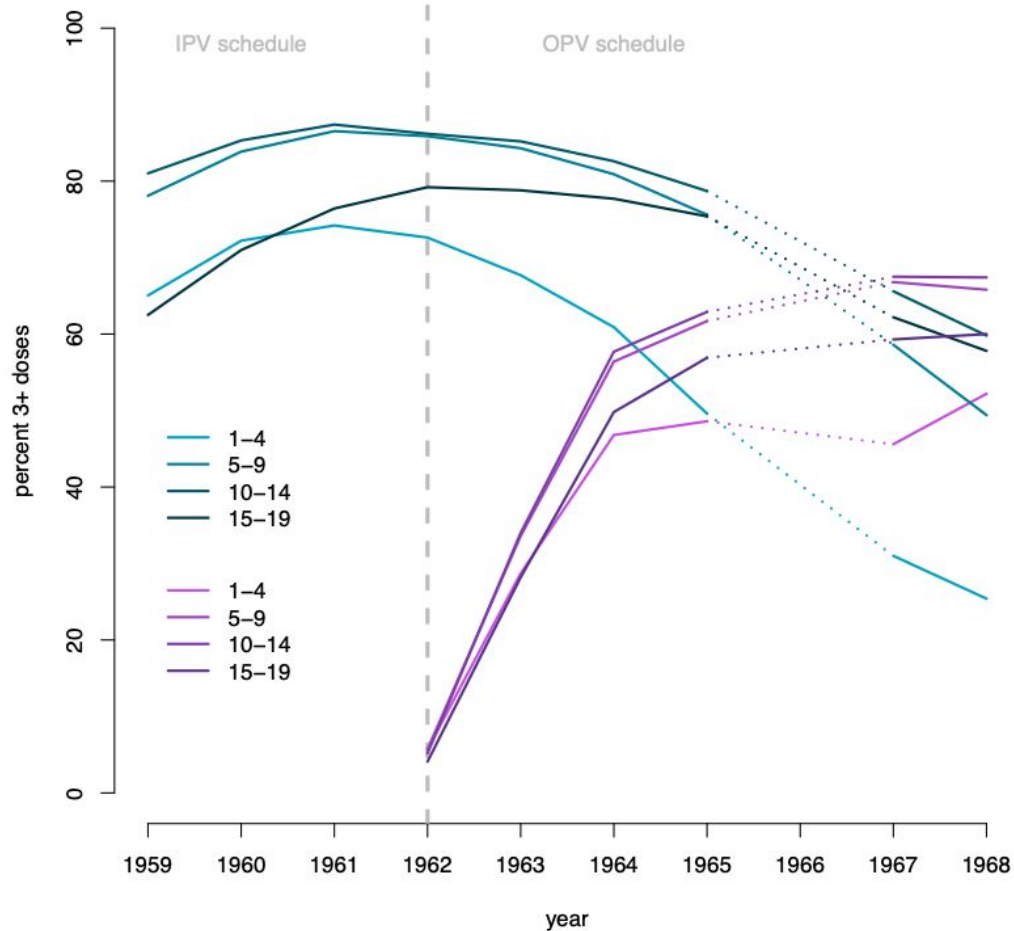
VACCINATION UPTAKE

IPV was initially studied and rolled-out in 2nd and 3rd graders before expanding across age groups and eventually becoming an infant vaccine
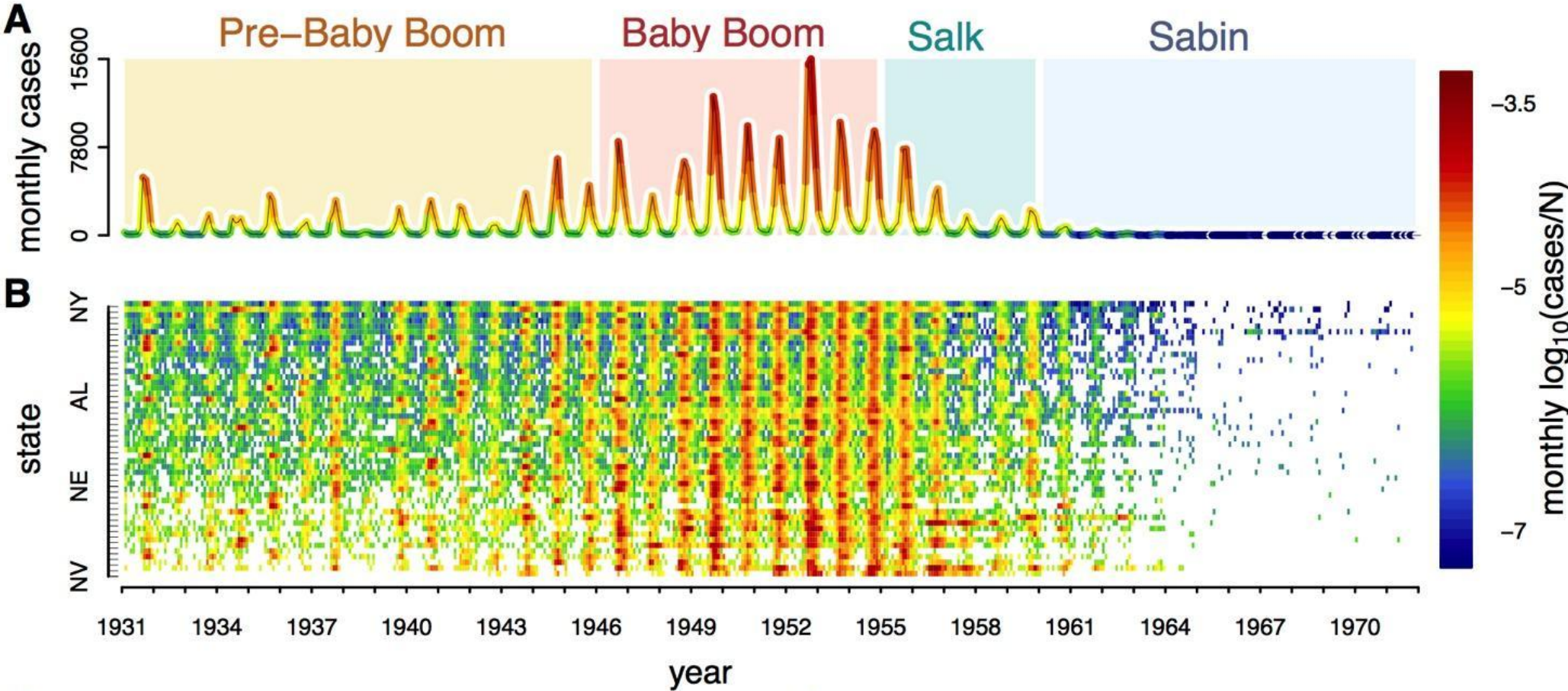
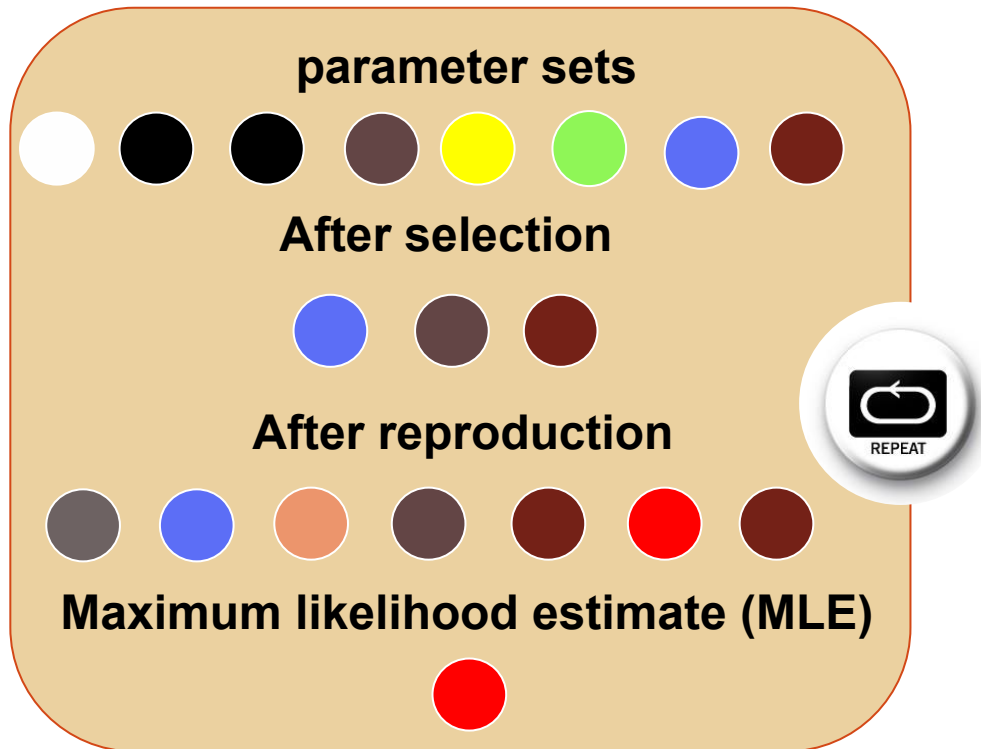# VACCINATION UPTAKE
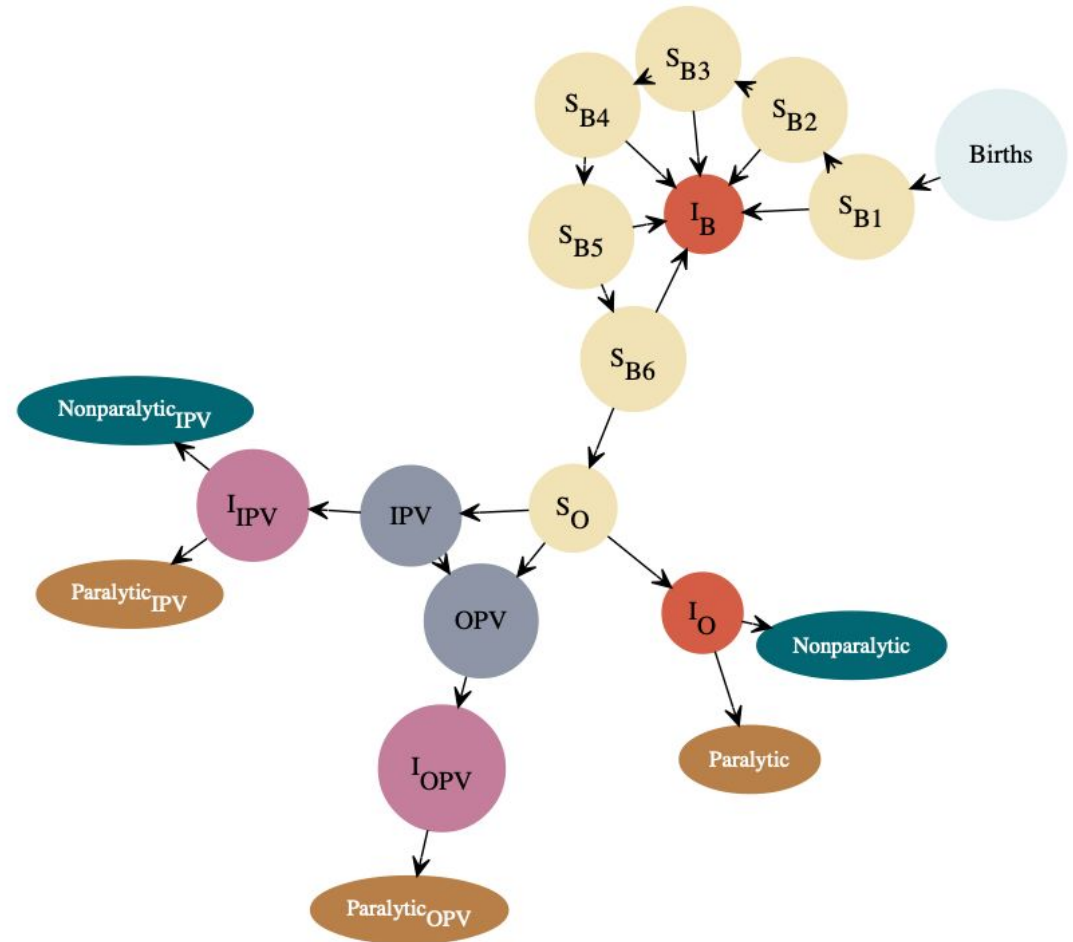
MODEL IS FIT TO CASE DATA

MAXIMUM LIKELIHOOD VIA ITERATED PARTICLE FILTERING (MIF)

Natural Selection on Parameter Sets

parameter sets

After selection

After reproduction

REPEAT

Maximum likelihood estimate (MLE)

- MIF is an algorithm that can be used to search parameter space to maximize likelihood

- The *"fitness"* of each parameter set is the likelihood of the data given the model with that parameter set, L(Θ) = p(y|Θ)
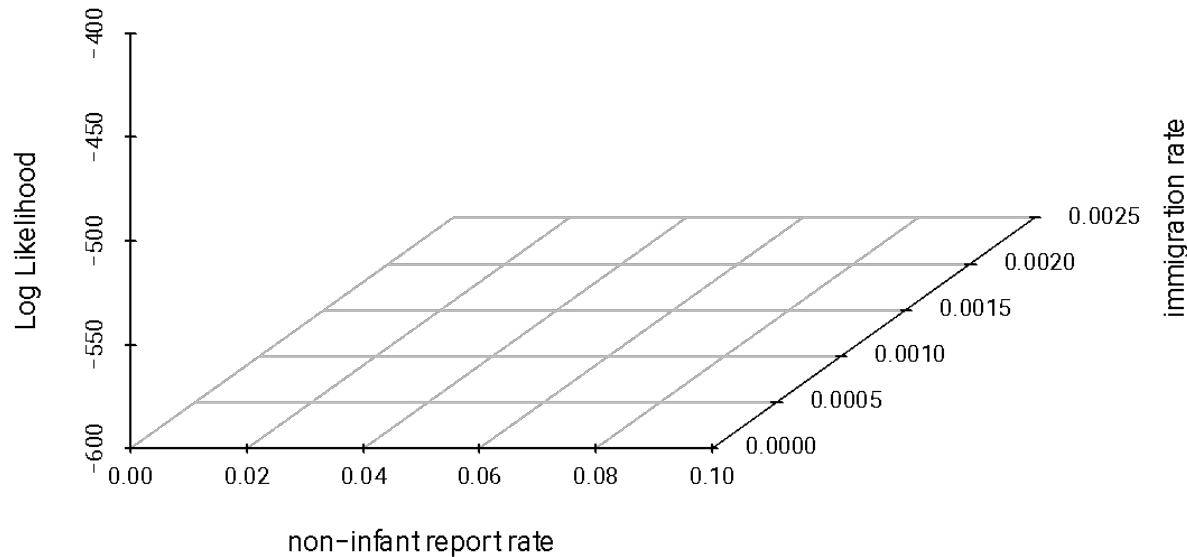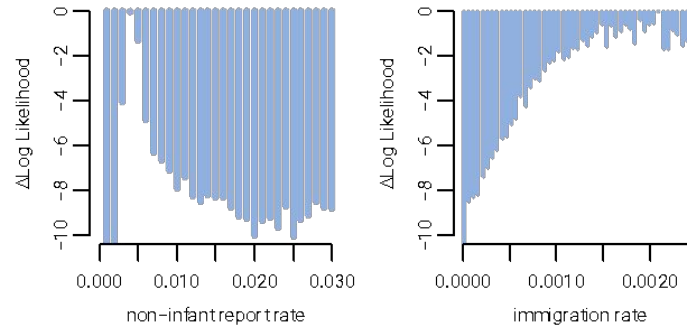
Likelihood Estimation

- Under such conditions, Maximum Likelihood Estimate, MLE, is simply parameter set with smallest deviation from data

- Equivalent to using least square errors, to decide on goodness of fit

  - Least Squares Statistic = SSE = $\Sigma(D_i - M_i)^2$

- Then, minimize SSE to arrive at MLE

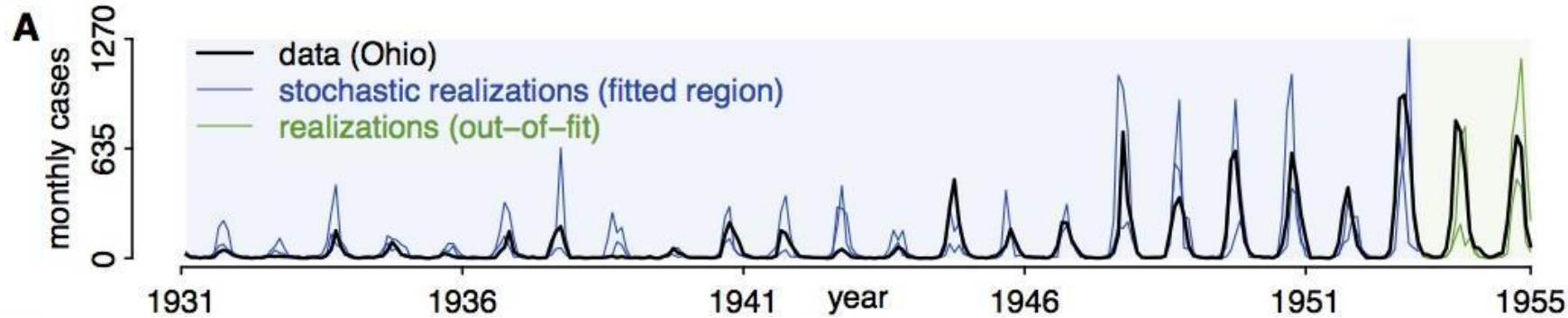# MAXIMUM LIKELIHOOD VIA ITERATED PARTICLE FILTERING (MIF)

District of Columbia

- MIF is an algorithm that can be used to search parameter space to maximize likelihood

- The *"fitness"* of each parameter set is the likelihood of the data given the model with that parameter set, L(y|Θ)
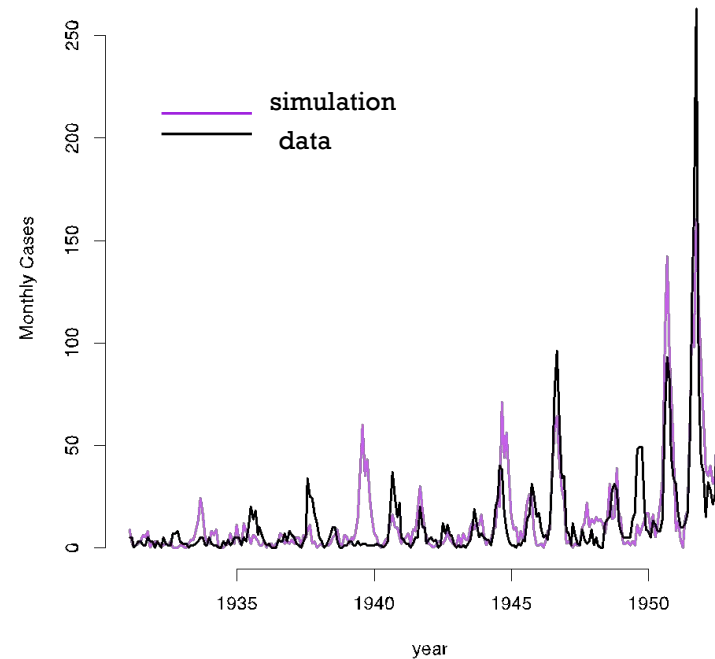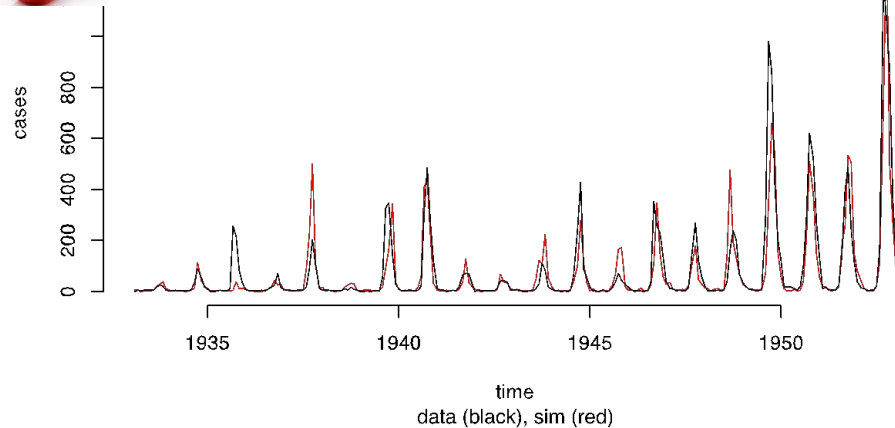
SIMULATING FROM THE MLE

The data we observe is a single realization of all the possible trajectories of a stochastic process.

Louisiana MLE pv6 pV, LogLik = −729

We can obtain the observed dynamics by simulating forward from initial conditions
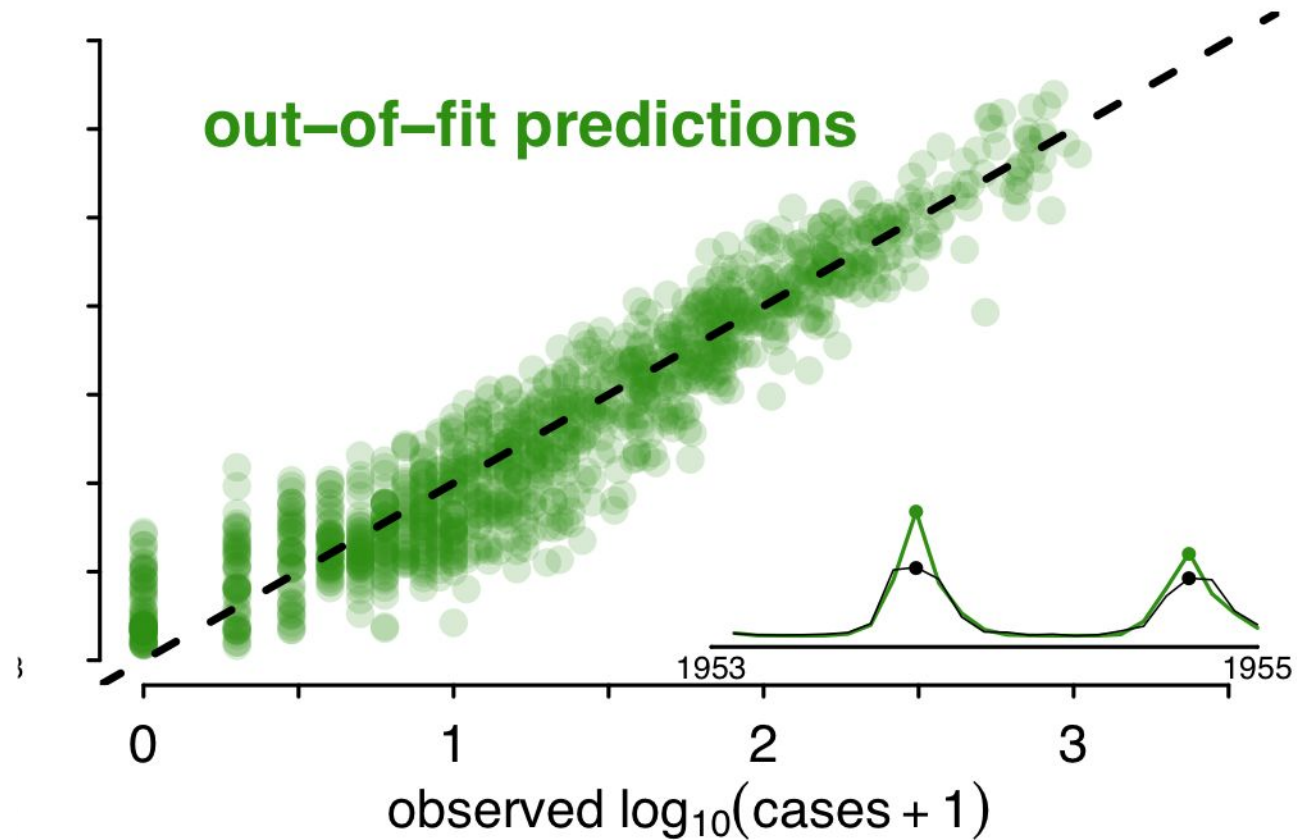

Michigan

data (black), sim (red)
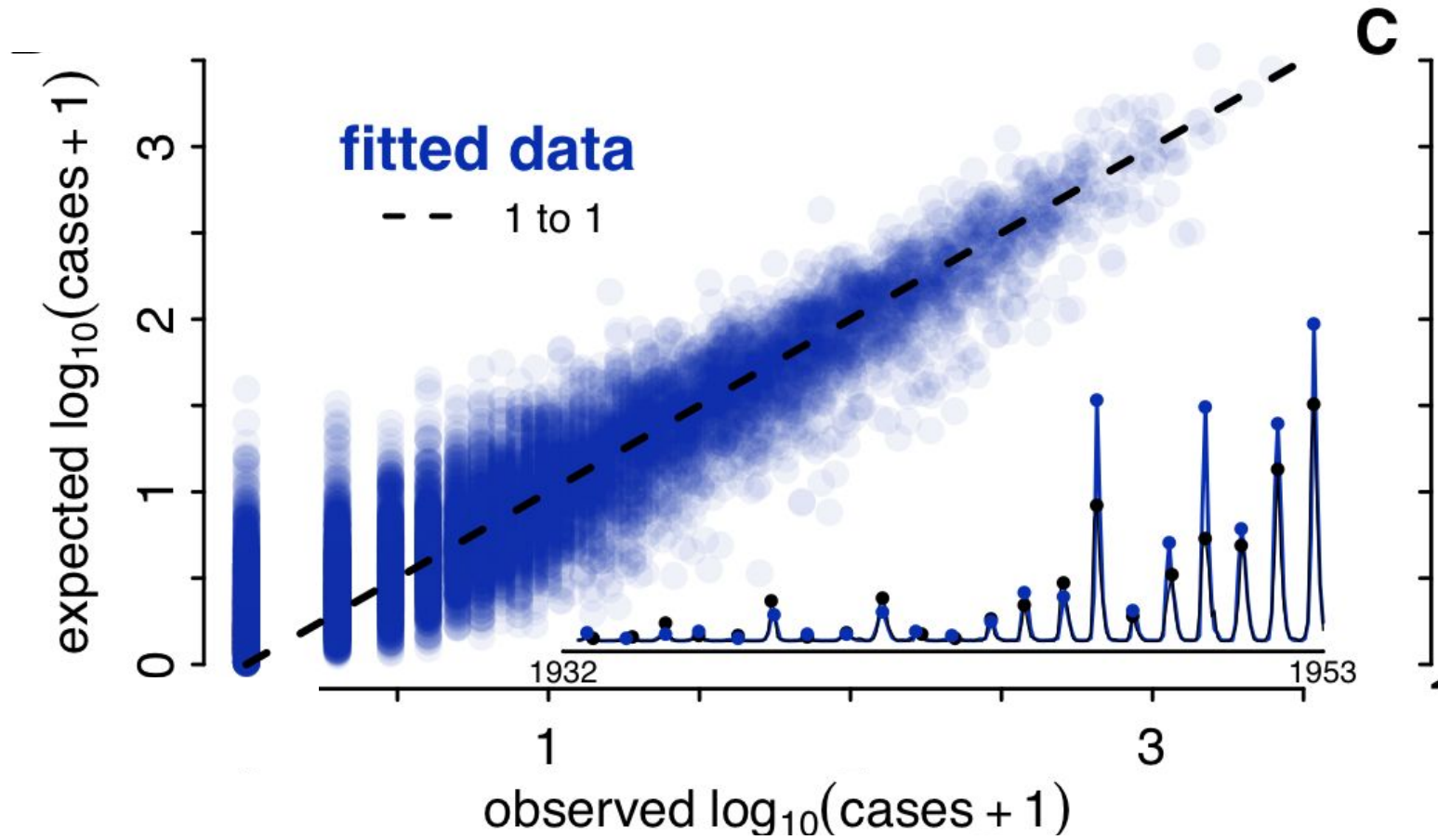
MODEL VALIDATION

MODEL VALIDATION

One-step-ahead predictions predict cases at time t using:

Model + MLE + cases$_0$+…+ cases$_{t-1}$

VACCINATION UPTAKE

(a) US

pre–vaccine — IPV

— December–to–December predictions
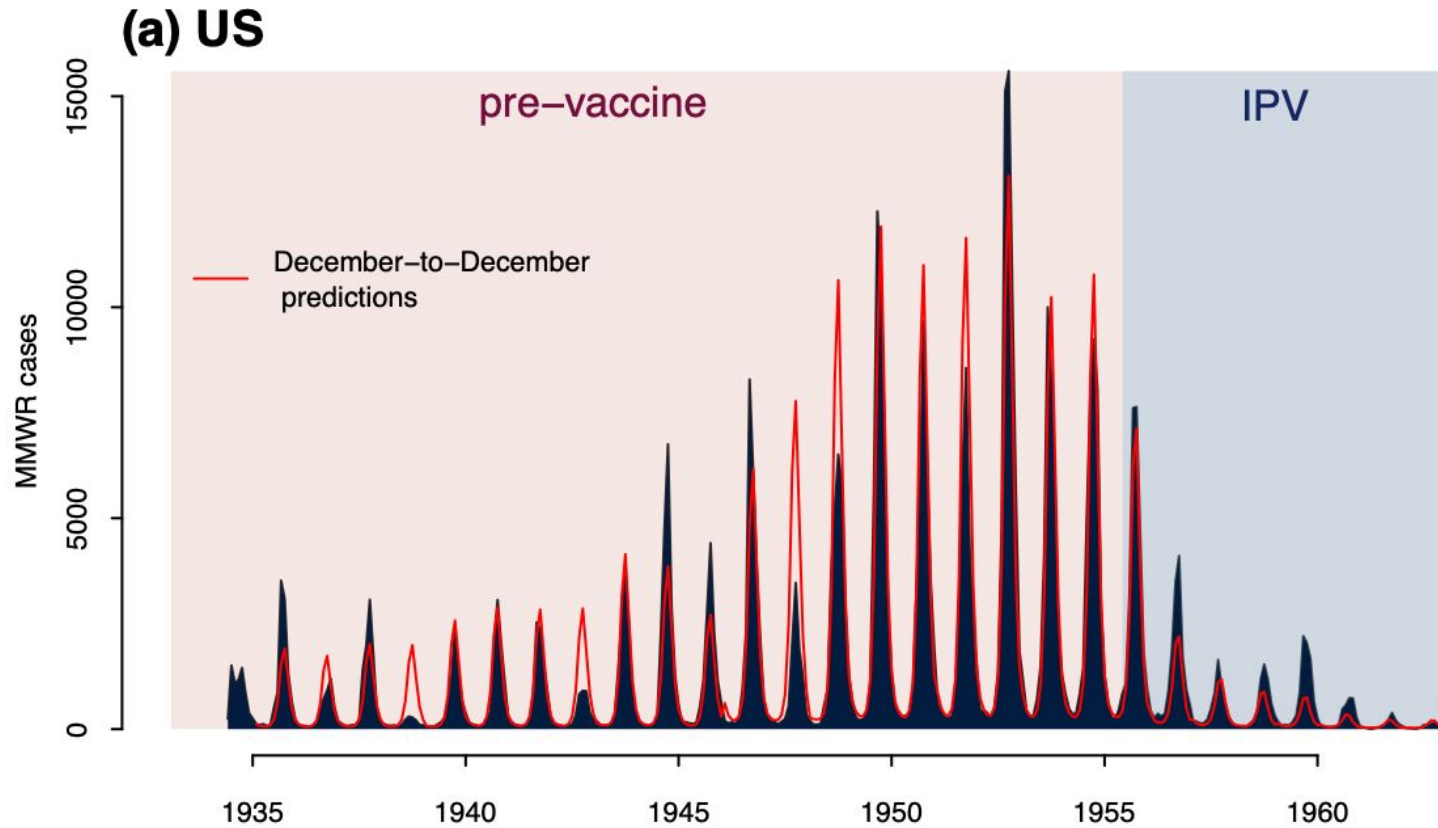
By having high quality vaccine uptake covariate data for IPV, we were able to infer that it reduces transmission by 69%

ME Martinez (Dissertation)

OTHER COVARIATES

- Report rate (reporting system, healthcare infrastructure, etc.)
- Treatment
- Interventions (e.g., bed nets, social distancing, quarantine, etc.)