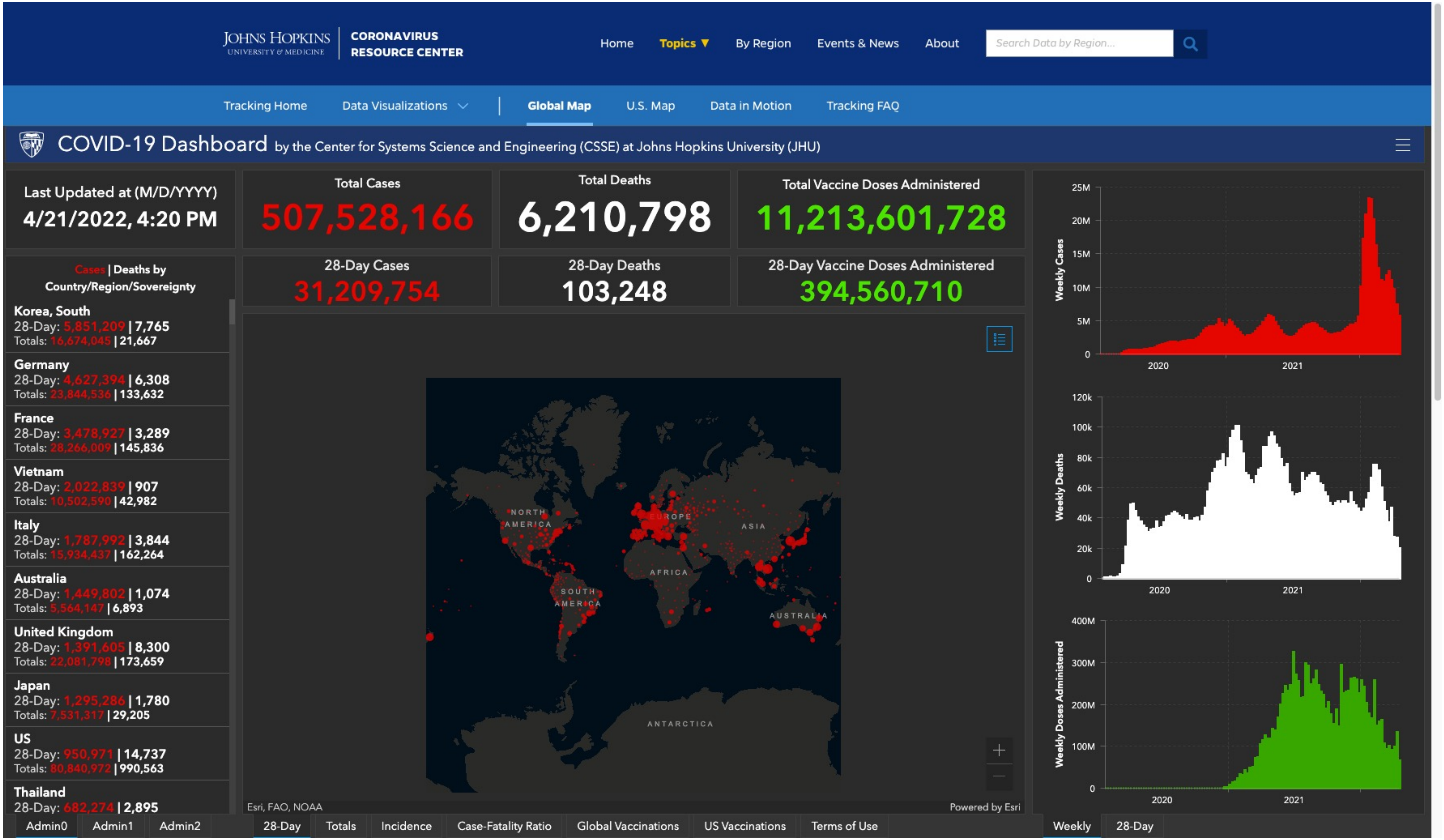
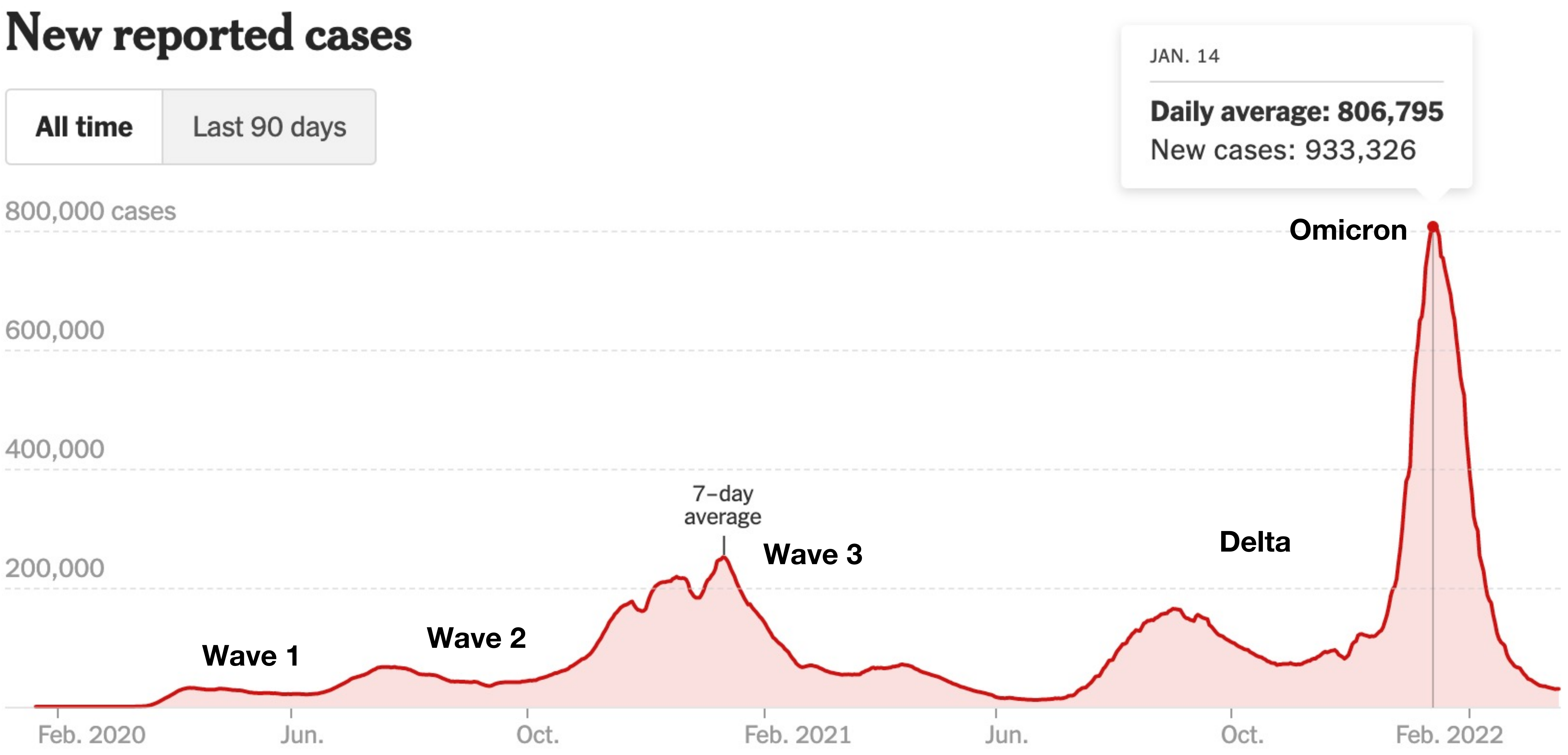
Prediction of unobserved SARS-CoV-2 Infection

Carson Chow, Alec Schwartz
Mathematical Biology Section, LBM, NIDDK, NIH

Shashaank Vattikuti, NIH/WRAIR, Joshua Chang, NIH, Richard Gerkin, ASU

Kaitlyn Sadtler, NIBIB, NIH


Johns Hopkins University & Medicine
CORONAVIRUS RESOURCE CENTER
Home
Topics ▼
By Region
Events & News
About
Search Data by Region...
Tracking Home
Data Visualizations
Global Map
U.S. Map
Data in Motion
Tracking FAQ
COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)
Last Updated at (M/D/YYYY)
4/21/2022, 4:20 PM
Total Cases
507,528,166
Total Deaths
6,210,798
Total Vaccine Doses Administered
11,213,601,728
28-Day Cases
31,209,754
28-Day Deaths
103,248
28-Day Vaccine Doses Administered
394,560,710
Cases | Deaths by Country/Region/Sovereignty
Korea, South
28-Day: 5,851,209 | 7,765
Totals: 16,674,045 | 21,667
Germany
28-Day: 4,627,394 | 6,308
Totals: 23,844,536 | 133,632
France
28-Day: 3,478,927 | 3,289
Totals: 28,266,009 | 145,836
Vietnam
28-Day: 2,022,839 | 907
Totals: 10,502,590 | 42,982
Italy
28-Day: 1,787,992 | 3,844
Totals: 15,934,437 | 162,264
Australia
28-Day: 1,449,802 | 1,074
Totals: 5,564,147 | 6,893
United Kingdom
28-Day: 1,391,605 | 8,300
Totals: 22,081,798 | 173,659
Japan
28-Day: 1,295,286 | 1,780
Totals: 7,531,317 | 29,205
US
28-Day: 950,971 | 14,737
Totals: 80,840,972 | 990,563
Thailand
28-Day: 682,274 | 2,895
NORTH AMERICA
EUROPE
ASIA
AFRICA
SOUTH AMERICA
AUSTRALIA
ANTARCTICA
Esri, FAO, NOAA
Powered by Esri
Weekly Cases
25M
20M
15M
10M
5M
0
2020
2021
Weekly Deaths
120k
100k
80k
60k
40k
20k
0
2020
2021
Weekly Doses Administered
400M
300M
200M
100M
0
2020
2021
Admin0
Admin1
Admin2
28-Day
Totals
Incidence
Case-Fatality Ratio
Global Vaccinations
US Vaccinations
Terms of Use
Weekly
28-Day



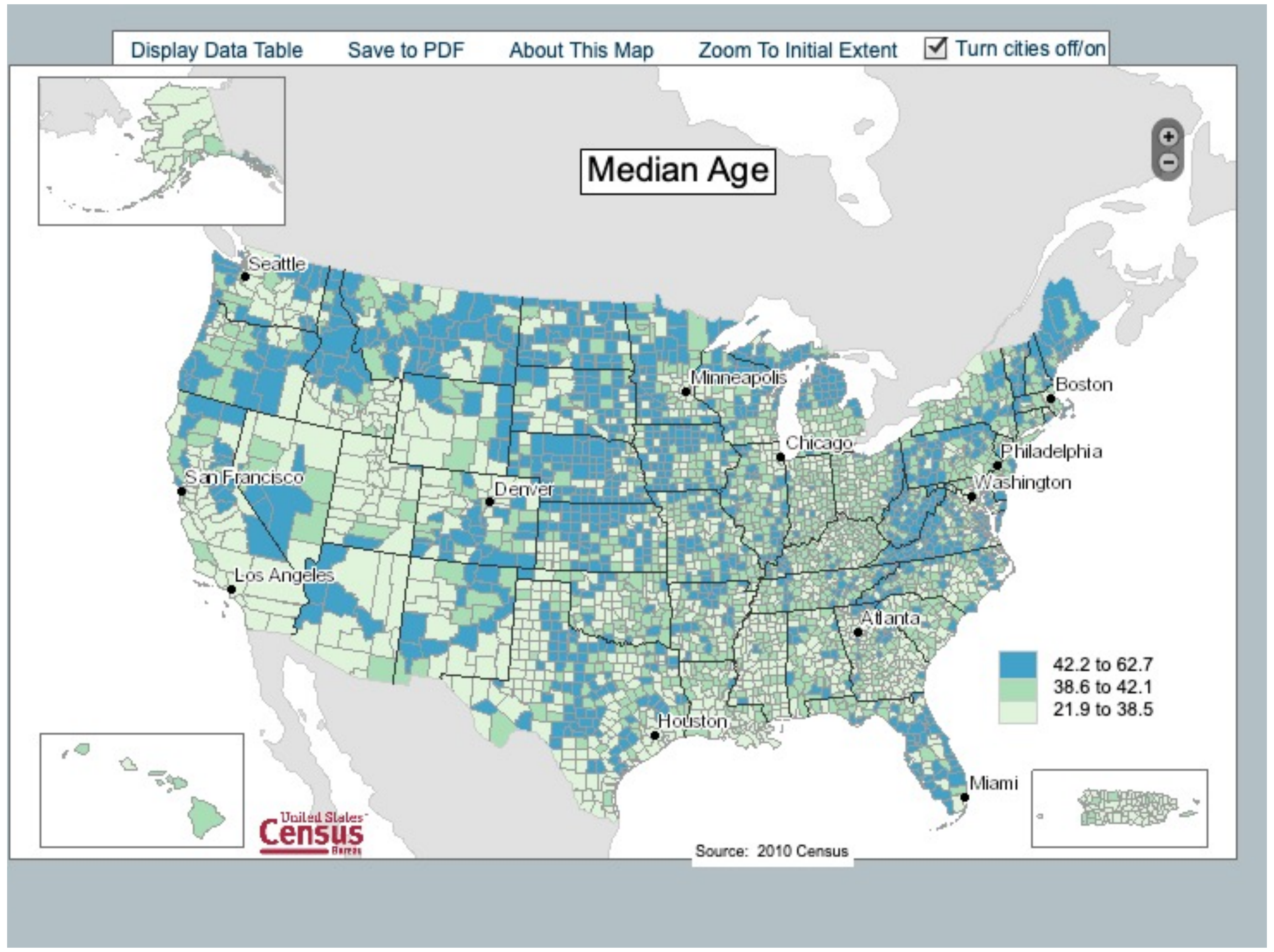
Source: NYTimes, https://www.nytimes.com/interactive/2021/us/covid-cases.html

But how many unreported?

Total in US ~ 80 million (Apr 21)

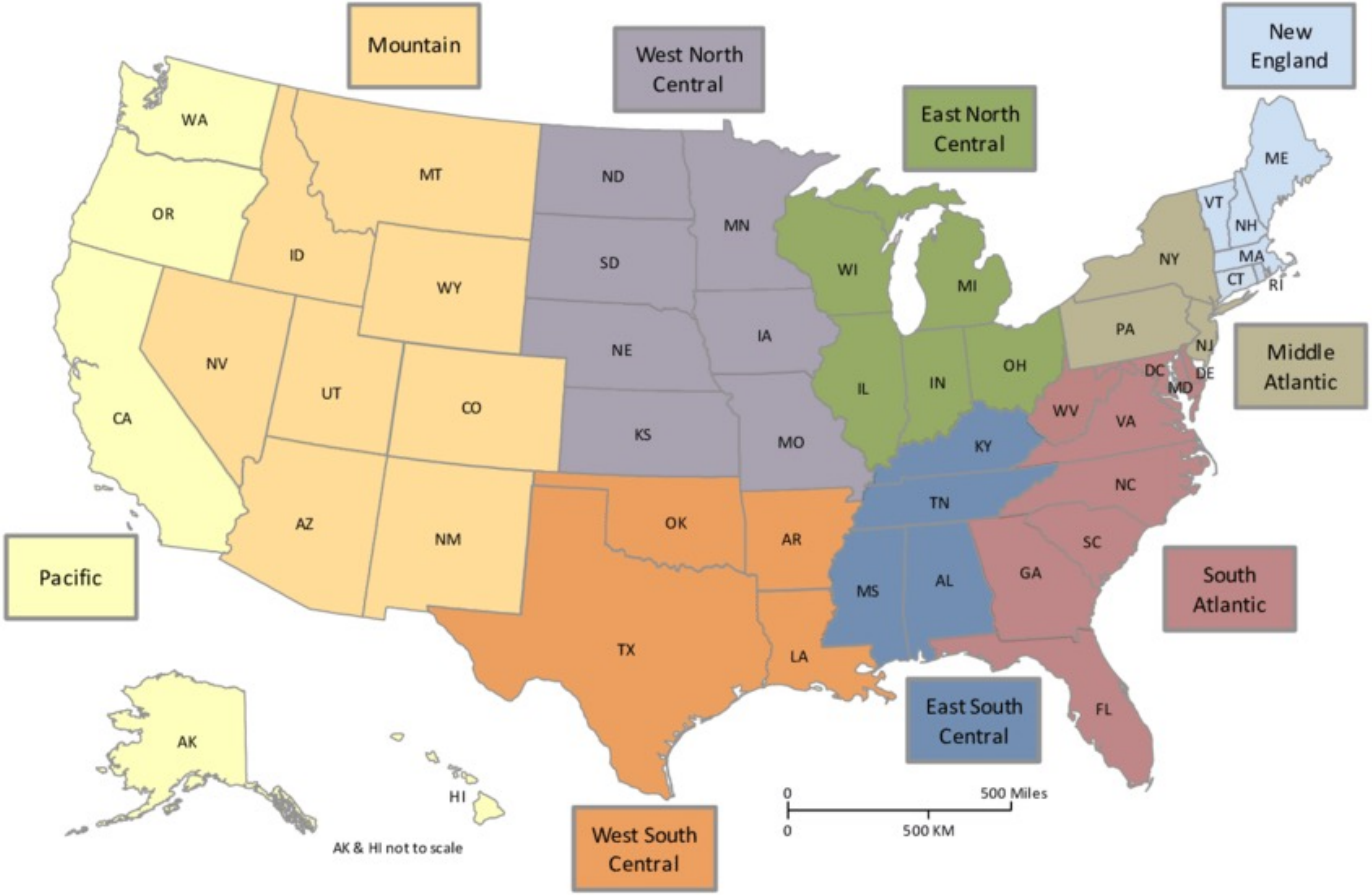
NIH Serosurvey (Kaitlyn Sadtler,NIBIB)

Neoteryx Mitra Volumetric Blood Sampling Kit
(at-home dried blood)



Quota-Based Sampling from large volunteer pool
(Goal: n = 10,000)

Participating IC’s & Sites: NIBIB, NCATS, NIAID, FNLCR, UPitt, UAB


Mountain
West North Central
East North Central
New England
Middle Atlantic
South Atlantic
East South Central
West South Central
Pacific
WA
OR
CA
ID
MT
WY
NV
UT
CO
AZ
NM
ND
SD
NE
KS
MN
IA
MO
WI
MI
IL
IN
OH
ME
VT
NH
MA
CT
RI
NY
PA
NJ
DE
MD
DC
WV
VA
NC
SC
GA
FL
KY
TN
MS
AL
OK
TX
AR
LA
AK
HI
AK & HI not to scale
0
500 Miles
0
500 KM

Can’t assay everyone all the time

Extrapolate with a model

Modeling a pandemic is hard

Biology is complicated

Data - incomplete, misreporting, changing standards

Network/spatial effects - not homogeneous

Population dependence, e.g. age, comorbidities, genetics, history

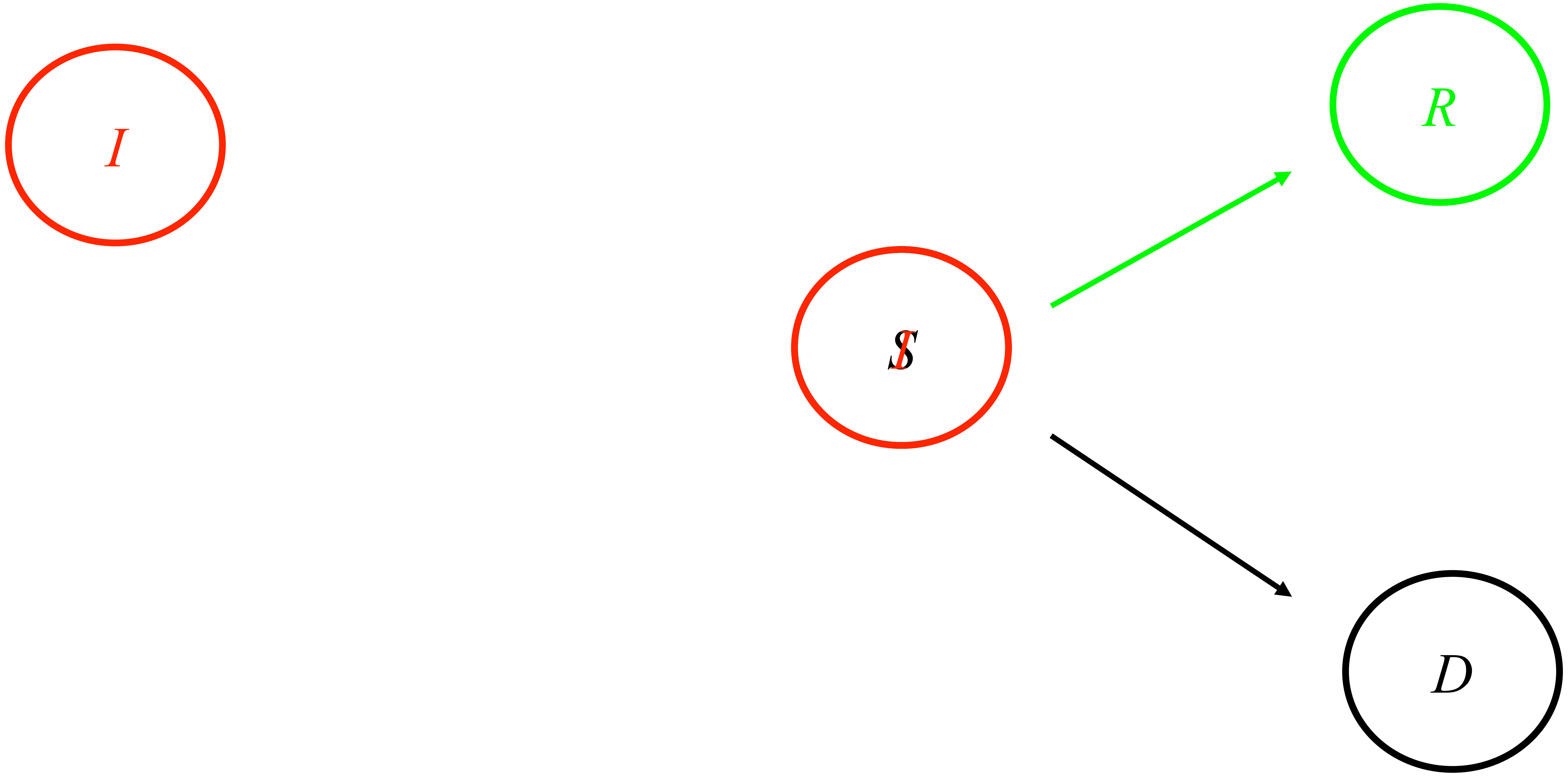
Physiology - duration and severity

Behavior and policy

What should a model include?

Minimal model

SIR(D)



Network dependent birth-death stochastic process

Mass Action SIR

$$\frac{dI}{dt} = \frac{\beta}{N}SI - \sigma I$$

$$\frac{dS}{dt} = -\frac{\beta}{N}SI$$

$$\frac{dR}{dt} = \sigma_R I$$

$$\frac{dD}{dt} = \sigma_D I$$

$$\sigma = \sigma_R + \sigma_D$$

Mass Action SIR

$$\frac{dI}{dt} = \beta SI - \sigma I \qquad \frac{dS}{dt} = -\beta SI$$

Scale by N

I peaks at

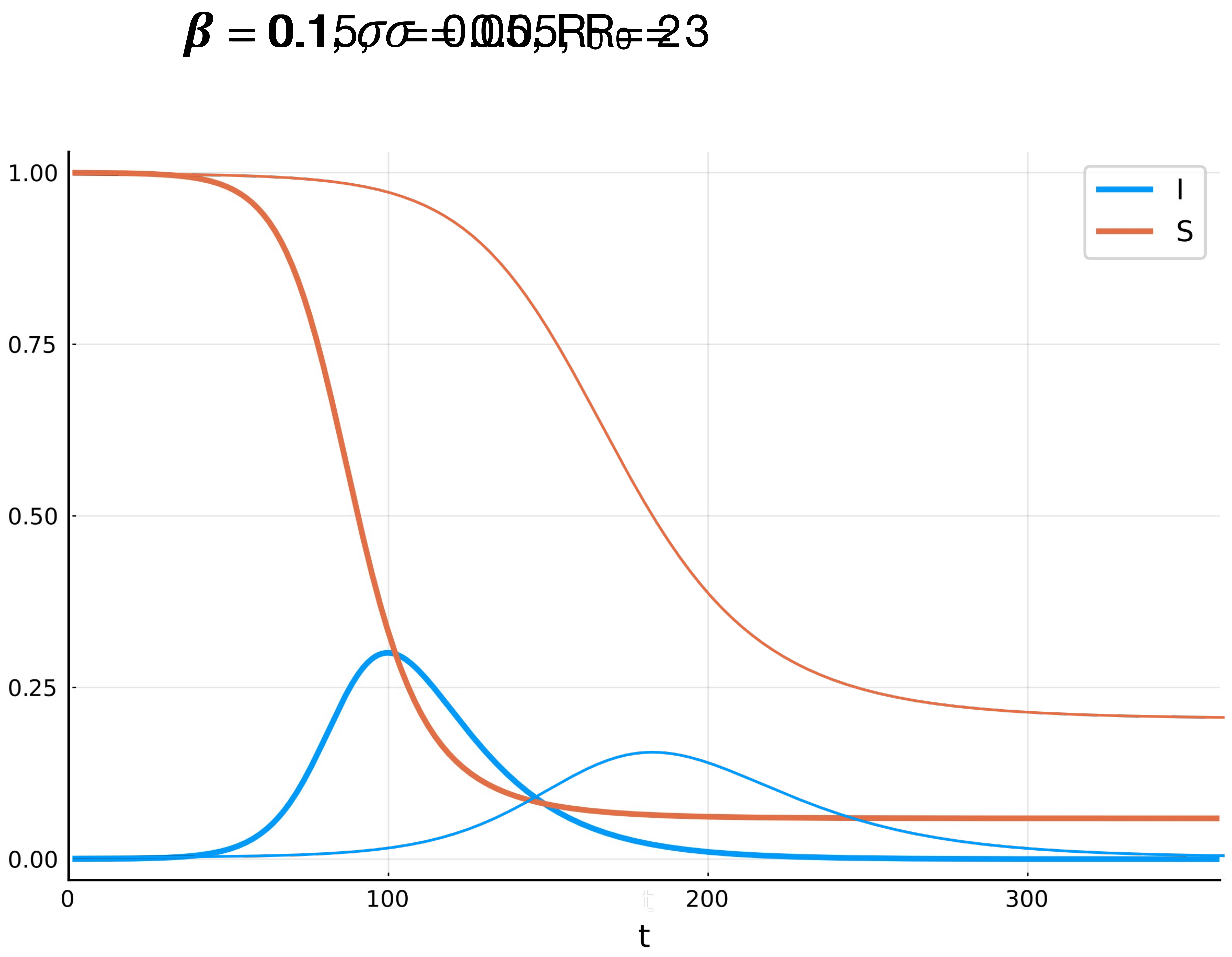
(I nullcline)

$$\frac{dI}{dt} = 0 = \beta SI - \sigma I$$

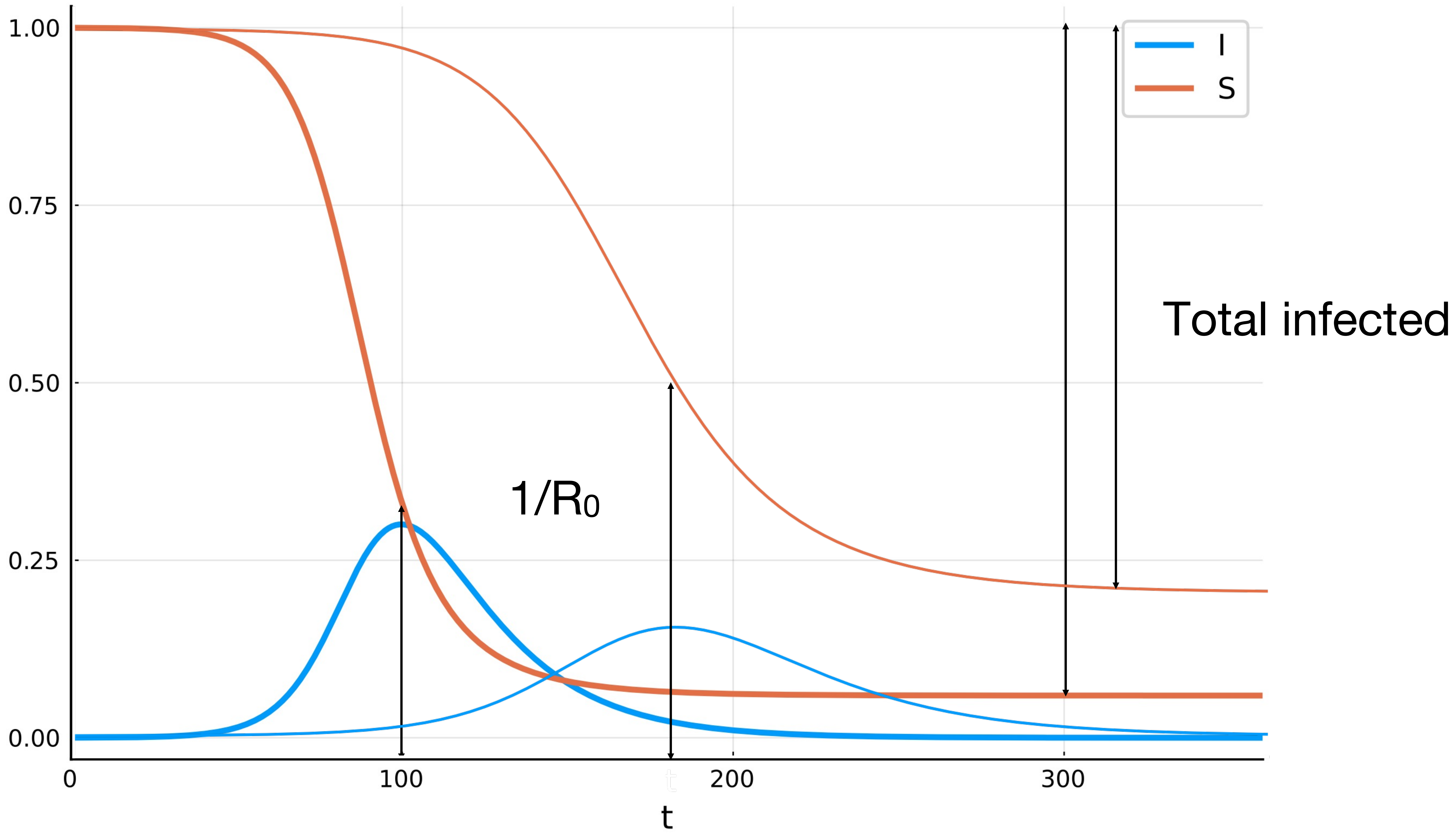
$$S^* = \frac{\sigma}{\beta} = \frac{1}{R_0}$$

Herd immunity threshold

R_0 = initial reproduction number

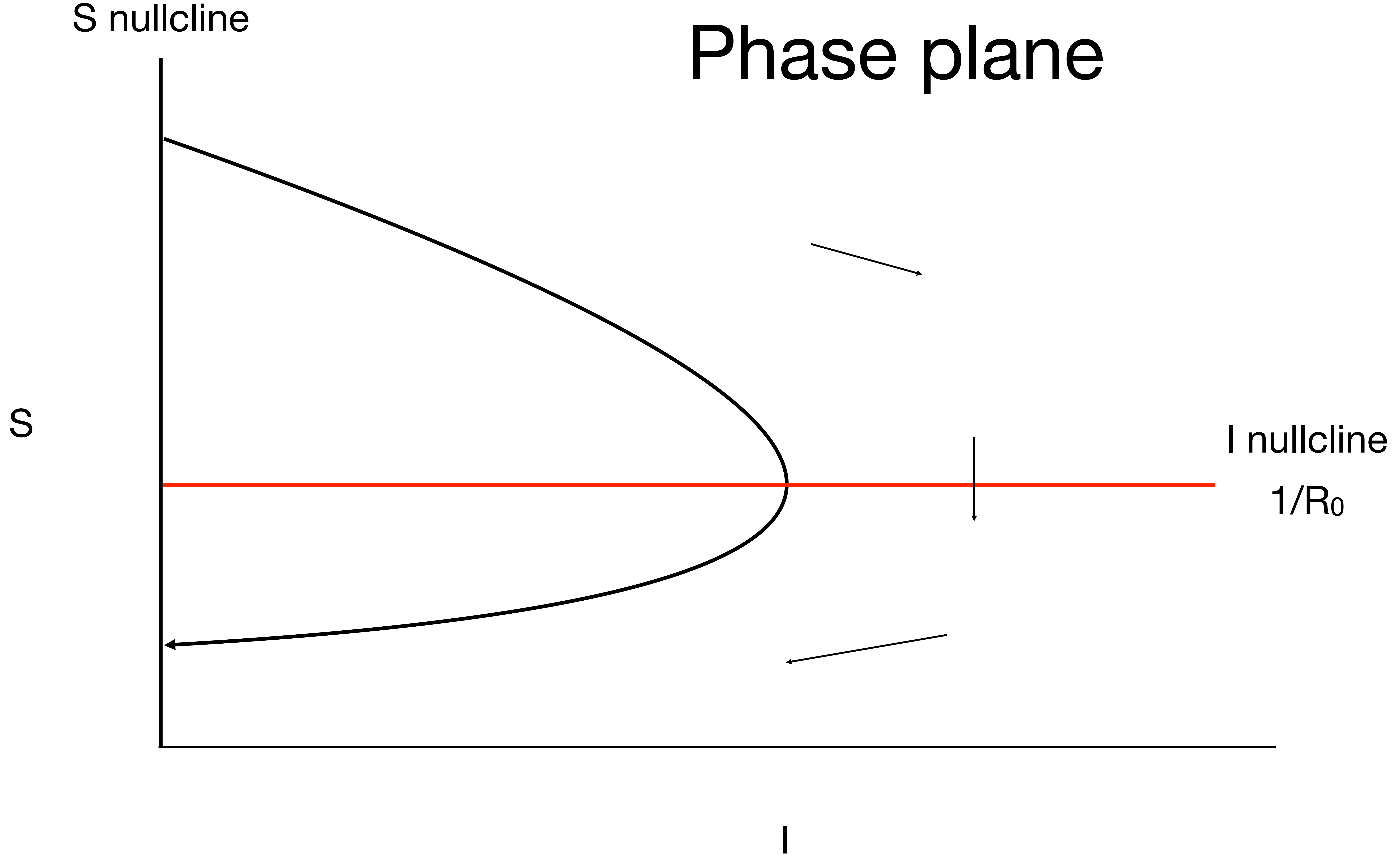


β = 0.15, σσ = 0.05, RR0 = 23
1.00
0.75
0.50
0.25
0.00
0
100
200
300
t
I
S



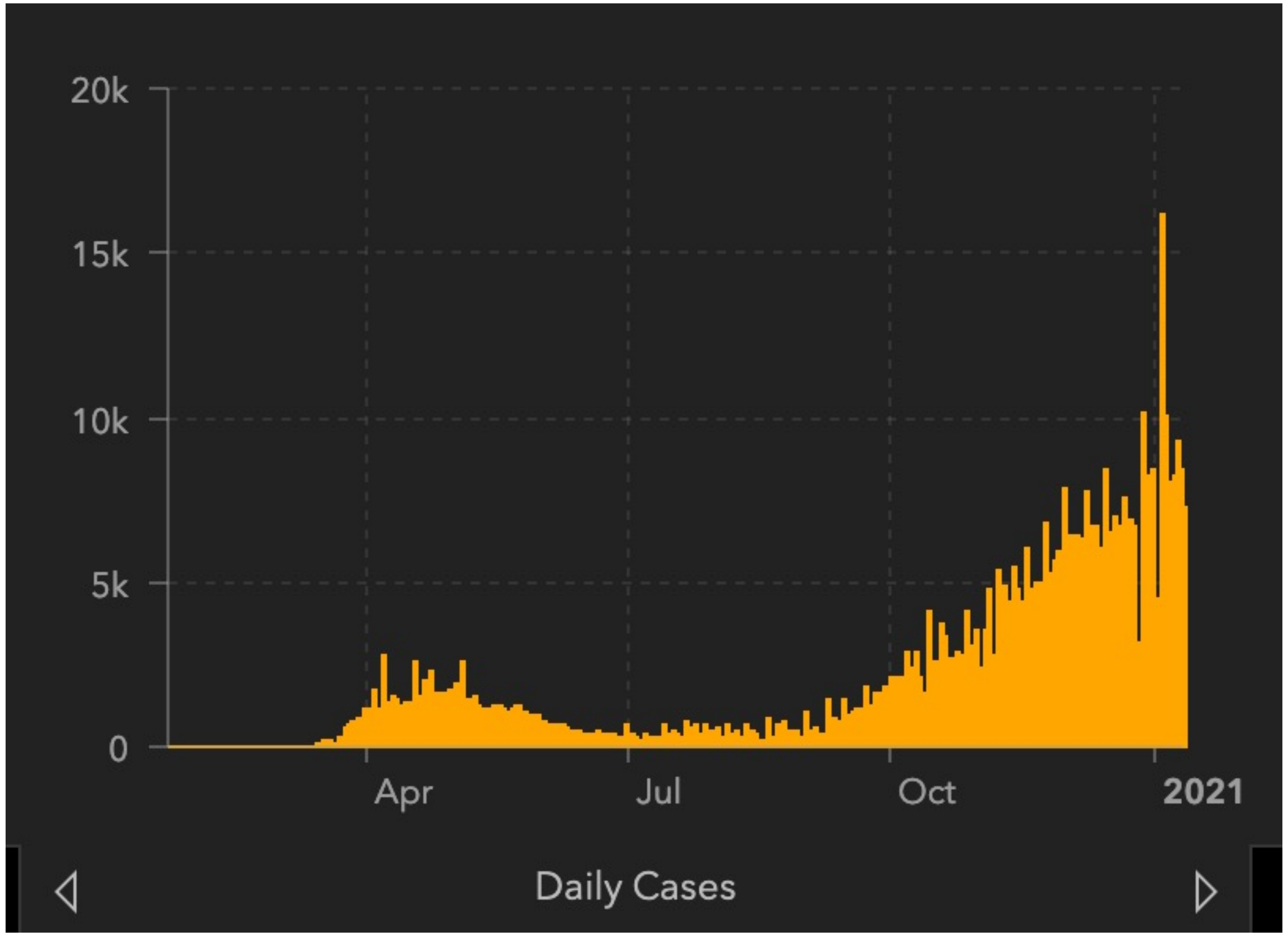
1.00
0.75
0.50
0.25
0.00
0
100
200
300
t
I
S
1/R_0
Total infected

Phase plane

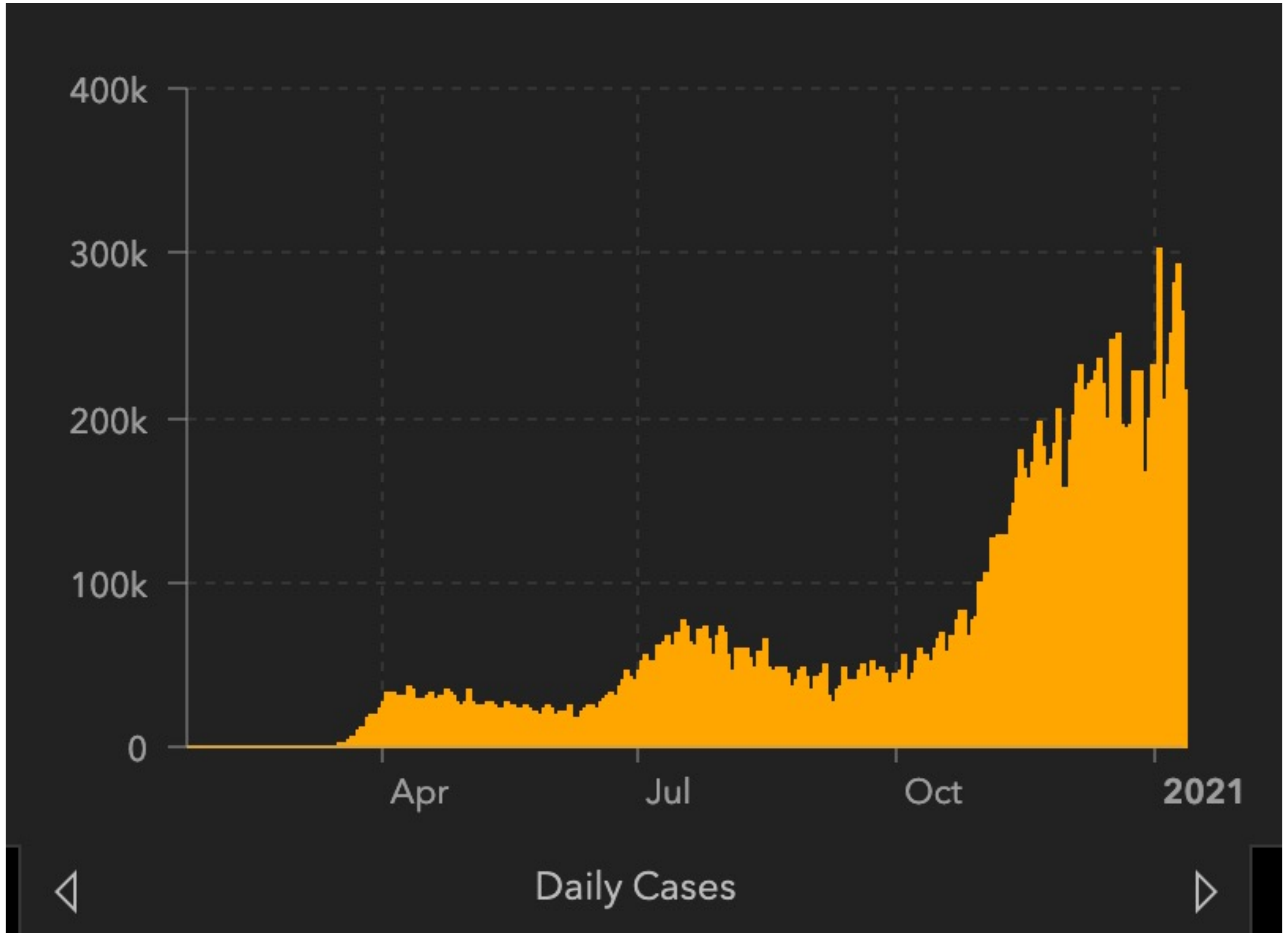


Recurrence and plateaus

Canada



USA



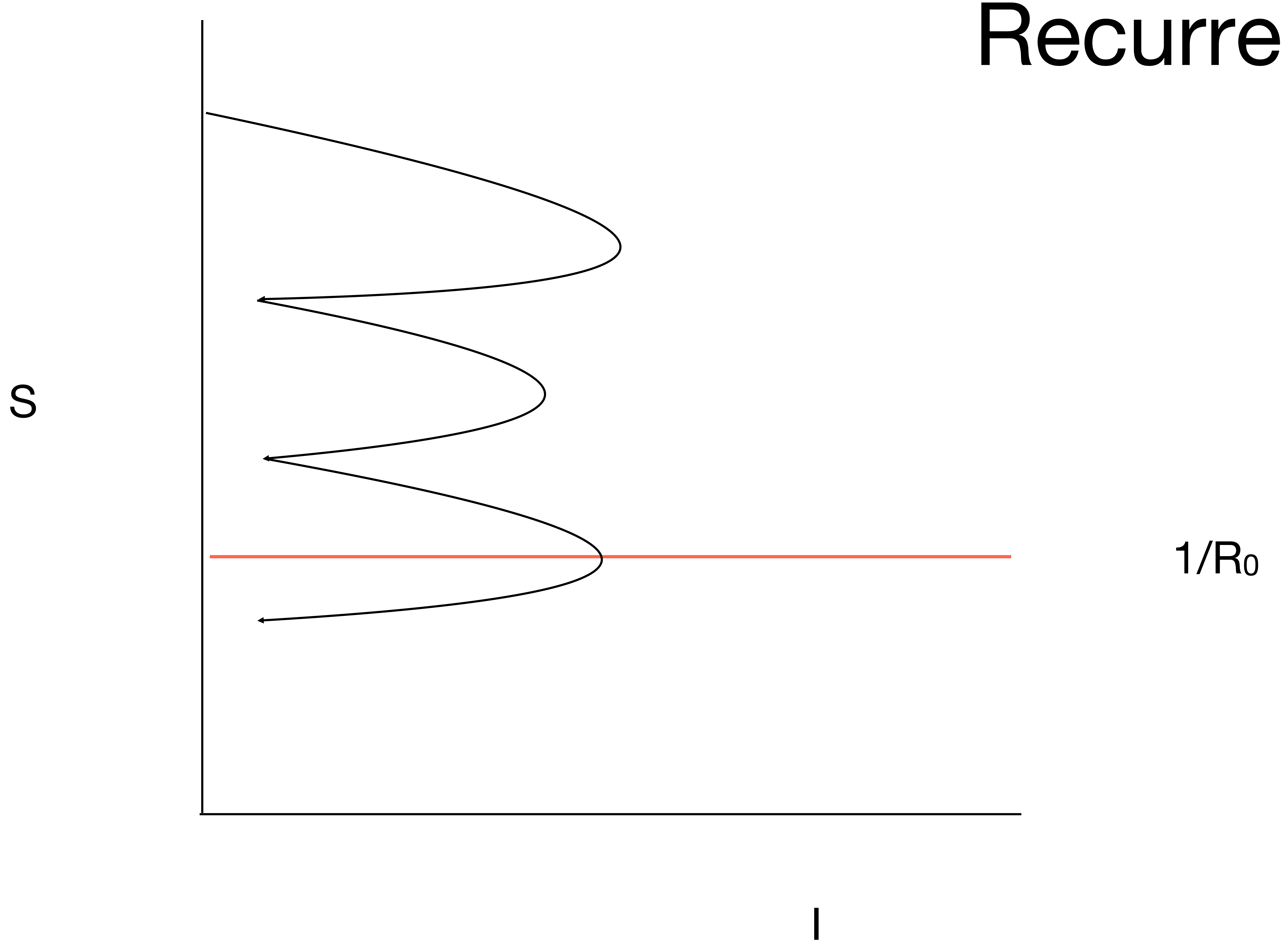
Time dependence

Policy, behavior, viral evolution, …

$$\frac{dI}{dt} = \beta(t)SI - \sigma I \qquad \frac{dS}{dt} = -\beta(t)SI$$

Move the nullcline

Recurrence



Plateaus

$$\frac{dI}{dt} = \beta(t)SI - \sigma I$$

$$\frac{dS}{dt} = -\beta(t)SI$$

$\frac{dI}{dt} = \beta(t)SI - \sigma I$=0

Unstable equilbrium

Plateaus

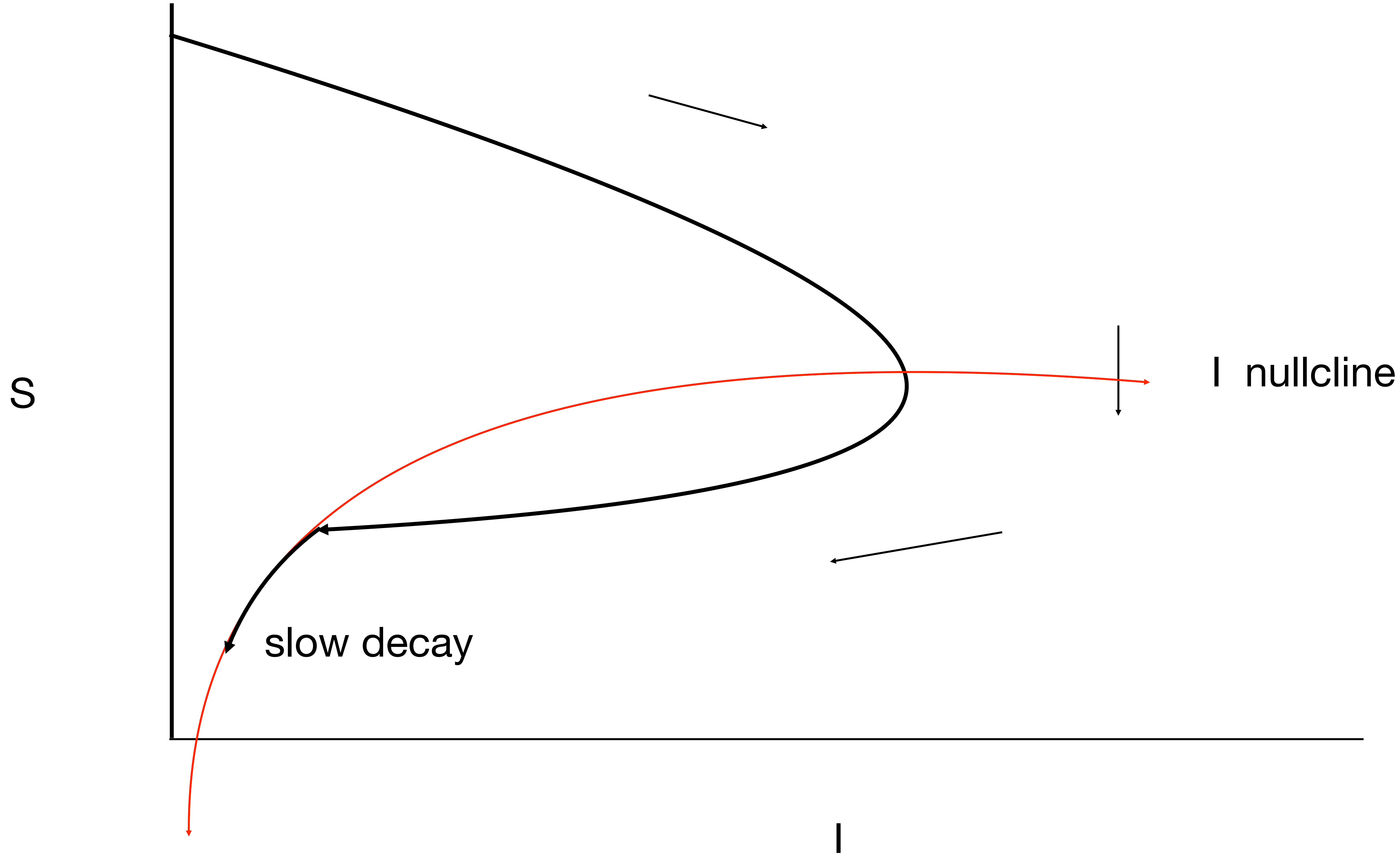
migration

$$\frac{dI}{dt} = \beta(t)SI - \sigma I + \alpha$$

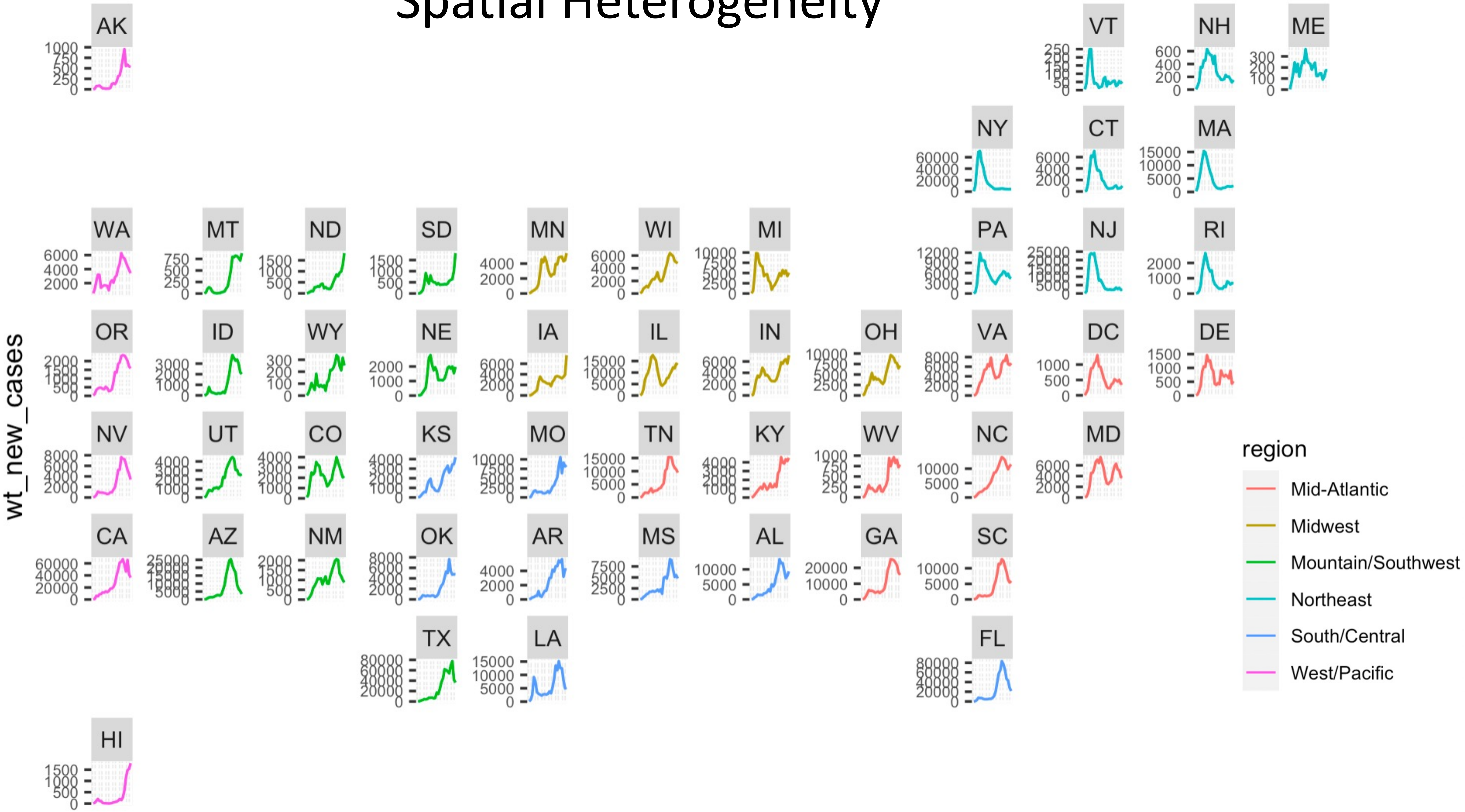
$$\frac{dS}{dt} = -\beta(t)SI$$

$$S = \frac{\sigma I - \alpha}{\beta I}$$

Quasi-stable equilibrium


S
I nullcline
slow decay
I

Spatial Heterogeneity



Weekly totals spanning from March 8, 2020 to September 4, 2020.

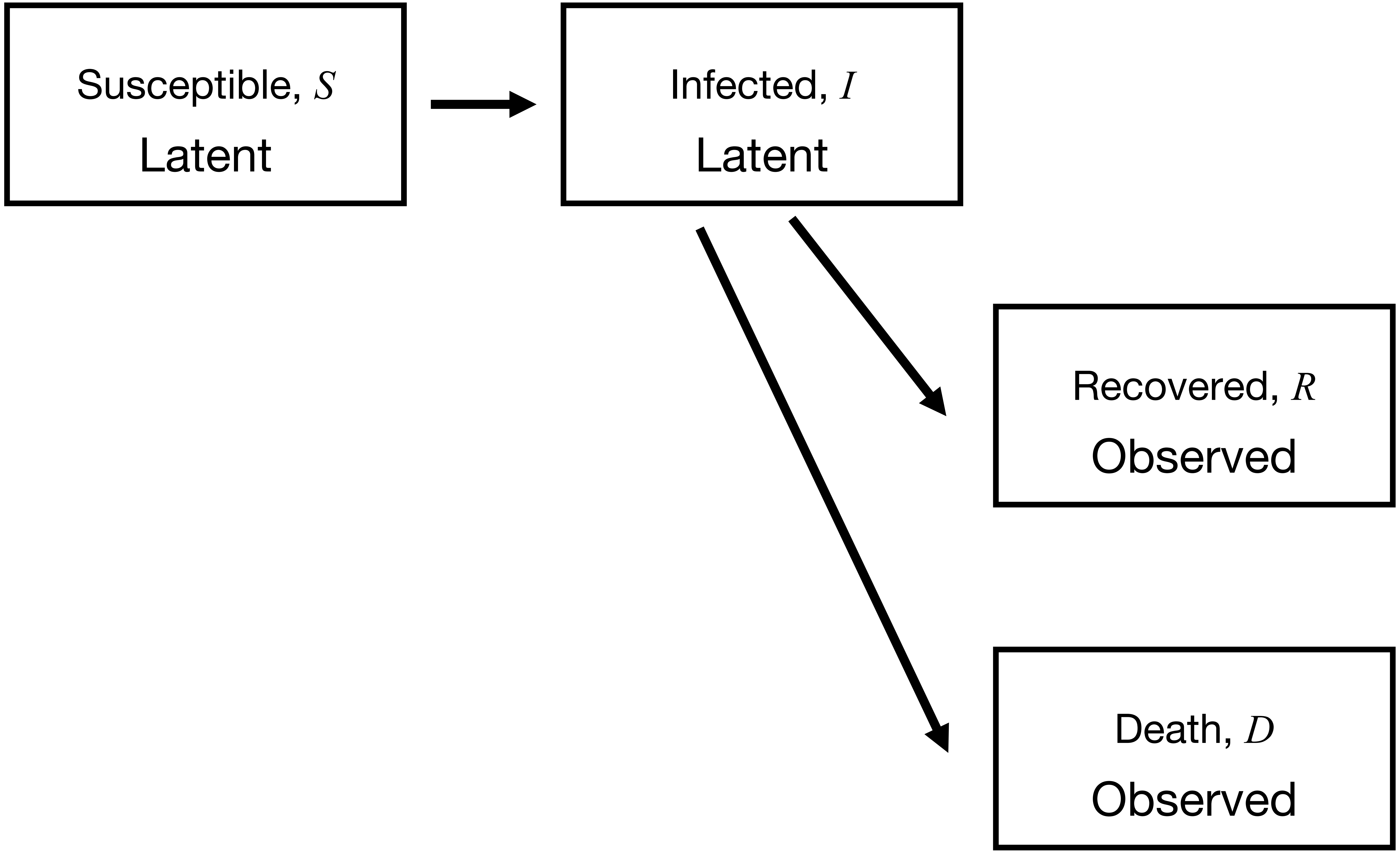
Model needs to include at least:

Time dependent infection rate, migration, spatial inhomogeneity

Reported cases are a fraction of total infected

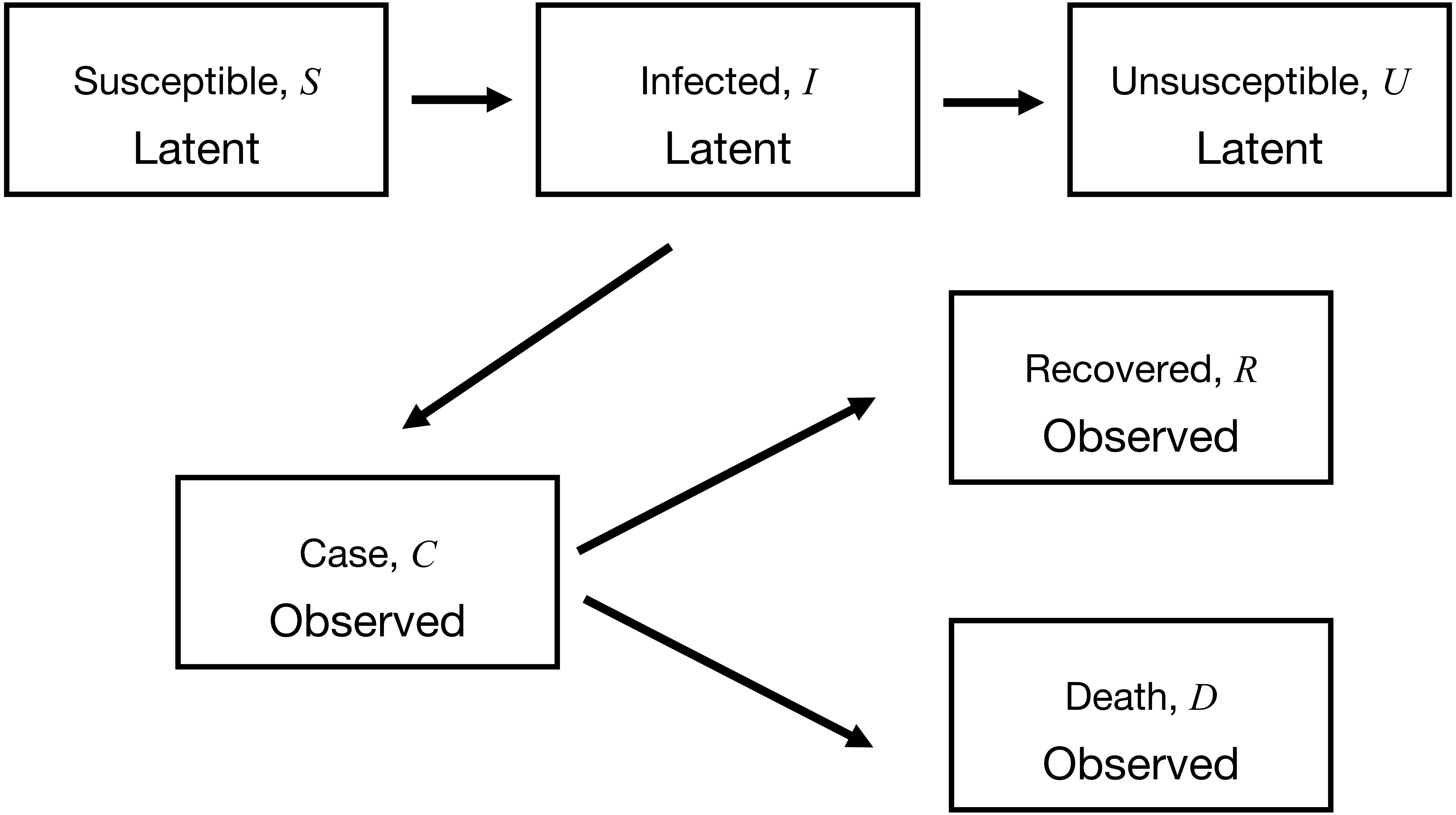
Case ascertainment ratio = total cases/total infections

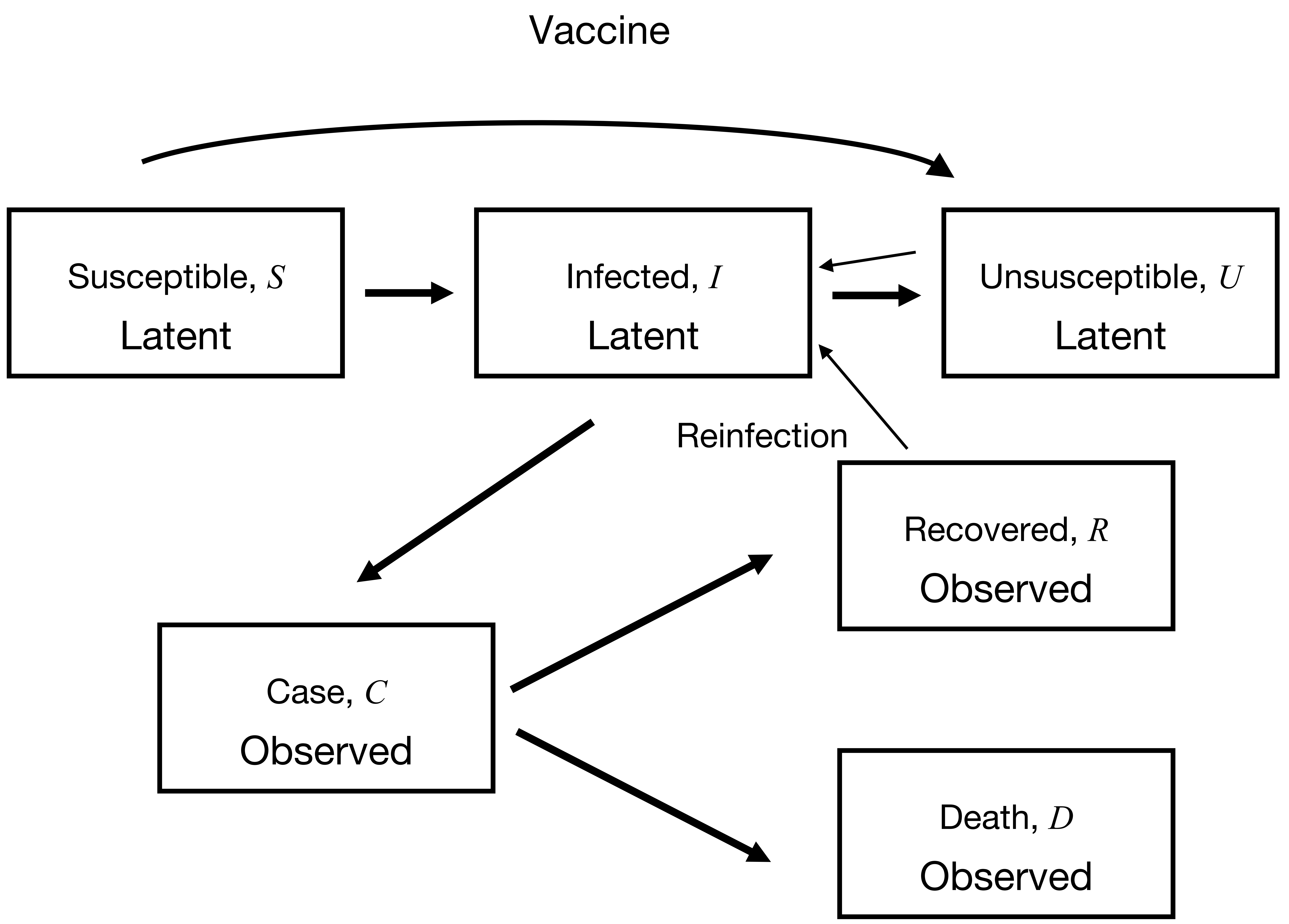
SIR model



Susceptible, S
Latent
Infected, I
Latent
Recovered, R
Observed
Death, D
Observed

Latent variable SICR model




Vaccine
Susceptible, S
Latent
Infected, I
Latent
Unsusceptible, U
Latent
Reinfection
Recovered, R
Observed
Case, C
Observed
Death, D
Observed

Measures

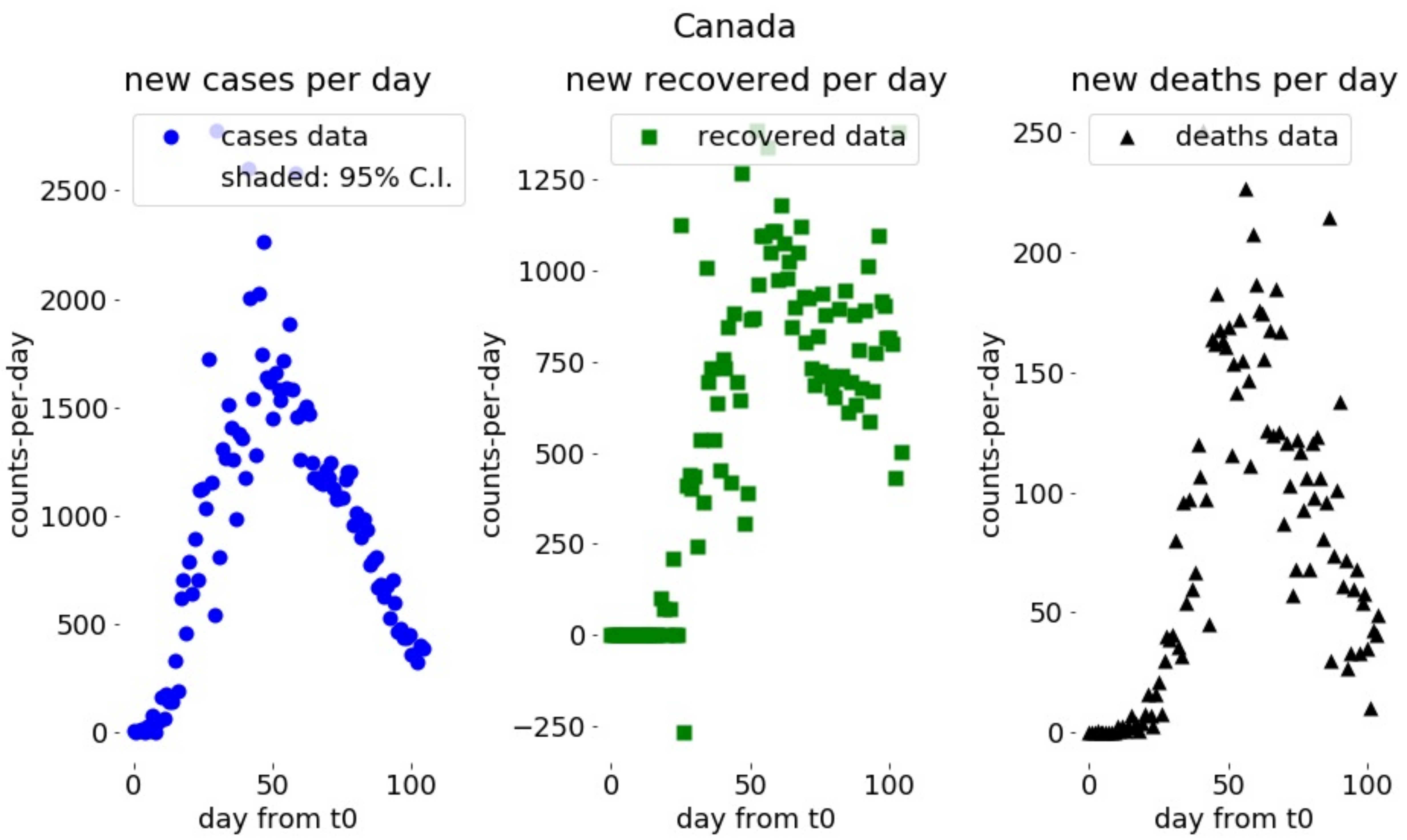
CAR_t = total cases / total infected

IR_t = total infected / total population

R_t = time dependent reproduction number

Bayesian Inference

Fitting models in the presence of data and parameter uncertainty



Bayesian Inference

Statistics by applying the rules of probability

$$P(A,B) = P(A|B)P(B) = P(B|A)P(A)$$

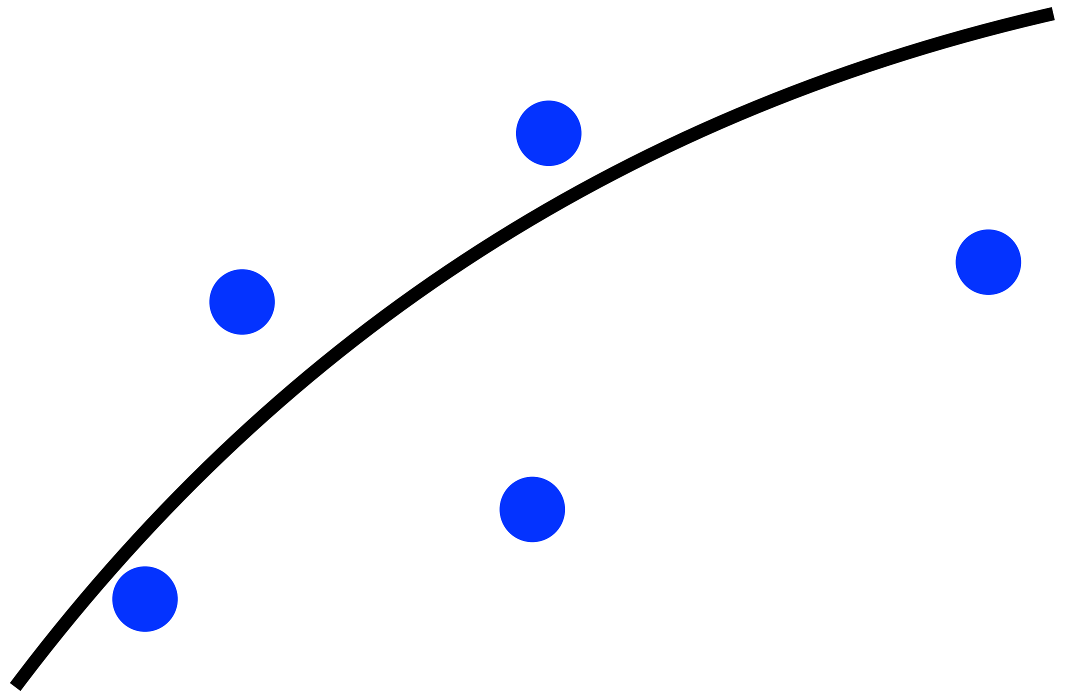
$$P(A) = \sum_B P(A|B)P(B)$$

Bayesian Inference

Fitting models in the presence of data and parameter uncertainty

Data has noise

$$D \sim P(D|y(\theta))$$



Model depends on parameters

$$y(\theta)$$

e.g. $$P(D|\theta) \propto e^{-(D-y)^2/2\sigma^2}$$

Likelihood: $$P(D|y(\theta)) \equiv P(D|\theta)$$

Bayesian Inference

Fitting models in the presence of data and parameter uncertainty

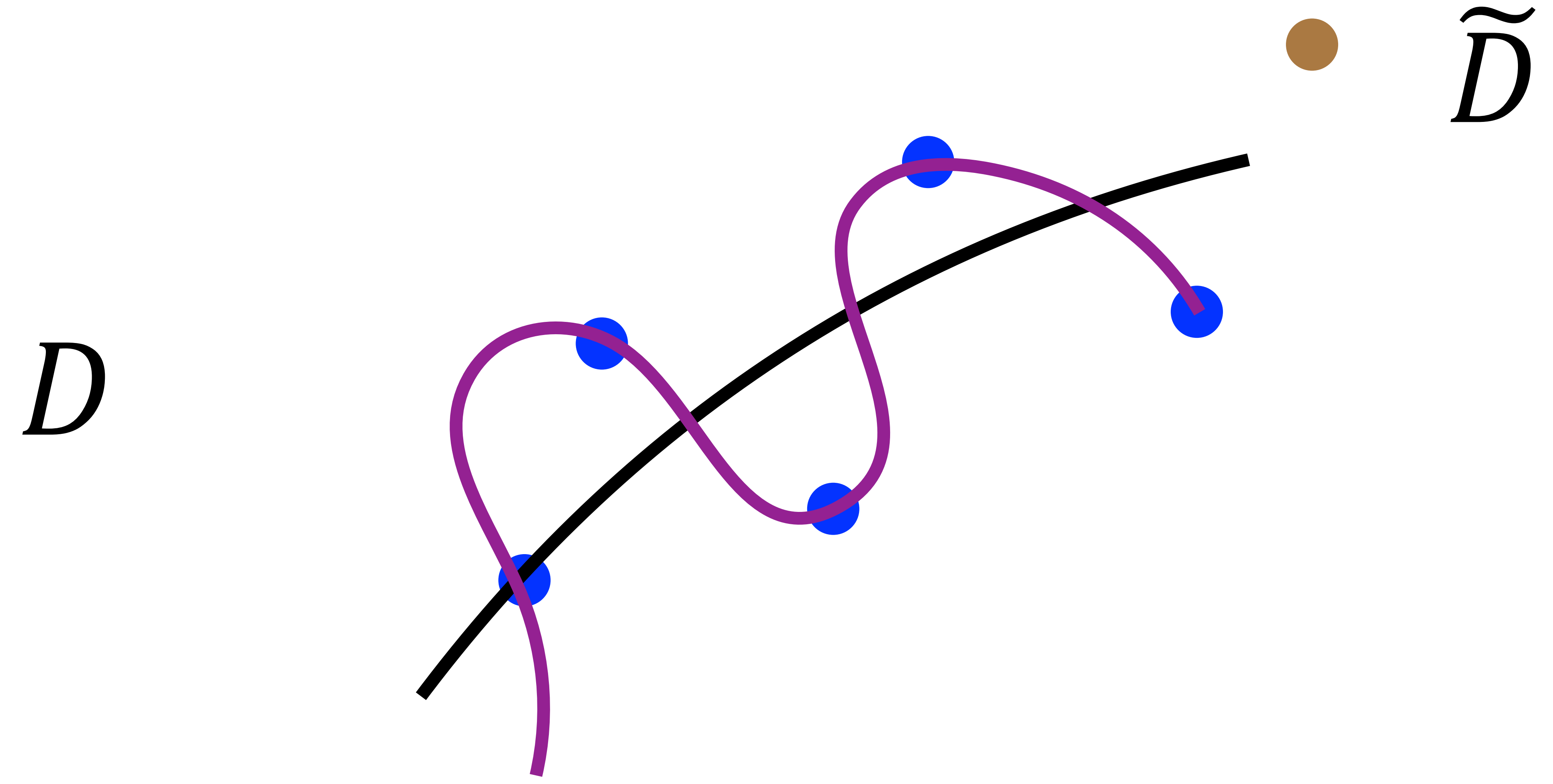
Prior

Parameter posterior $$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{P(D|\theta)P(\theta)}{\int P(D|\theta)P(\theta)d\theta}$$

Projections $$P(\widetilde{D}|D) = \int P(\widetilde{D}|\theta)P(\theta|D)d\theta$$

Bayesian model comparison and model averaging

Try lots of models, M



Balance fit with complexity

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

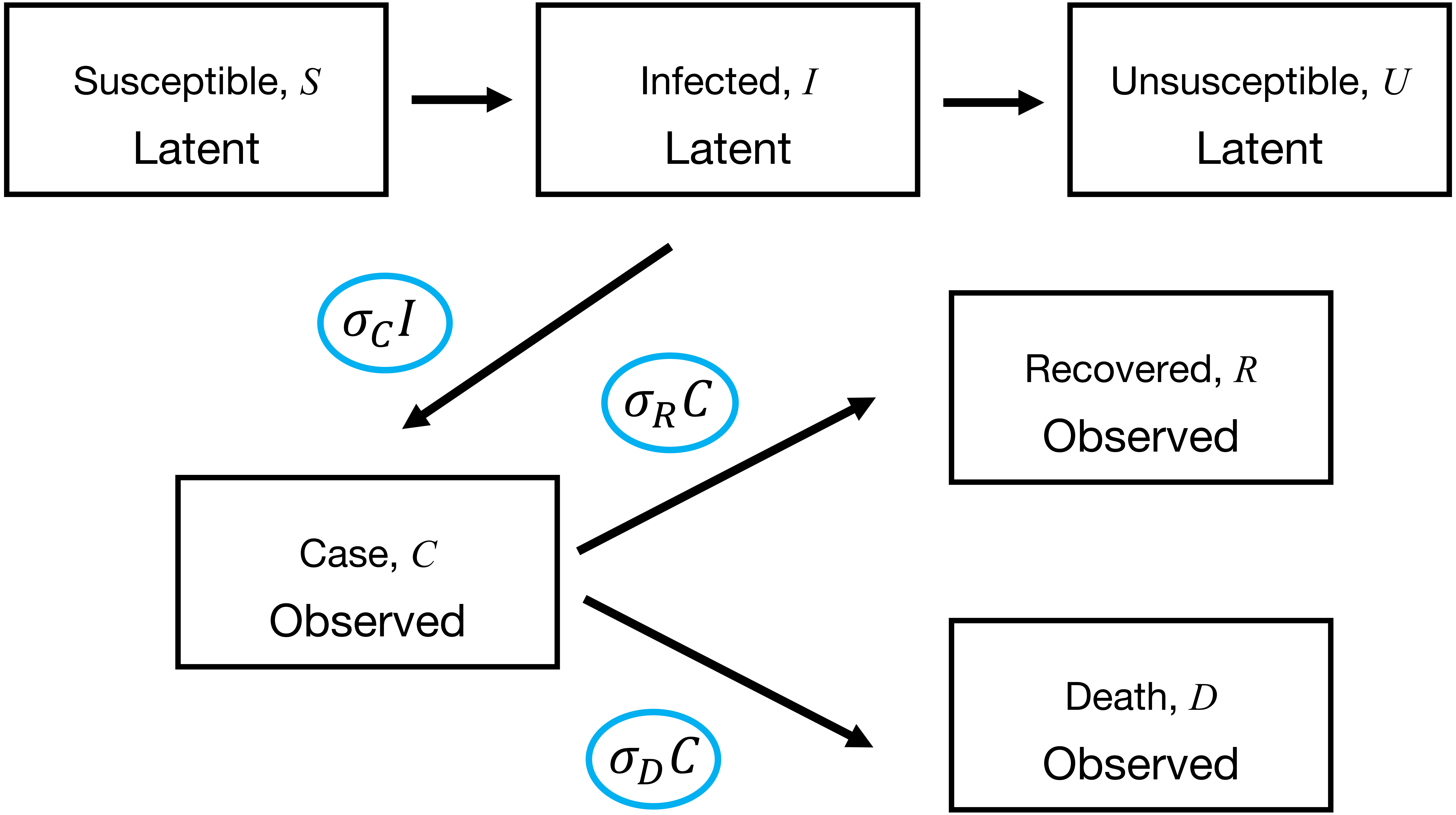
Maximize model likelihood $P(D|M)$

Bayes factor

Maximize prediction $P(\widetilde{D}|M)$

cross-validation, Information measures, e.g. AIC, WAIC,…

Pick the right data to fit



Noise model

If events arrive randomly then counts are Poisson distributed

Add extra variance due to heterogeneity and misreporting

New cases / day	$\sim \mathrm{NegBinom}(\sigma_c I, \phi)$
New recoveries / day	$\sim \mathrm{NegBinom}(\sigma_R C, \phi)$
Daily deaths / day	$\sim \mathrm{NegBinom}(\sigma_D C, \phi)$

$$var = \lambda + \frac{\lambda^2}{\phi}$$

Over dispersion compared to Poisson

Include as prior

NIH seropositivity estimates for May 9 - June 25, 2020

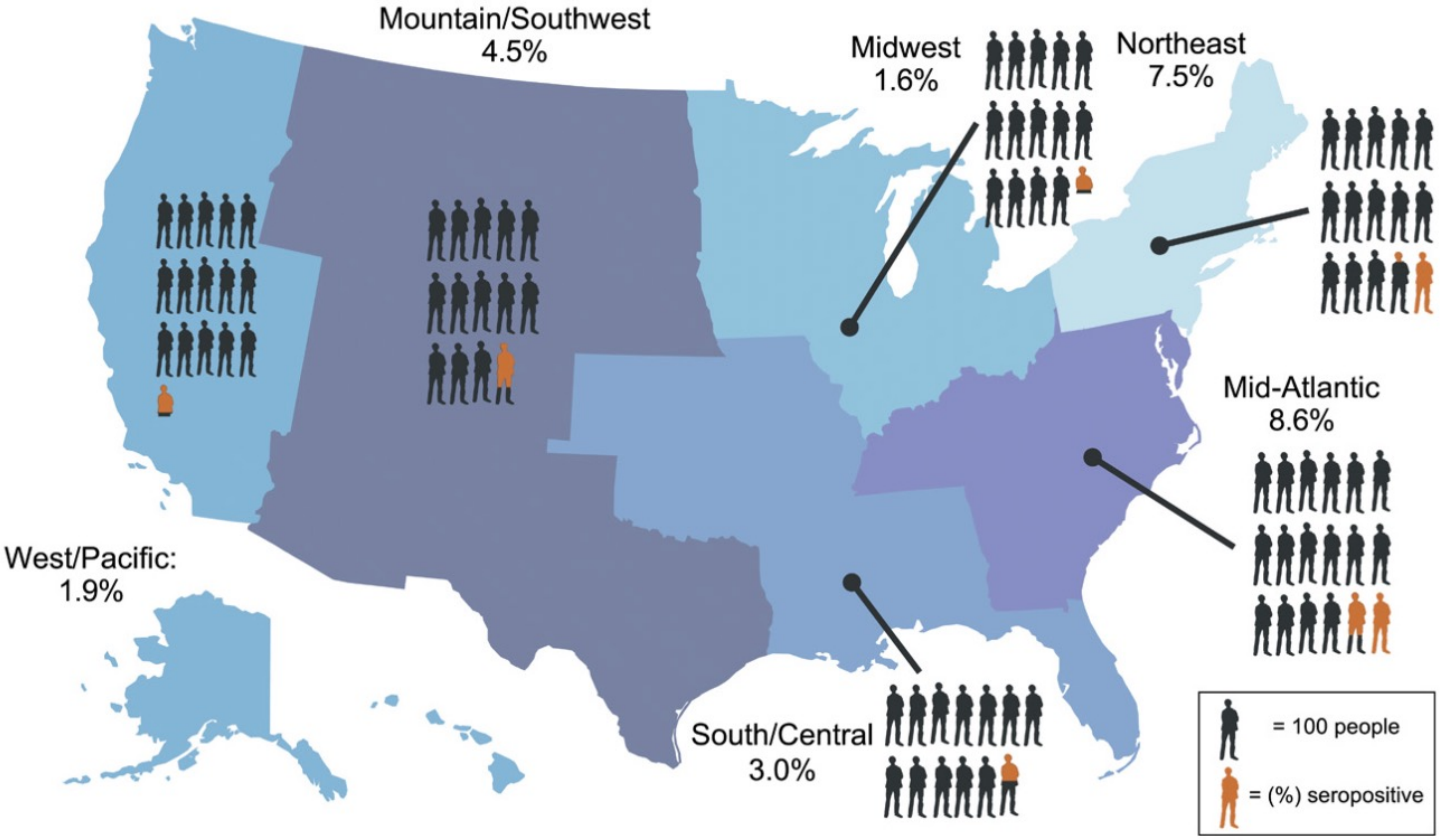


Figure from "Undiagnosed SARS-CoV-2 seropositivity during the first 6 months of the COVID-19 pandemic in the United States" by Heather Kalish et al, Science Translational Medicince https://www.science.org/doi/10.1126/scitranslmed.abh3826

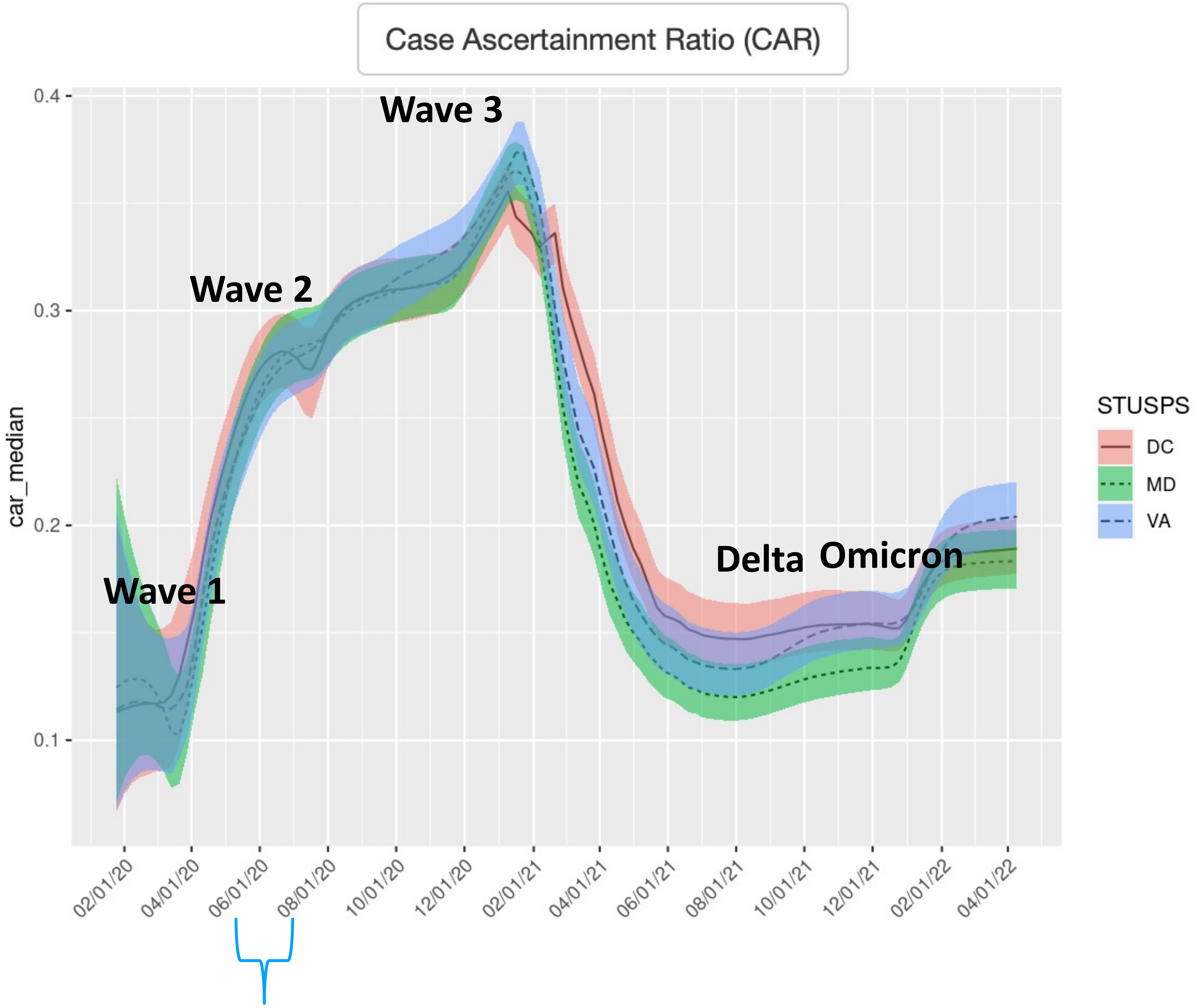
Hierarchical (mixed) model using Stan

https://mc-stan.org/

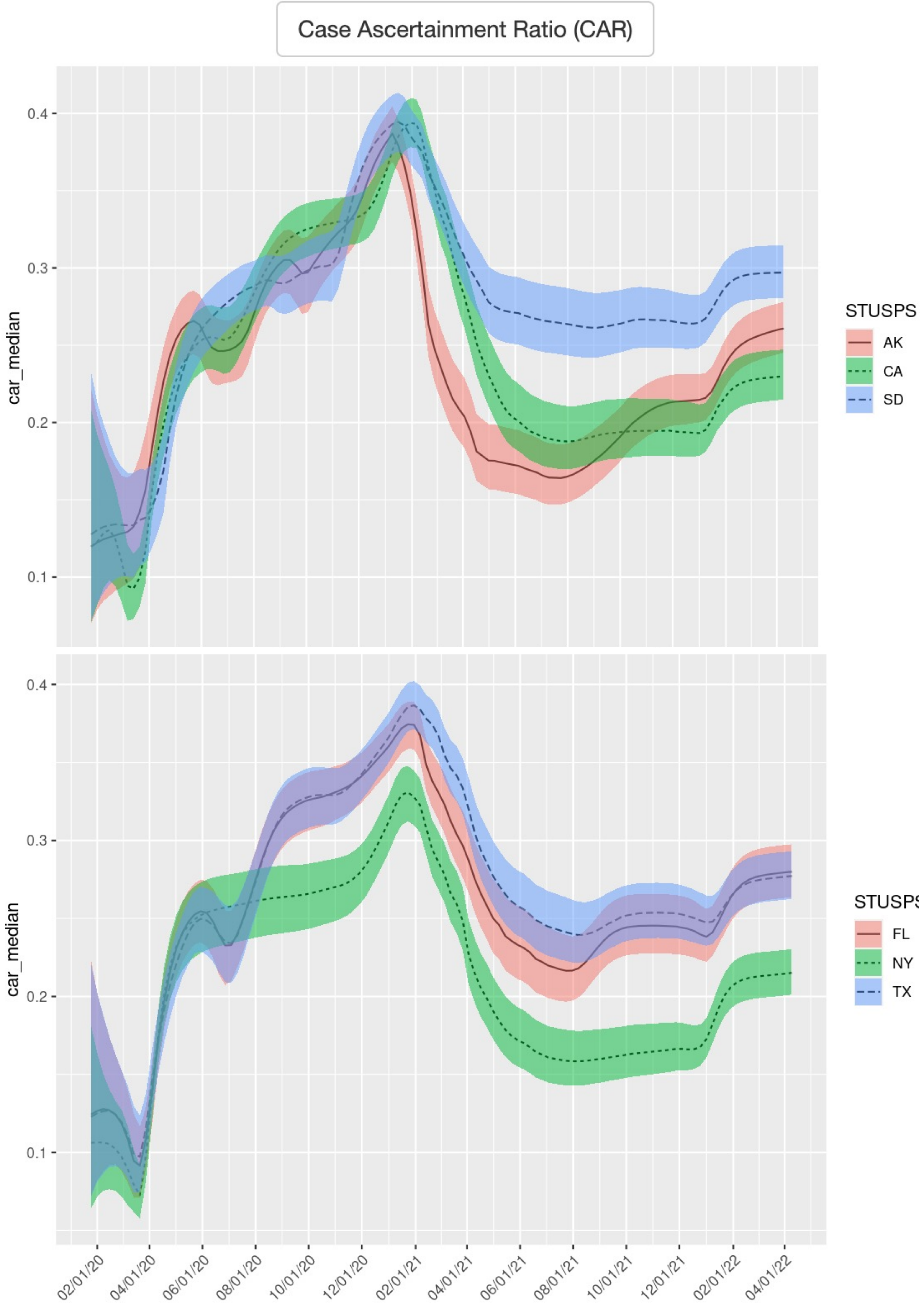
Markov Chain Monte Carlo (NUTS)

https://github.com/nih-niddk-mbs/covid-sicr/tree/Hierarchical

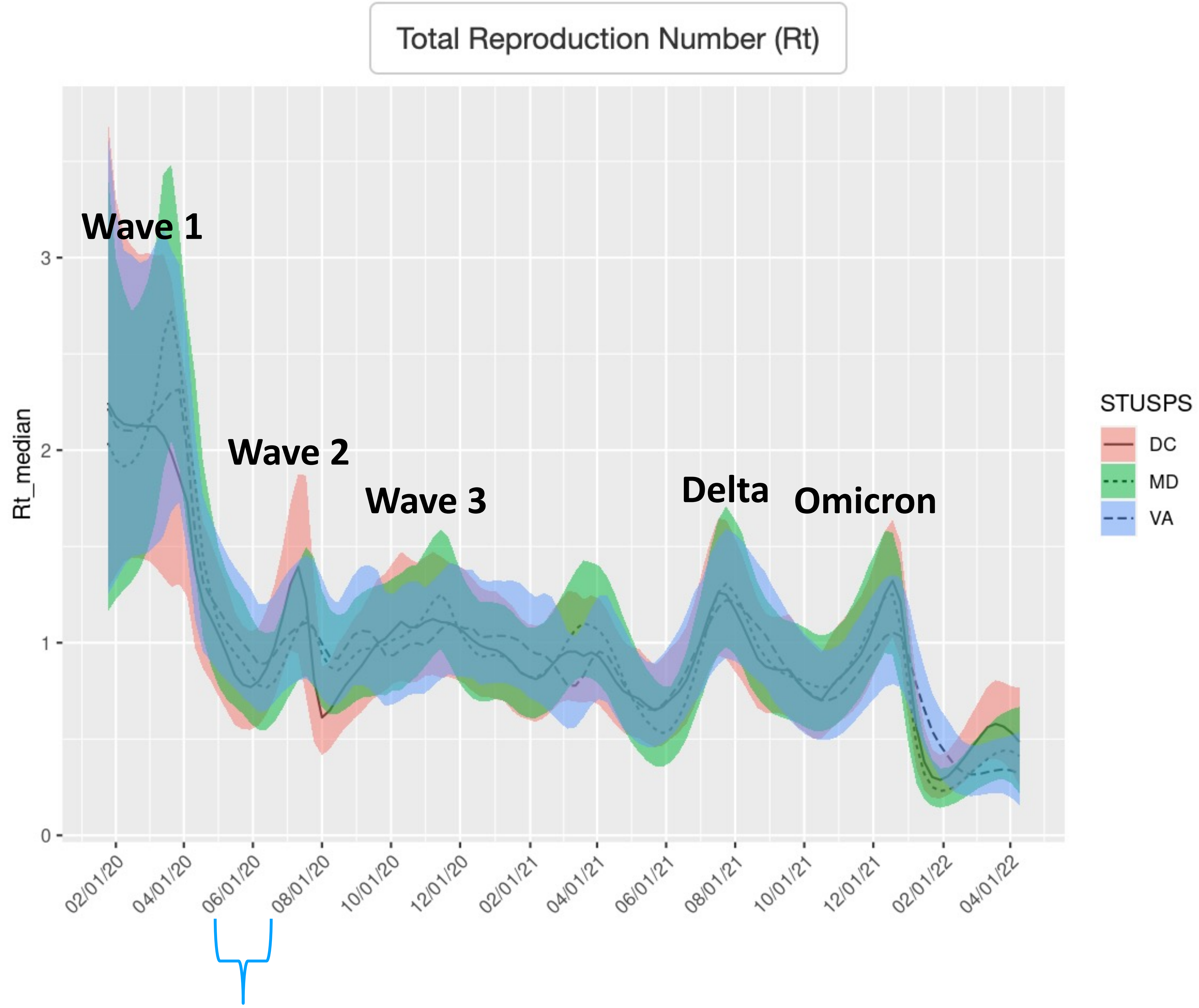
Case Ascertainment Ratio (CAR_t)



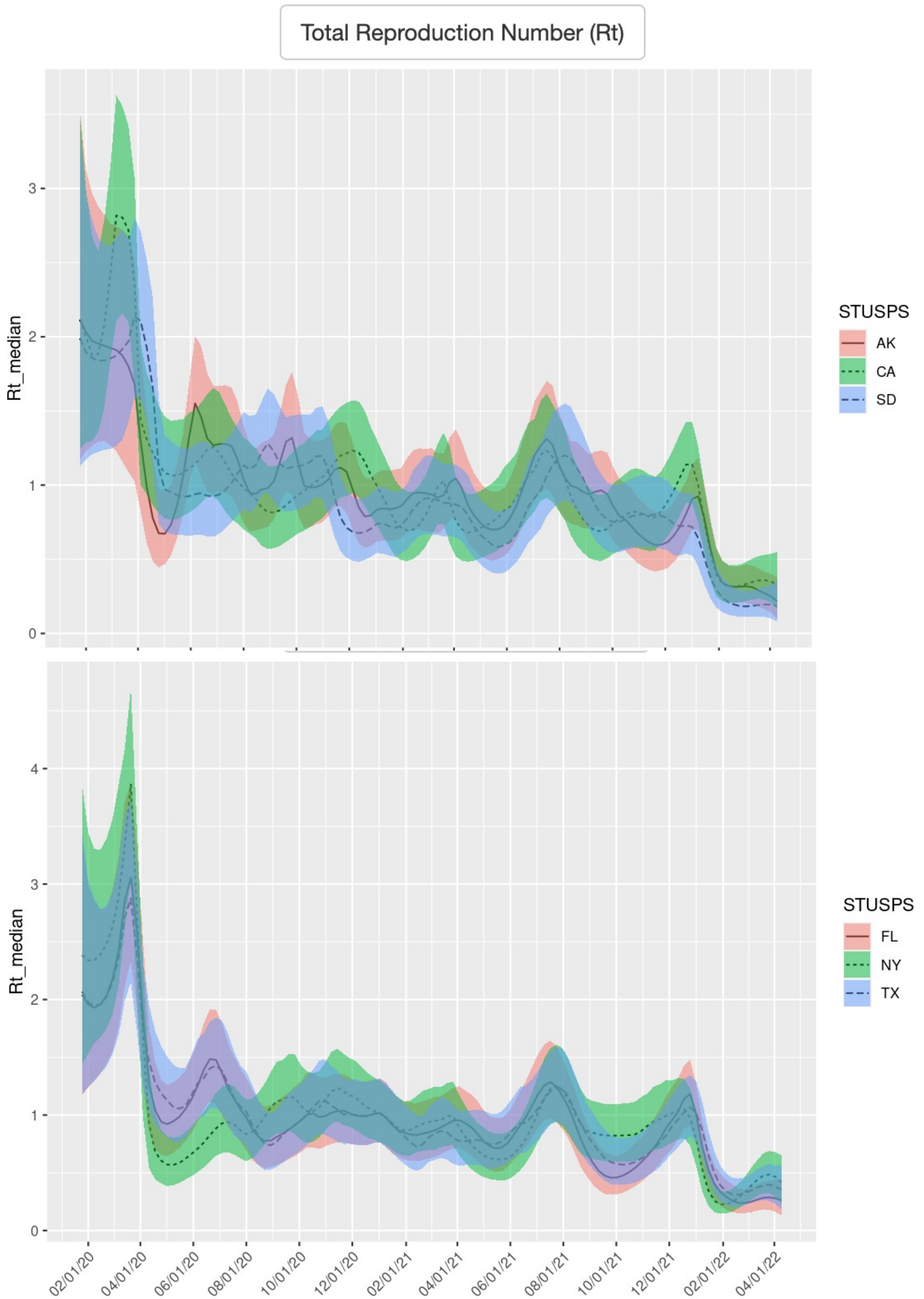
*Seropositivity priors extend from May 9 - June 25, 2020



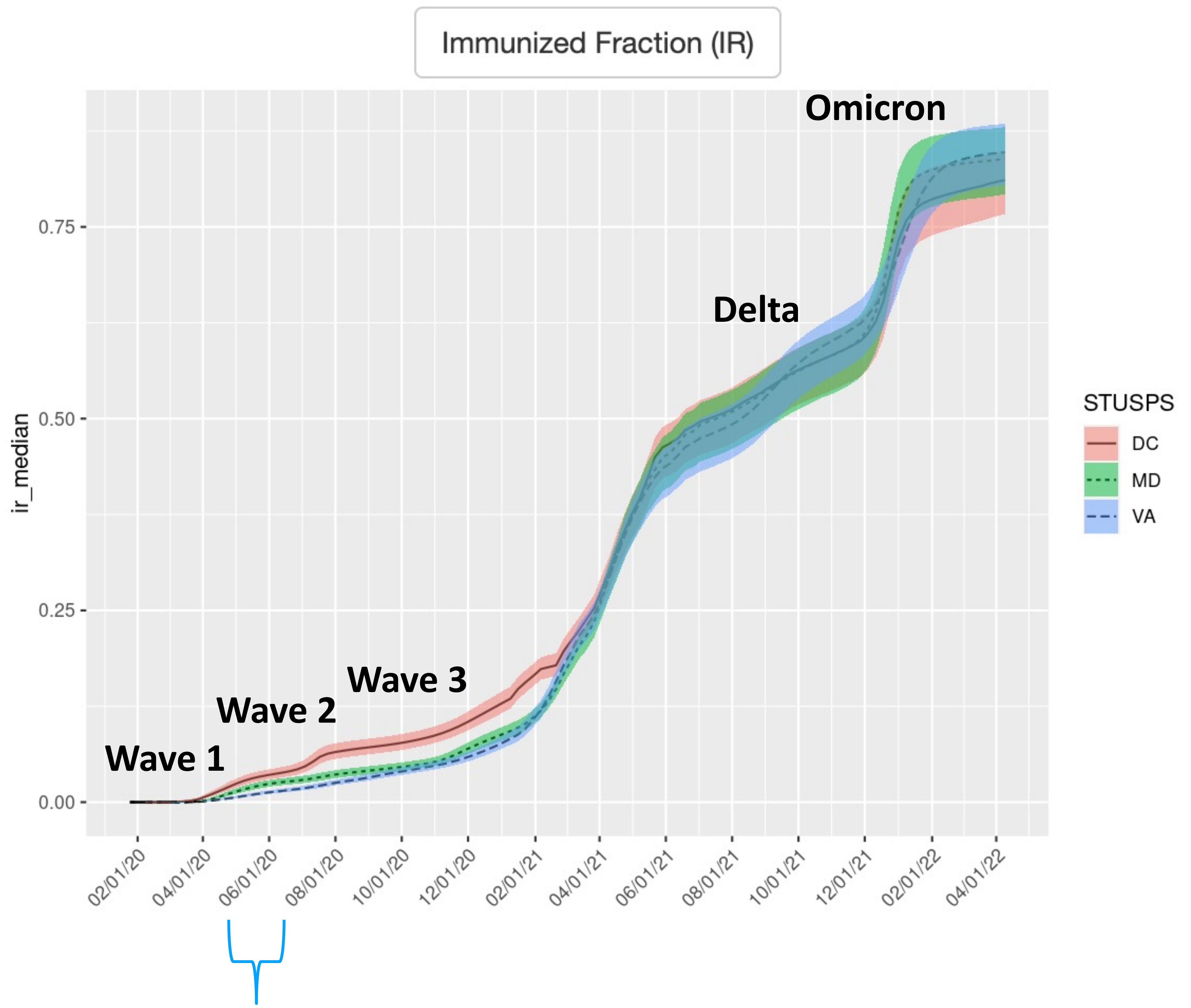
Reproduction Number (R_t)



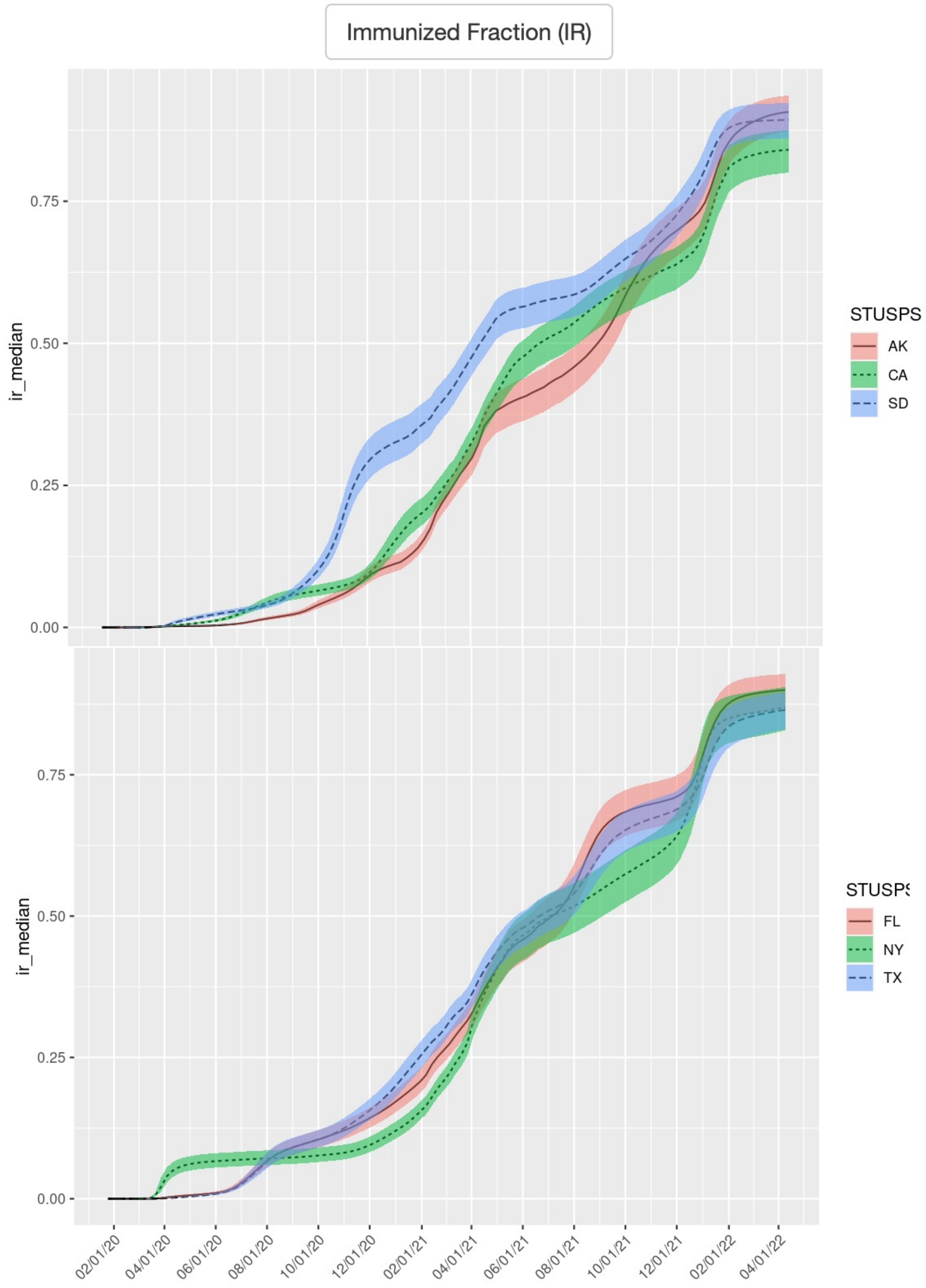
*Seropositivity priors extend from May 9 - June 25, 2020



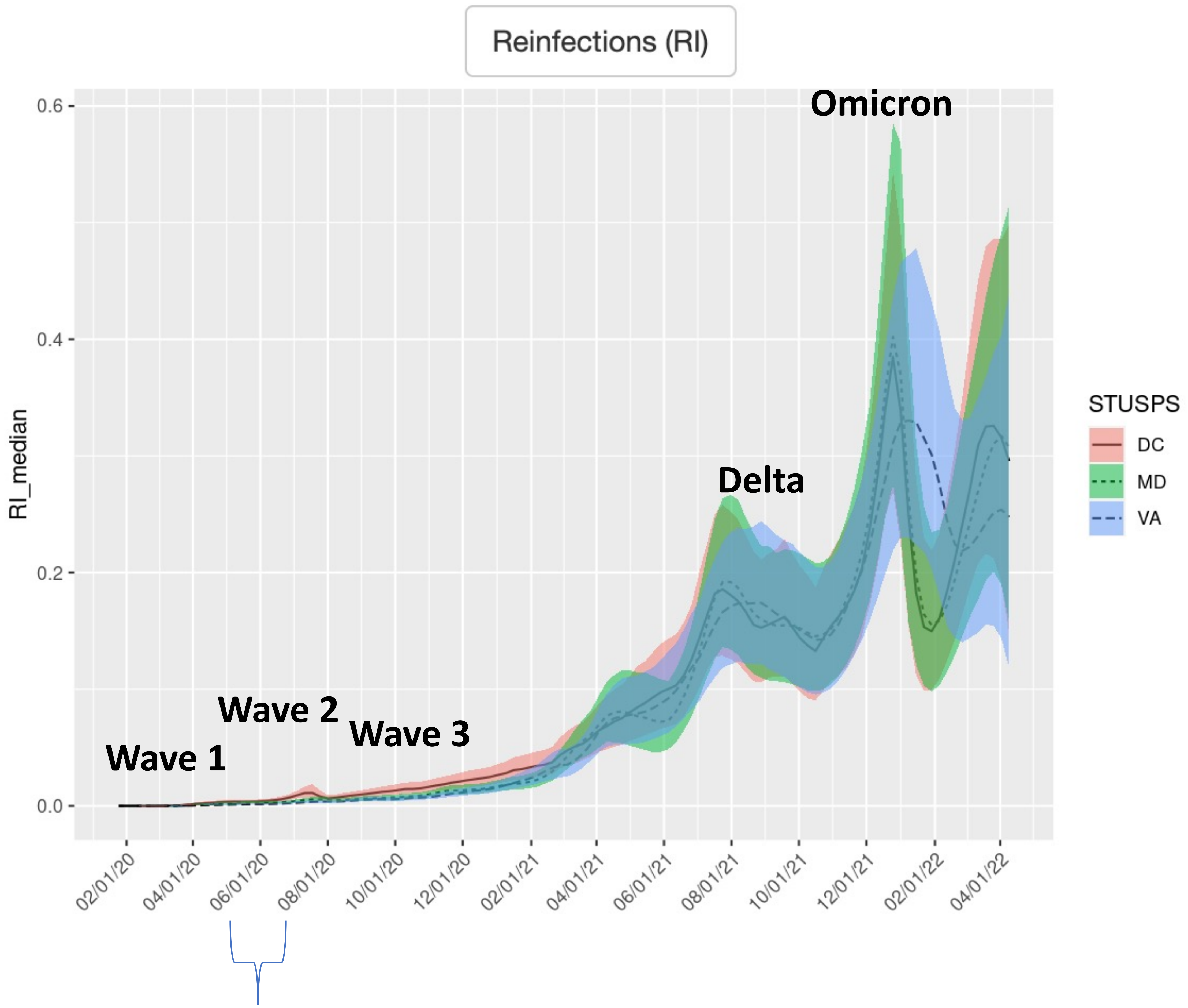
Fraction infected and/or vaccinated



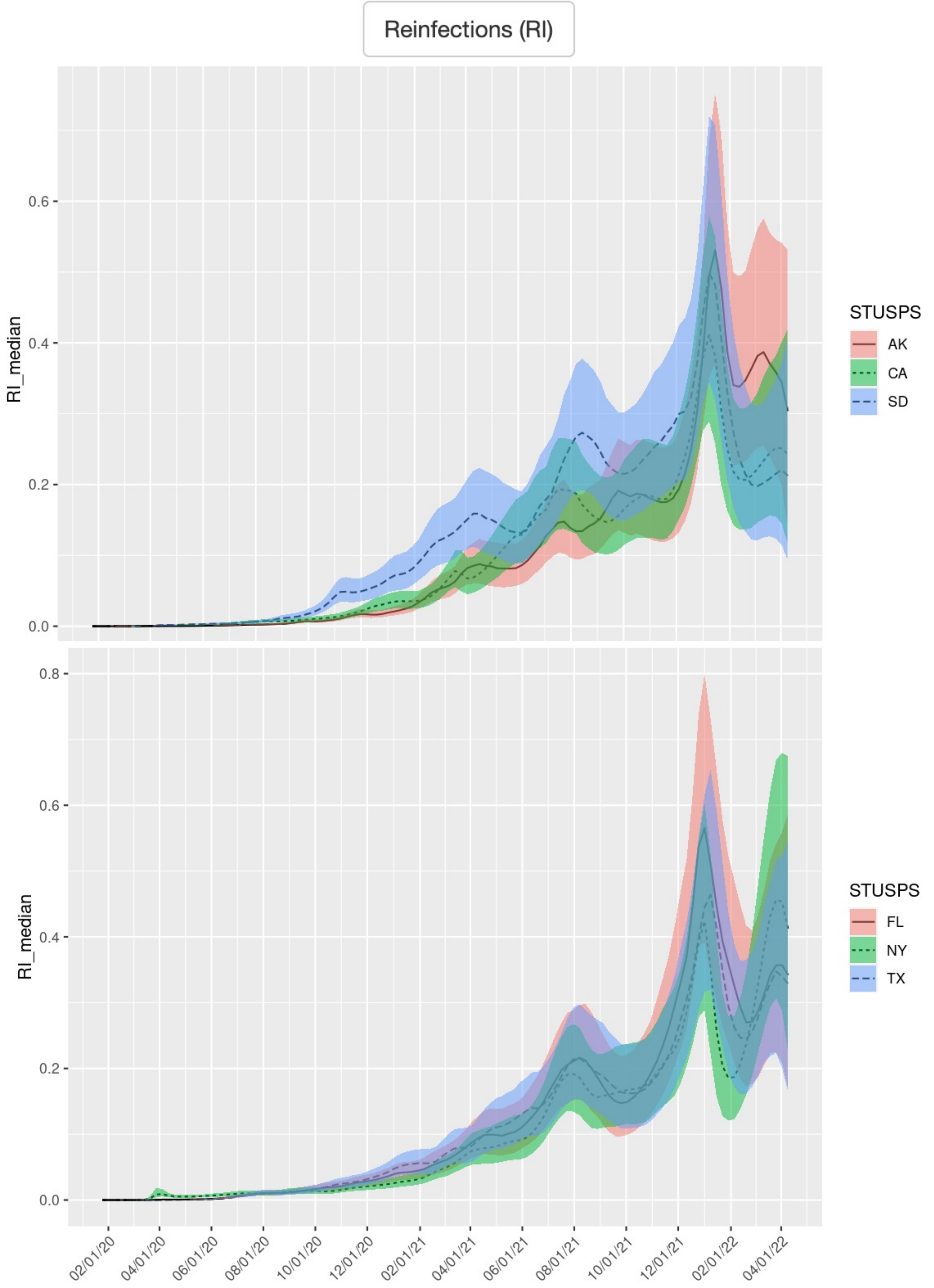
*Seropositivity priors extend from May 9 - June 25, 2020



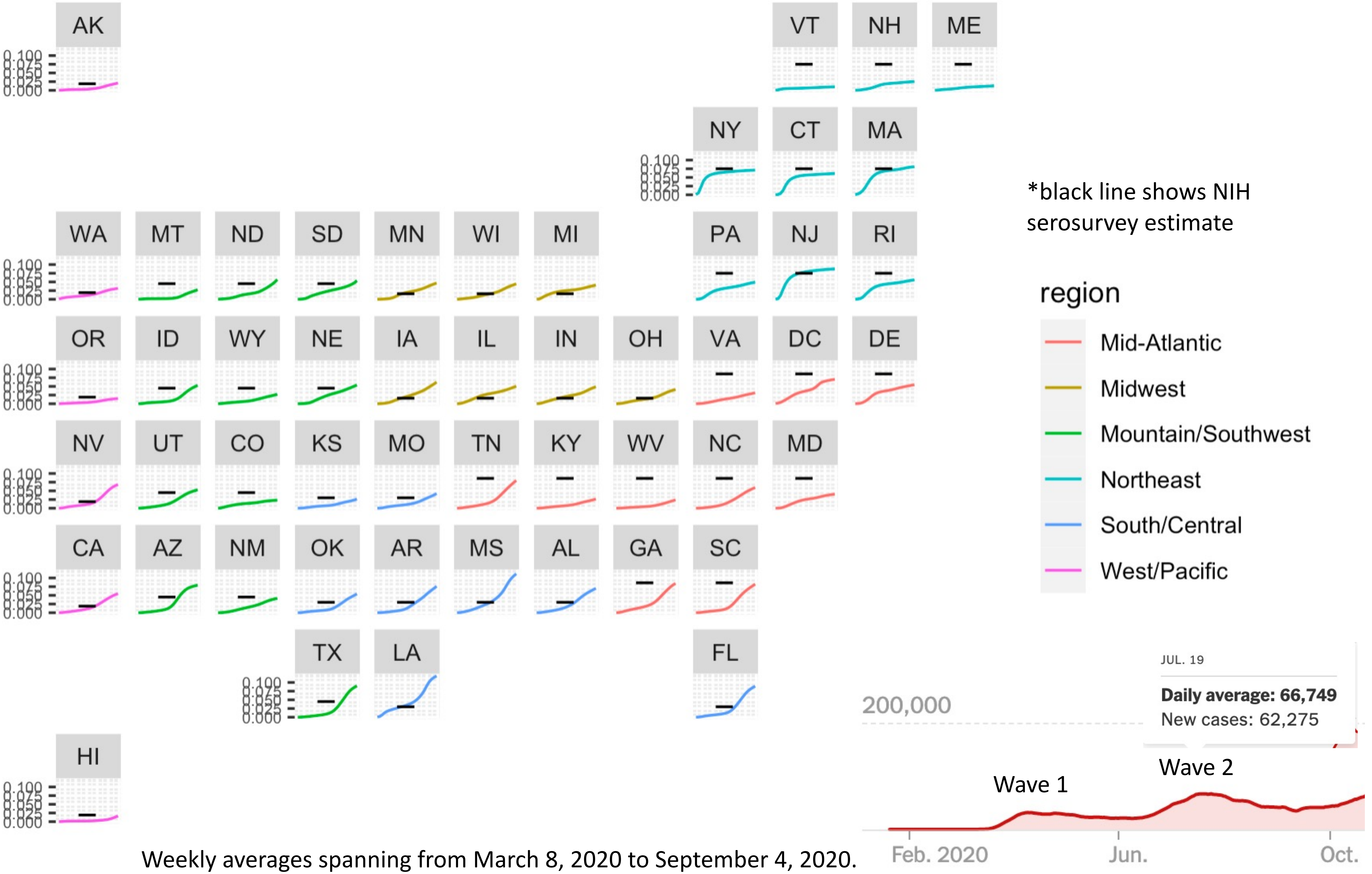
Fraction of "immunized" population infected (RI_t)



*Seropositivity priors extend from May 9 - June 25, 2020

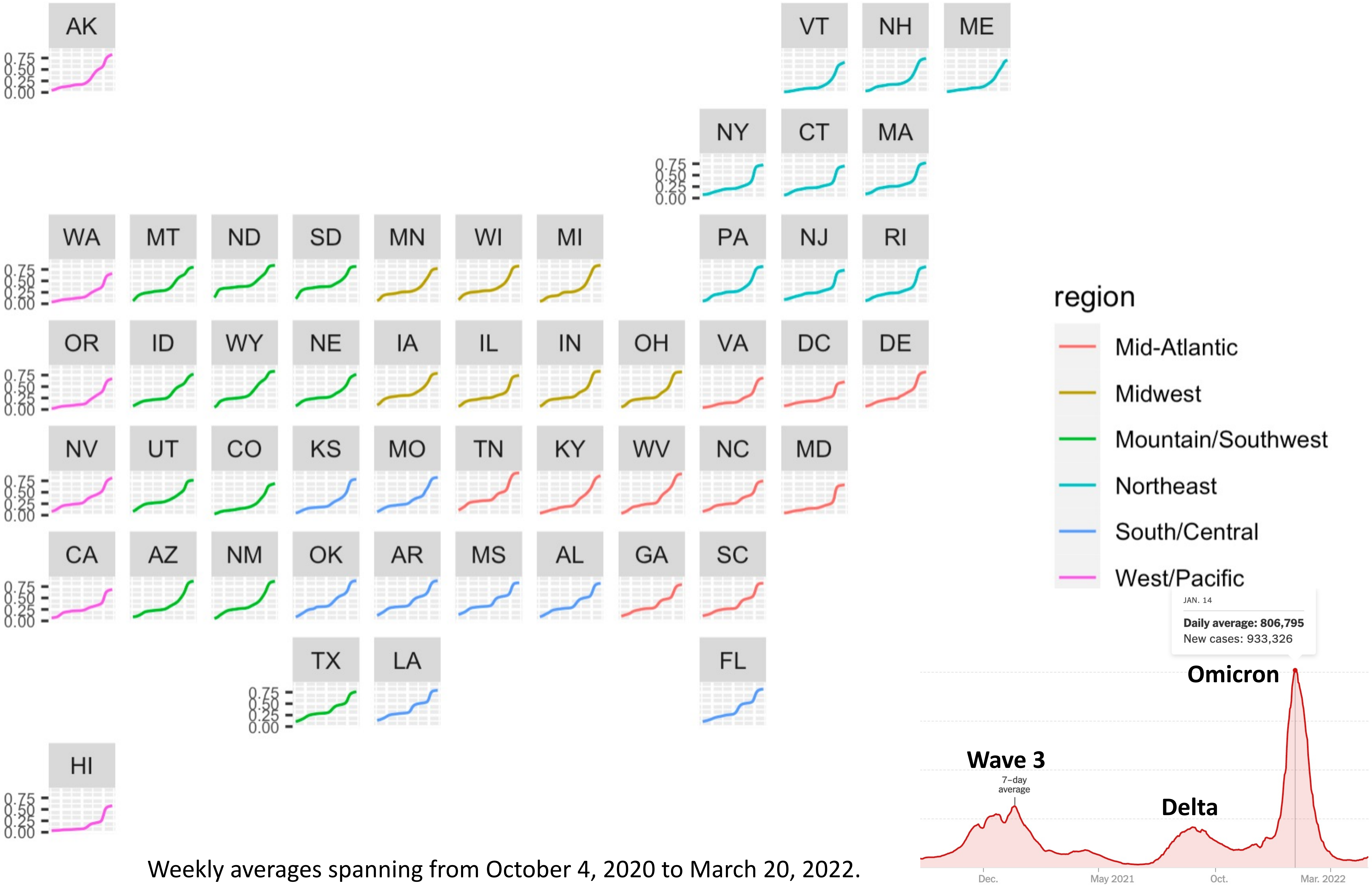


median predicted total infected fraction during Waves 1 and 2

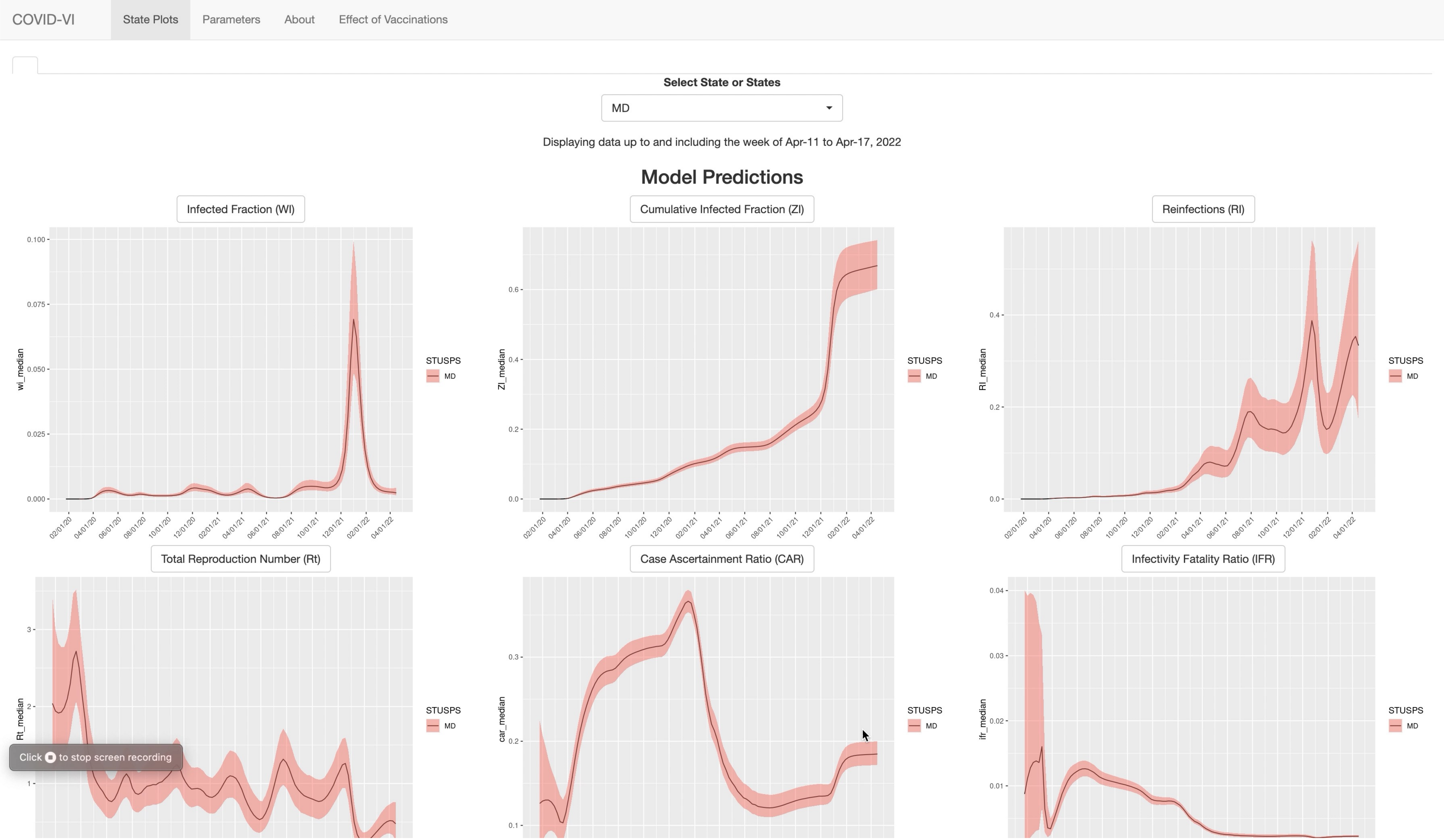


Weekly averages spanning from March 8, 2020 to September 4, 2020.

median predicted total infected fraction during Waves 3, Delta, and Omicron



Weekly averages spanning from October 4, 2020 to March 20, 2022.



https://aoschwartz.shinyapps.io/COVID-ShinyAnimated/

https://www.medrxiv.org/content/10.1101/2020.04.29.20083485v1

https://github.com/nih-niddk-mbs/covid-sicr

Slides on sciencehouse.wordpress.com