

# **Bayesian semi-mechanistic modelling of COVID-19**

## **StanCon 2020**

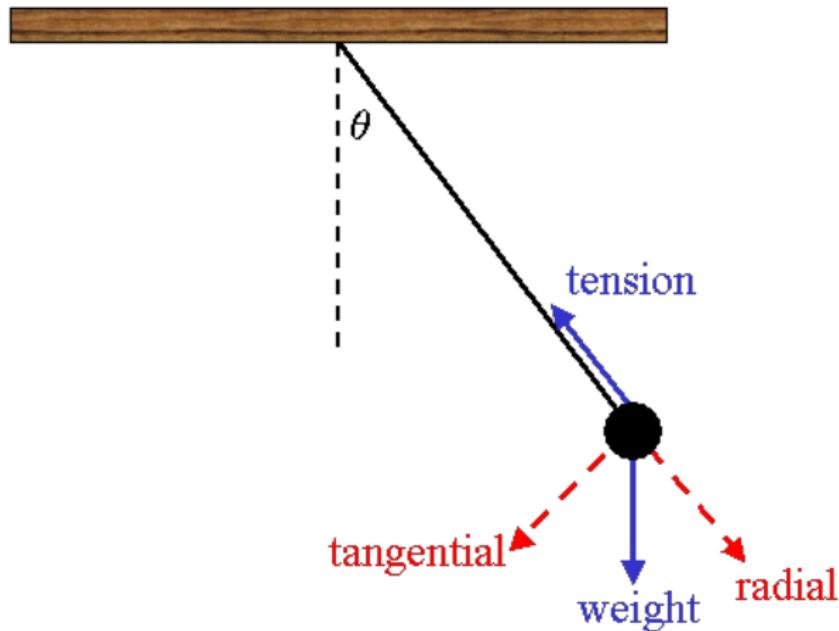
**Samir Bhatt, Helen Coupland, Seth Flaxman, Axel Gandy,  
Thomas Mellan, Swapnil Mishra, Mélodie Monod, Oliver  
Ratmann, Jamie Scott, Juliette Unwin**

**Imperial College  
London**

# SARS-CoV2 and COVID-19 Context

# SARS-CoV2 and COVID-19: Context

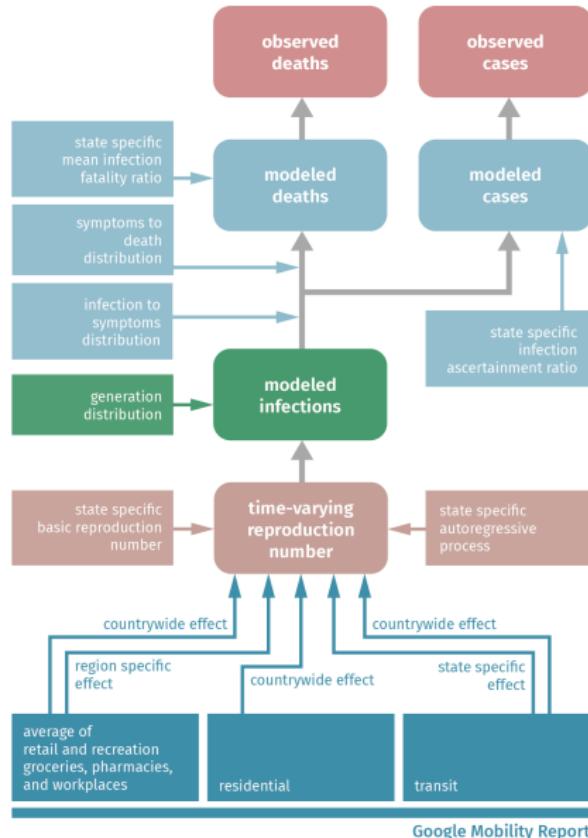
- Once-in-a-century pathogen we've been worried about
- Right away the world began curve fitting
- Epidemiological mechanism needed to understand dynamics



# Challenges and Key Questions

- How to model infection process? *Bellman-Harris Counting Process*
- Modelling epidemics requires case or infection data
- Early on (maybe still now), had to rely on COVID-19 death data
- We build a model connecting deaths to infections to transmission
- Semi-mechanistic: we fix some key epidemiological parameters
- **Estimate how transmission, based on predictors, is changing**

# Model Schematic: General Version



# Talk Overview

- 1 SARS-CoV2 and COVID-19 Context**
- 2 Model Details and Stan Implementation**
- 3 A tale of Scandinavia**
- 4 Case studies: Brazil, USA, NYS**
- 5 *epidemia*: a package for flexible epidemic modelling**
- 6 Age structure and contact patterns**
- 7 Where do we go next?**
- 8 References**

# **Model Details and Stan Implementation**

# Infection process (Renewal Equation)<sup>1</sup>

$$i_{t,m} = S_{t,m} R_{t,m} \sum_{\tau=0}^{t-1} i_{\tau,m} g_{t-\tau},$$

$$S_{t,m} = 1 - \frac{\sum_{j=1}^{t-1} i_{j,m}}{N_m}$$

---

<sup>1</sup> Swapnil Mishra: <https://vimeo.com/447243523>

# Infection process (Renewal Equation)<sup>1</sup>

$$i_{t,m} = S_{t,m} R_{t,m} \sum_{\tau=0}^{t-1} i_{\tau,m} g_{t-\tau}, \quad S_{t,m} = 1 - \frac{\sum_{j=1}^{t-1} i_{j,m}}{N_m}$$

```
real convolution = dot_product(sub_col(prediction, 1, m,
                                         i-1),
                               tail(SI_rev, i-1));

cumm_sum[i, m] = cumm_sum[i-1, m] + prediction[i-1, m];
Rt_adj[i, m] = ((pop[m]-cumm_sum[i, m]) / pop[m]) * Rt[i, m];

prediction[i, m] = Rt_adj[i, m] * convolution;
```

---

<sup>1</sup> Swapnil Mishra: <https://vimeo.com/447243523>

# Death process (Survival Process)

$$d_{t,m} = \text{ifr}_m^* \sum_{\tau=0}^{t-1} i_{\tau,m} \pi_{t-\tau}$$

$$\pi \sim \text{Gamma}(5.1, 0.86) + \\ \text{Gamma}(17.8, 0.45)$$

# Death process (Survival Process)

$$d_{t,m} = \text{ifr}_m^* \sum_{\tau=0}^{t-1} i_{\tau,m} \pi_{t-\tau}$$

$$\pi \sim \text{Gamma}(5.1, 0.86) + \\ \text{Gamma}(17.8, 0.45)$$

```
E_deaths[i,m] = dot_product(sub_col(prediction, 1, m, i-1),  
                           tail(f_rev[m], i-1)  
                           ) * ifr_noise[m];
```

# Reproduction Number

$$R_{t,m} = R_{0,m} \cdot f \left( - \left( \sum_{k=1}^{K_{full}} X_{t,m,k} \alpha_k \right) - \left( \sum_{k=1}^{K_{partial}} Y_{t,m,k} \alpha_{m,k}^{partial} \right) - \epsilon_{m,w_m(t)} \right)$$

# Reproduction Number

$$R_{t,m} = R_{0,m} \cdot f \left( - \left( \sum_{k=1}^{K_{full}} X_{t,m,k} \alpha_k \right) - \left( \sum_{k=1}^{K_{partial}} Y_{t,m,k} \alpha_{m,k}^{partial} \right) - \epsilon_{m,w_m(t)} \right)$$

```
Rt[,m] = mu[m] * 2 * inv_logit(-X[m] * alpha  
-X_partial[m] * alpha_partial[m]  
-weekly_effect[week_index[m],m]);  
  
kappa ~ normal(0,0.5);  
mu ~ normal(3.28, kappa);
```

# Death Likelihood

$$D_{t,m} \sim \text{Negative Binomial} \left( d_{t,m}, d_{t,m} + \frac{d_{t,m}^2}{\psi} \right)$$

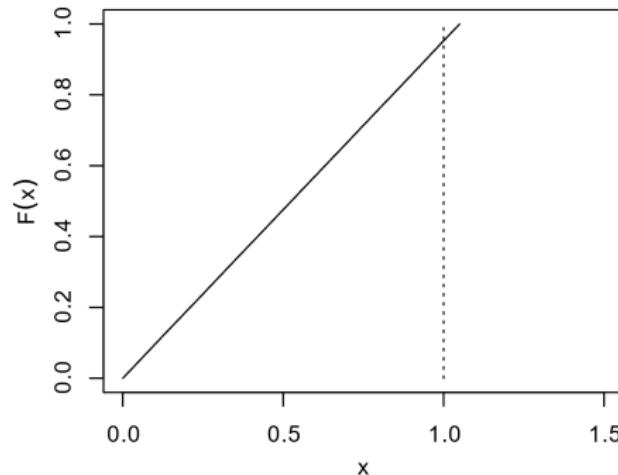
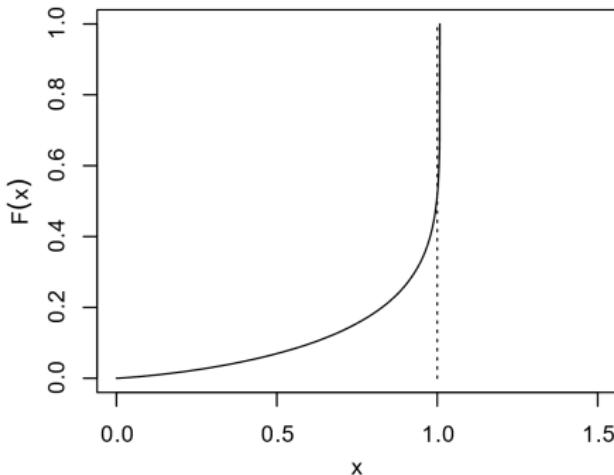
# Death Likelihood

$$D_{t,m} \sim \text{Negative Binomial} \left( d_{t,m}, d_{t,m} + \frac{d_{t,m}^2}{\psi} \right)$$

```
deaths[, m] ~ neg_binomial_2(E_deaths[EpidemicStart[, m],  
phi);
```

# Prior Choice<sup>2</sup>

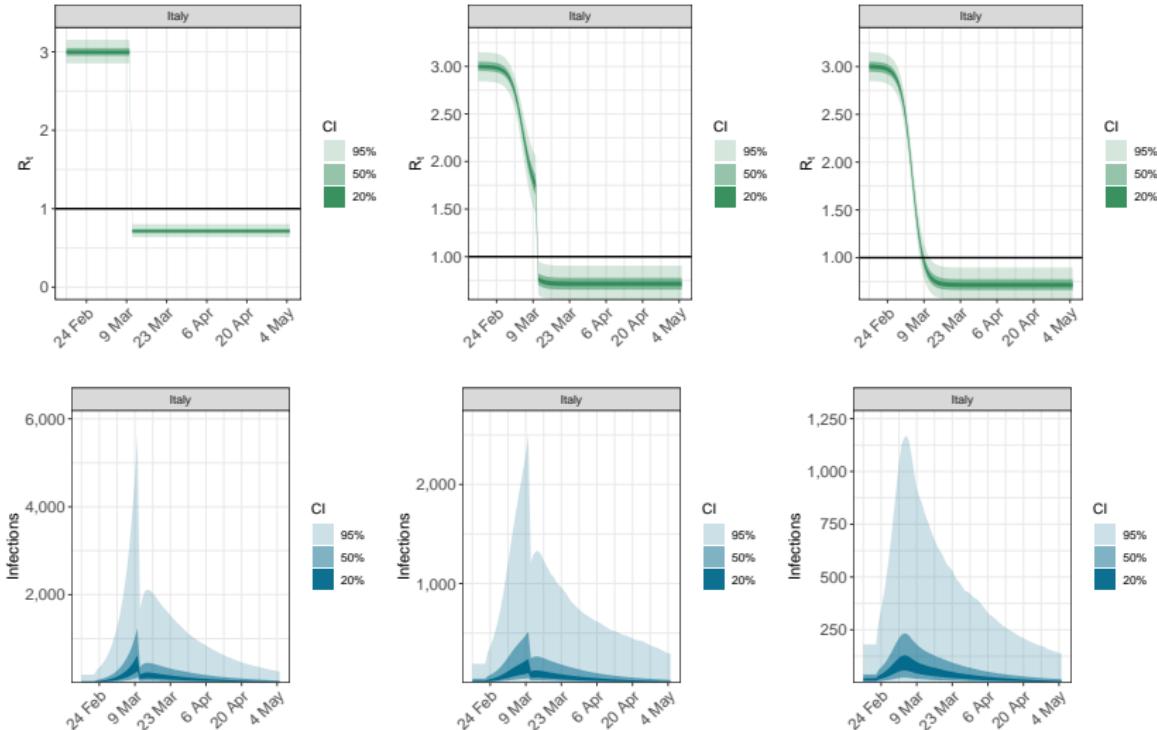
- Effect of interventions
- Situation: 2-3 weeks after lockdown in Italy



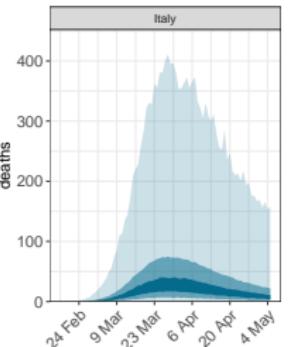
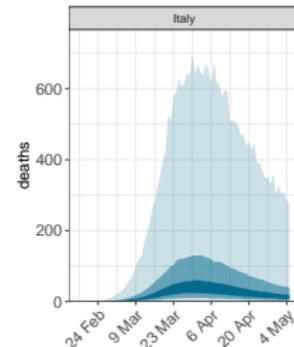
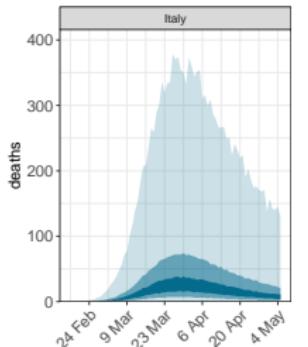
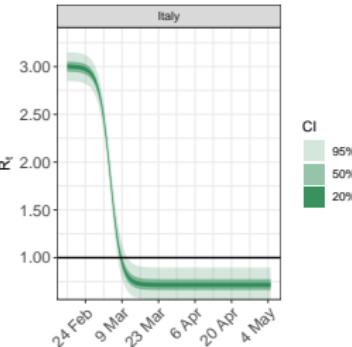
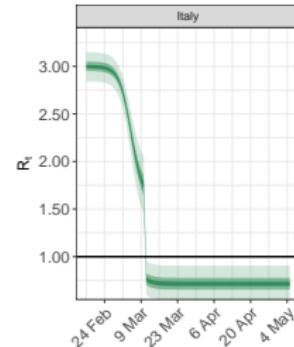
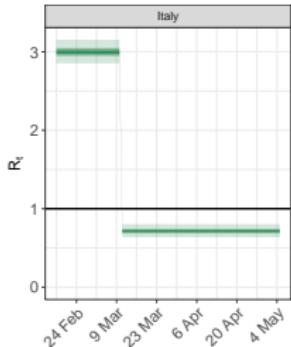
Left: CDF of prior on total reduction through one intervention [ $\exp(-\alpha_k)$ ]  
Right: prior on all interventions together [ $\exp(-\sum_{k=1}^6 \alpha_k)$ ]

<sup>2</sup>Axel Gandy: <https://vimeo.com/447135085>

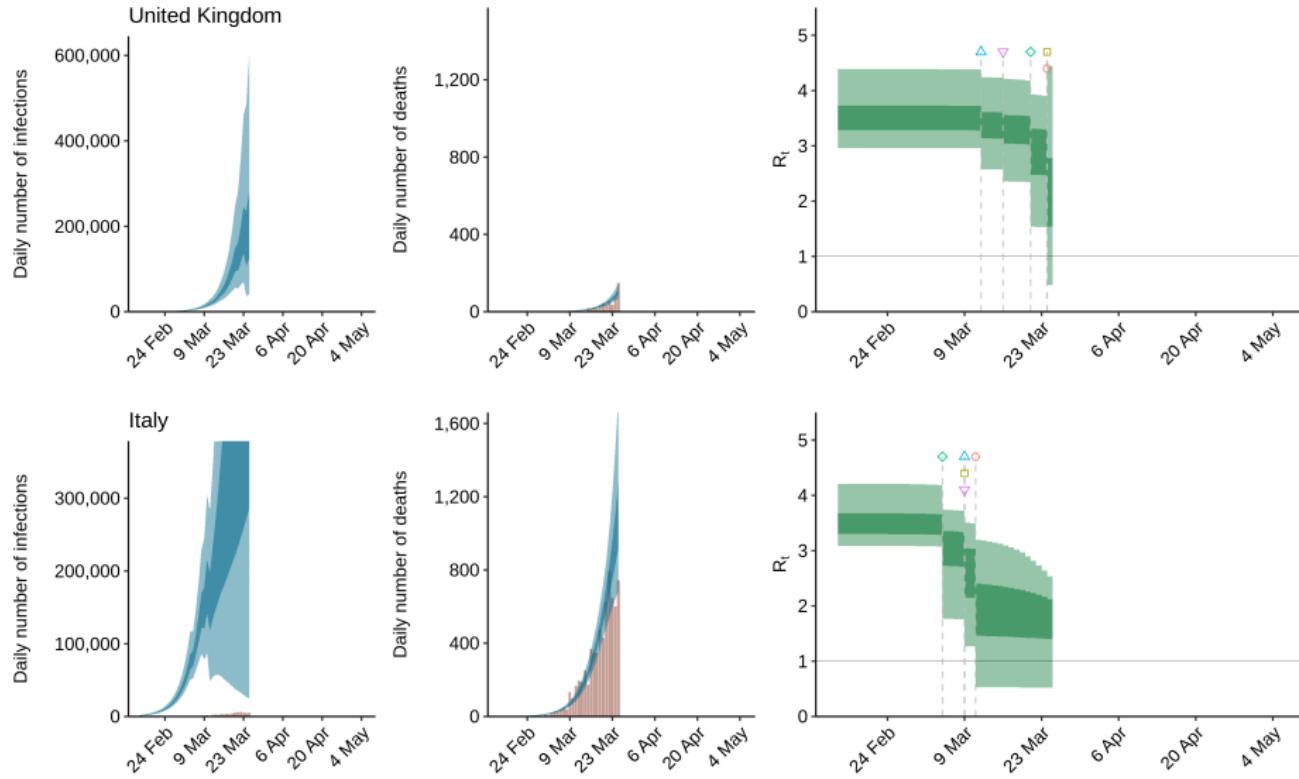
# Prior predictive checks



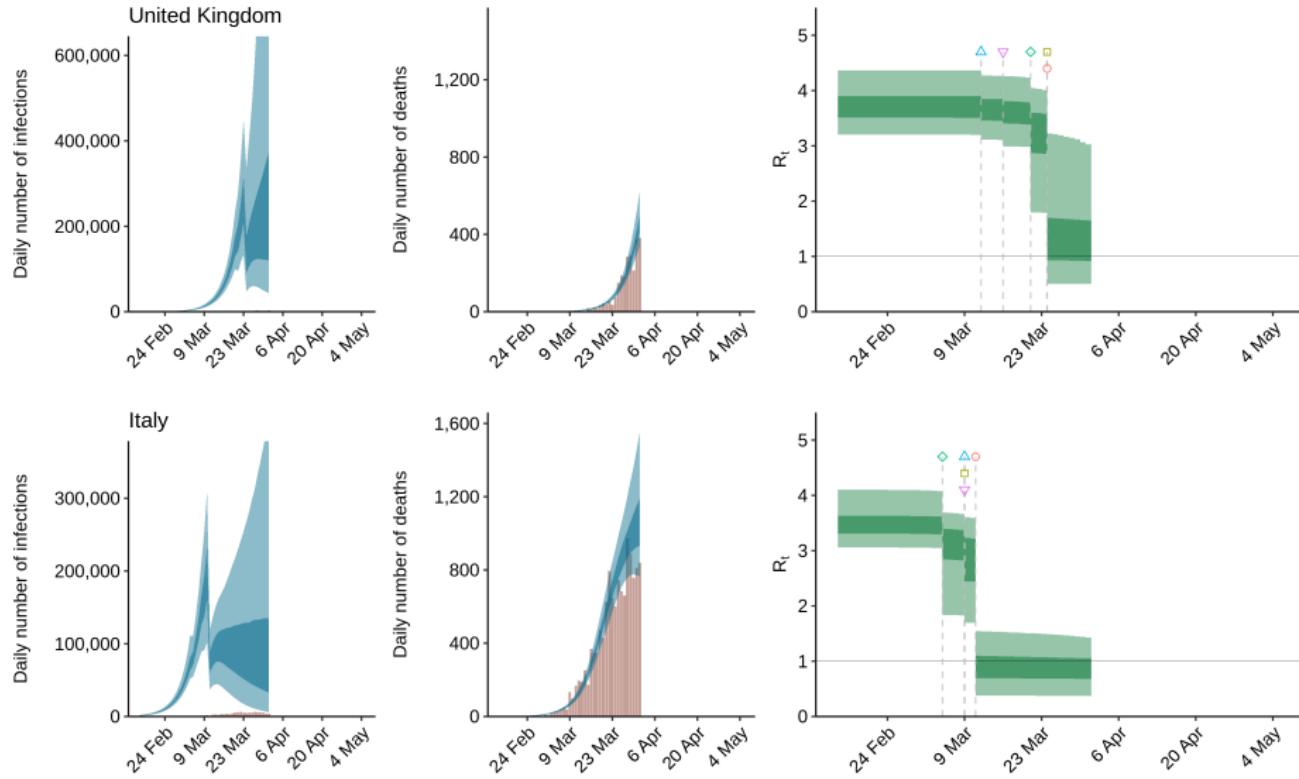
# Prior predictive checks



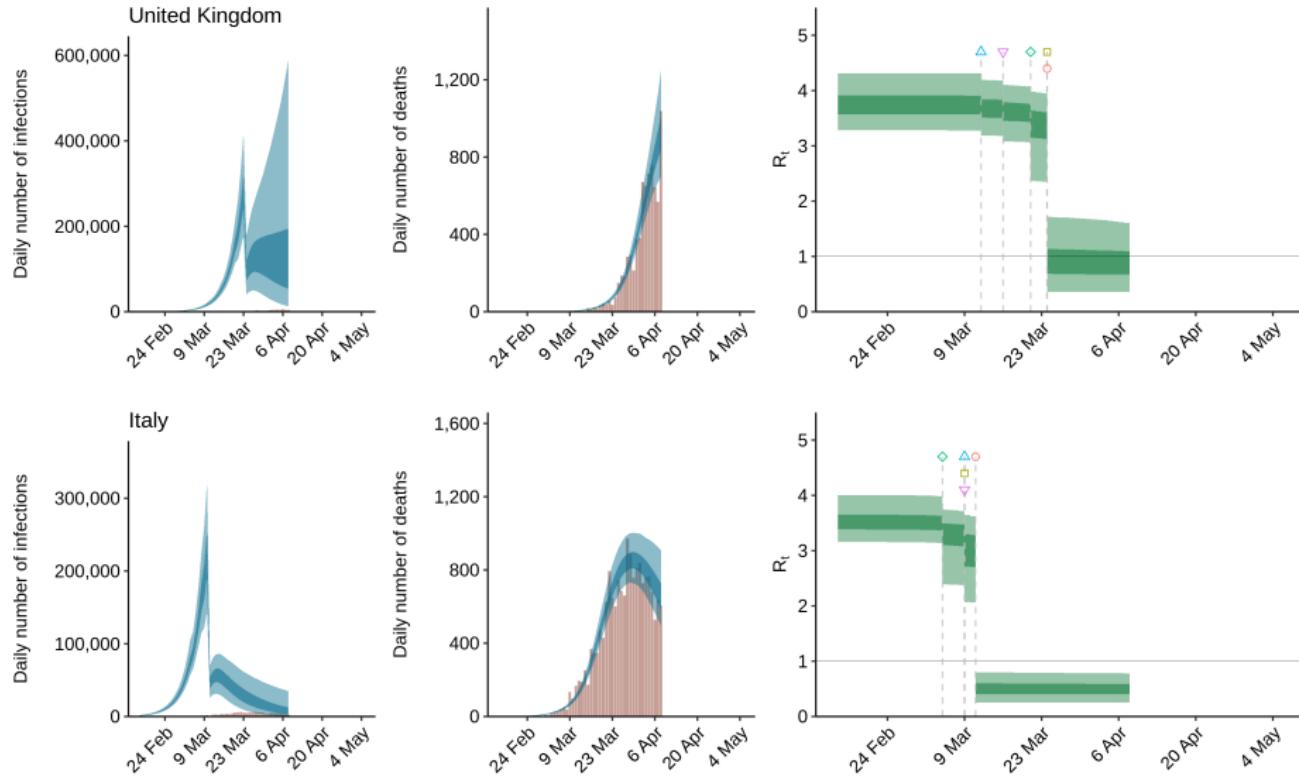
# Partial Pooling



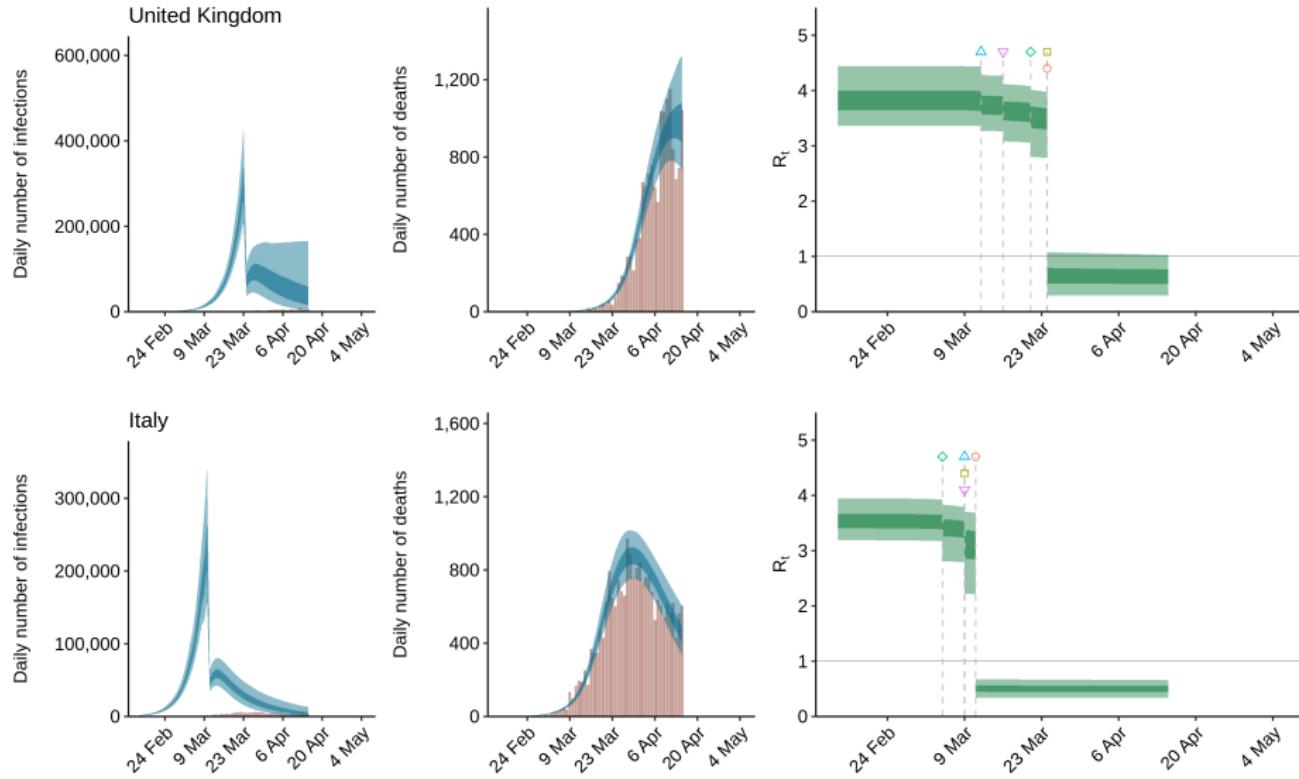
# Partial Pooling



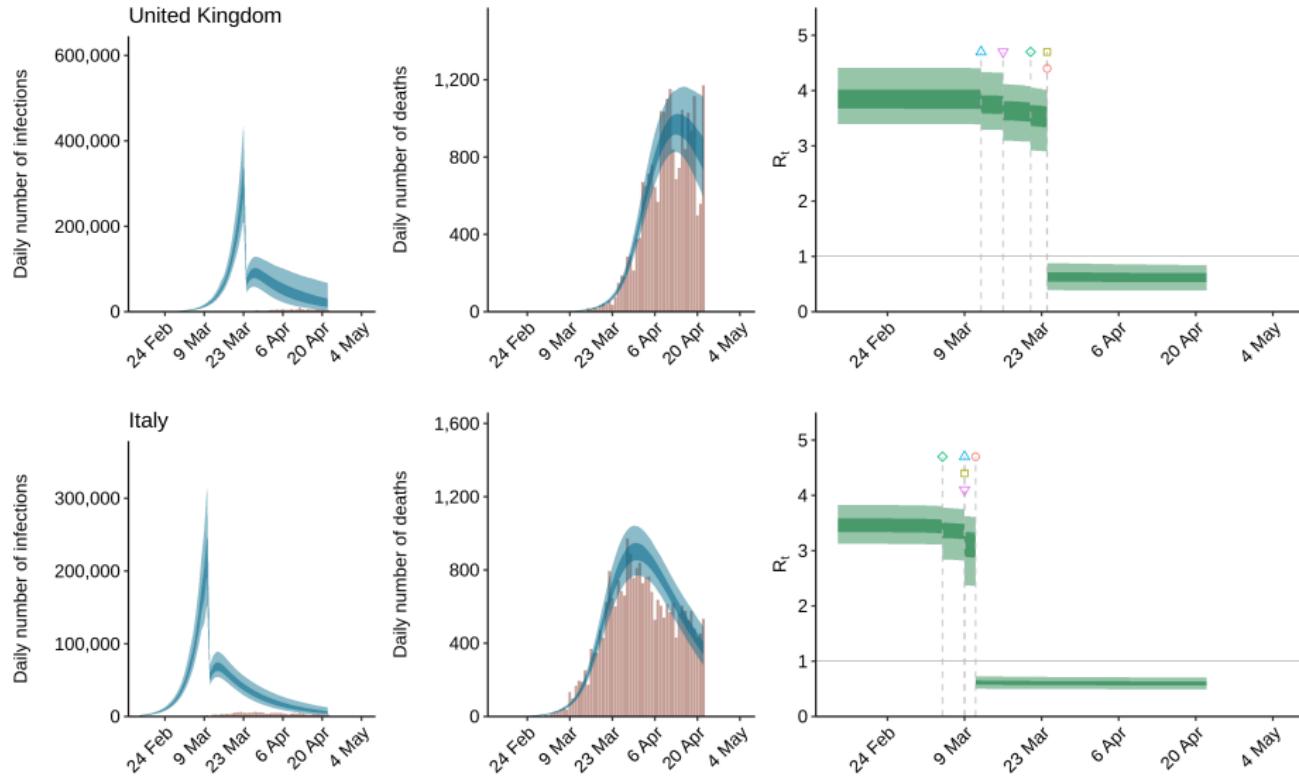
# Partial Pooling



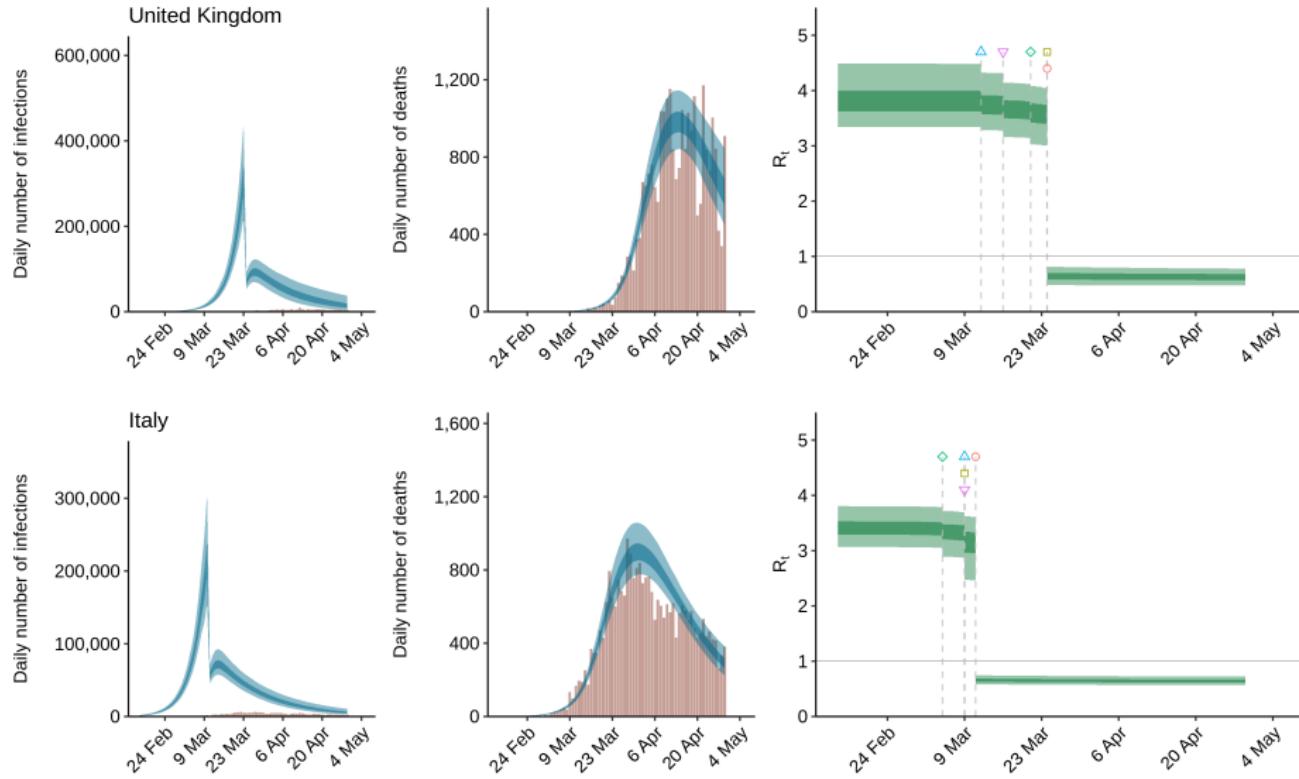
# Partial Pooling



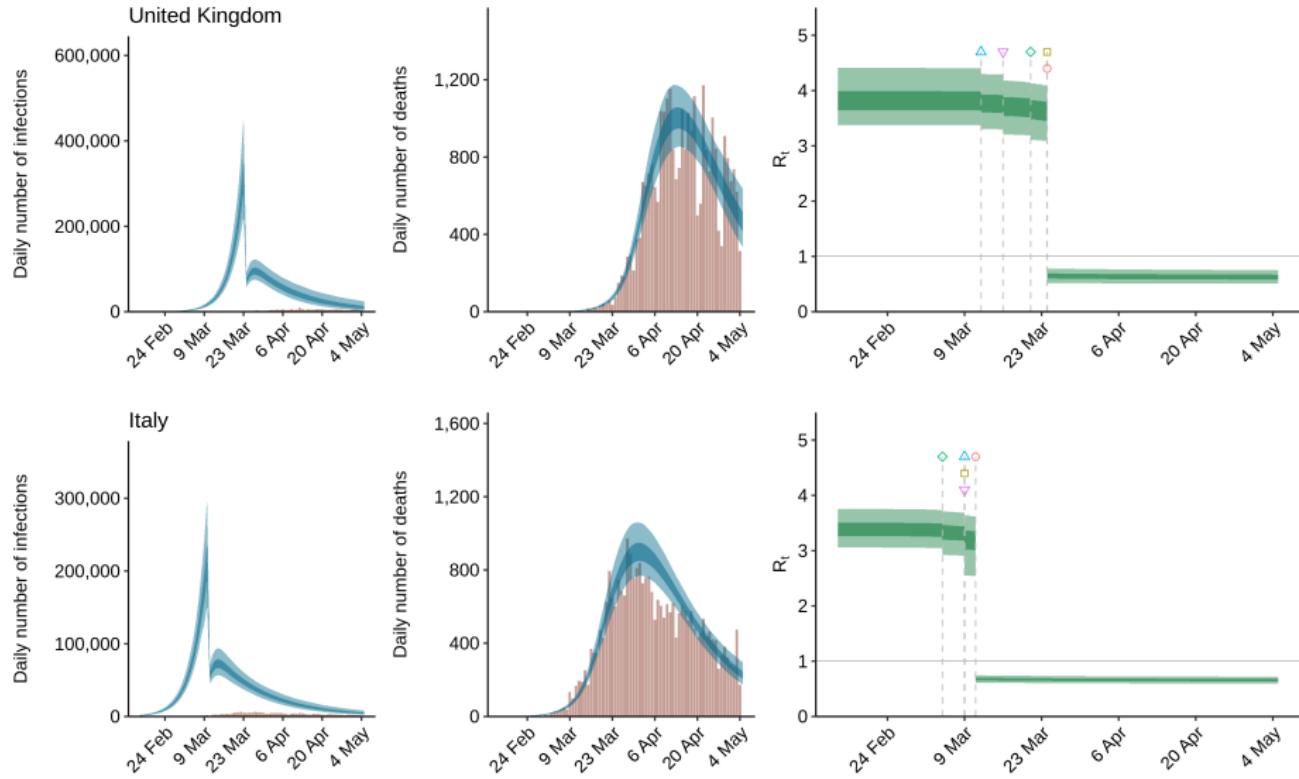
# Partial Pooling



# Partial Pooling

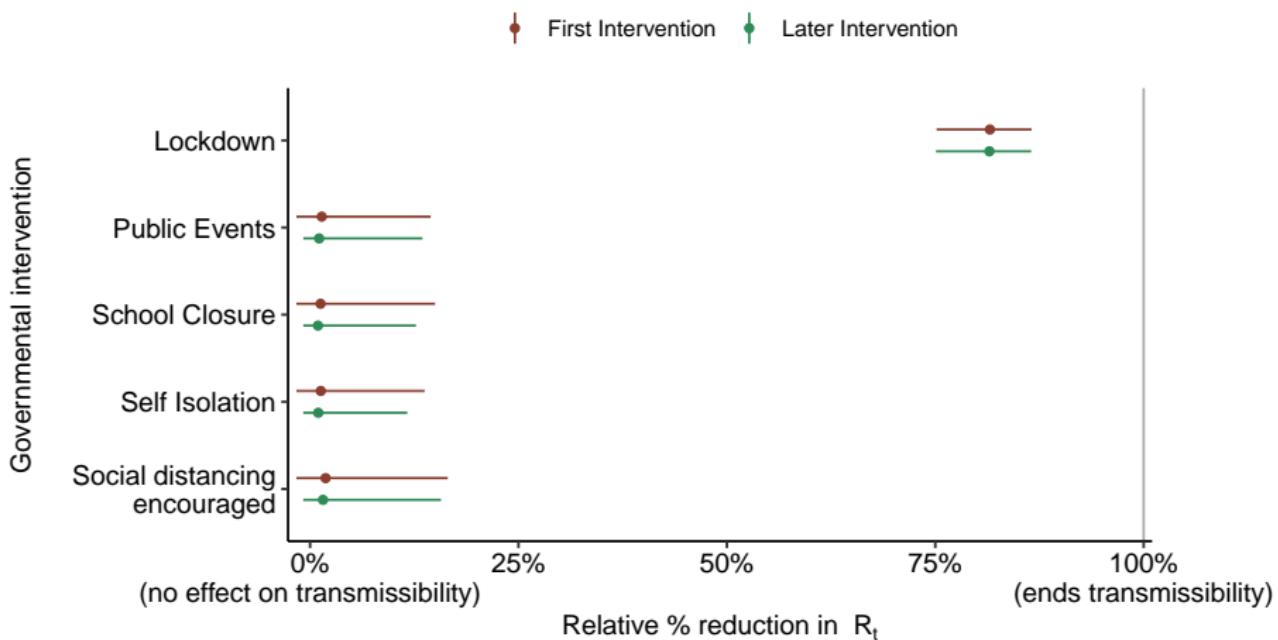


# Partial Pooling

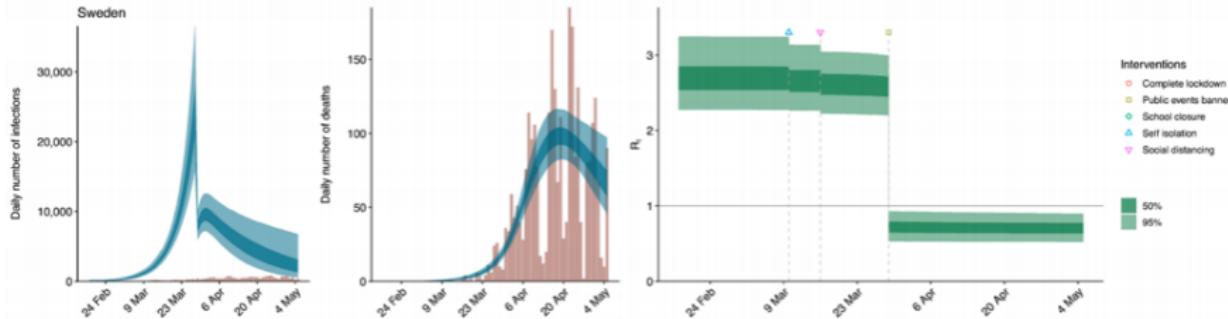


# A tale of Scandinavia

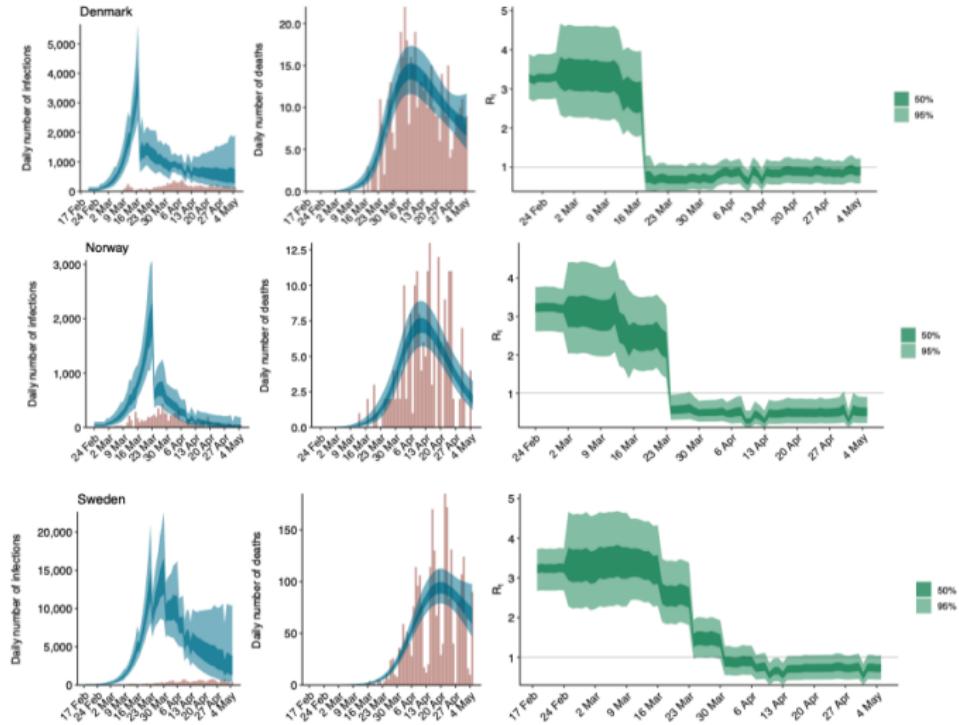
# What we found, and what we got wrong: A tale of Scandinavia 1



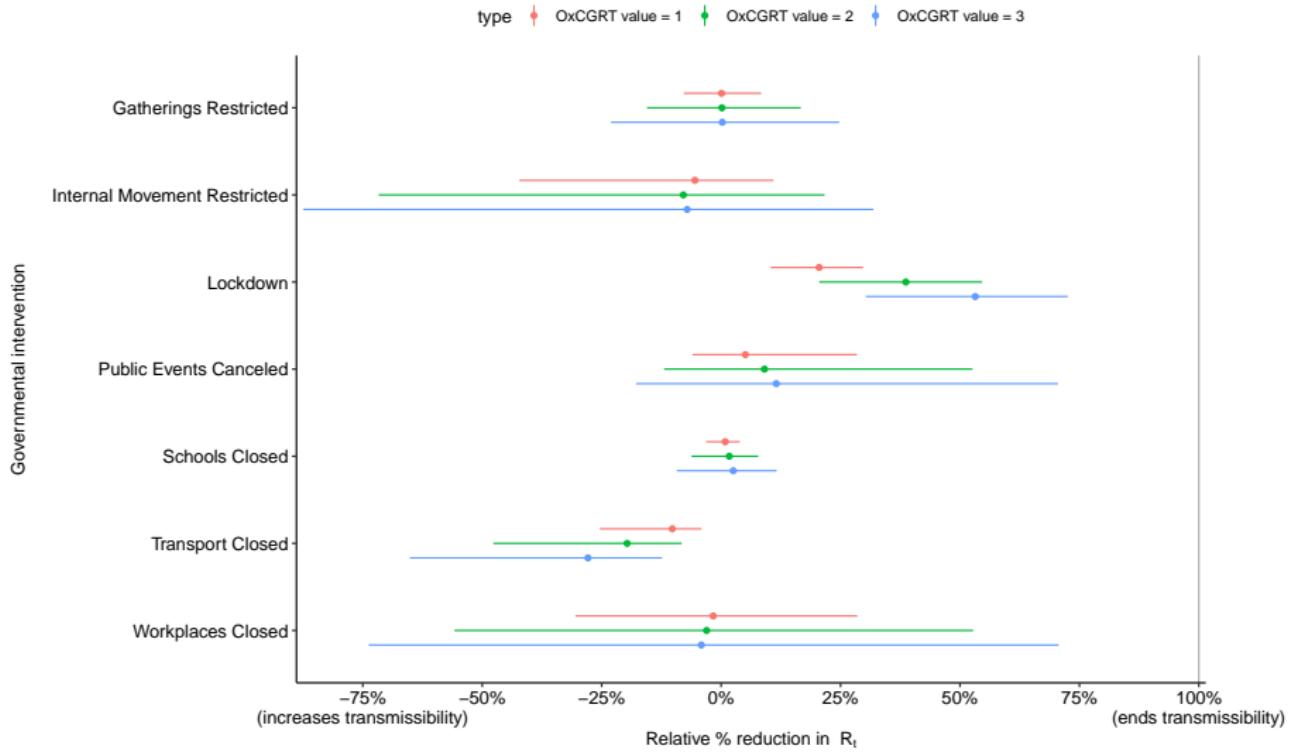
# A tale of Scandinavia



# A tale of Scandinavia

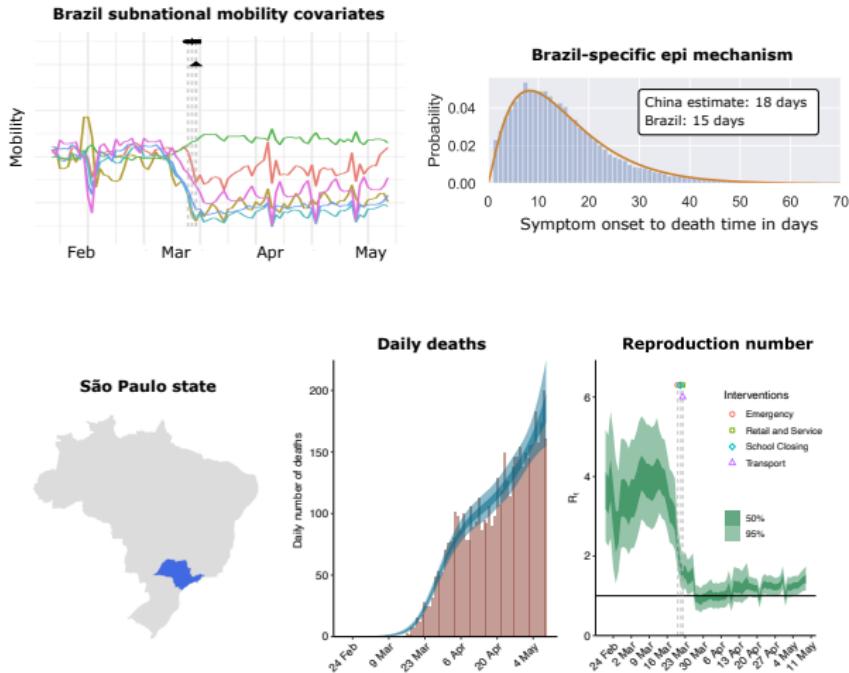


# An expanded analysis



# **Case studies: Brazil, USA, NYS**

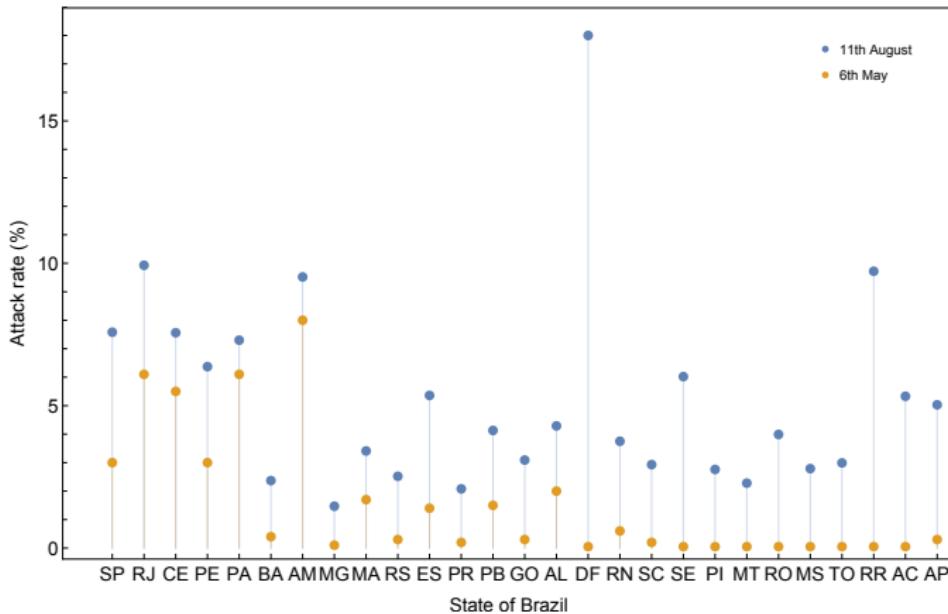
# Brazil: Subnational Model for $R_t$ <sup>3</sup>



Report (6 May): measures have been partially effective, but  $R_t > 1$  in 17/18 states analysed – the epidemic in Brazil is not under control.

<sup>3</sup>Thomas Mellan <https://vimeo.com/447268450>

# Brazil: Attack Rates



6th May – 6,633 deaths. 11th August – 101,049. The Brazilian epidemic is geographically heterogeneous and still nascent in many locations – thousands of deaths can still be averted.

# Brazil: Epidemiological Parameters

Onset-to-death distribution for the  $i^{\text{th}}$  state is given by

$$\text{Gamma}(\alpha_i, \beta_i), \quad (1)$$

with

$$\alpha_i \sim N(\alpha_{\text{Brazil}}, \sigma_1), \beta_i \sim N(\beta_{\text{Brazil}}, \sigma_2) \quad (2)$$

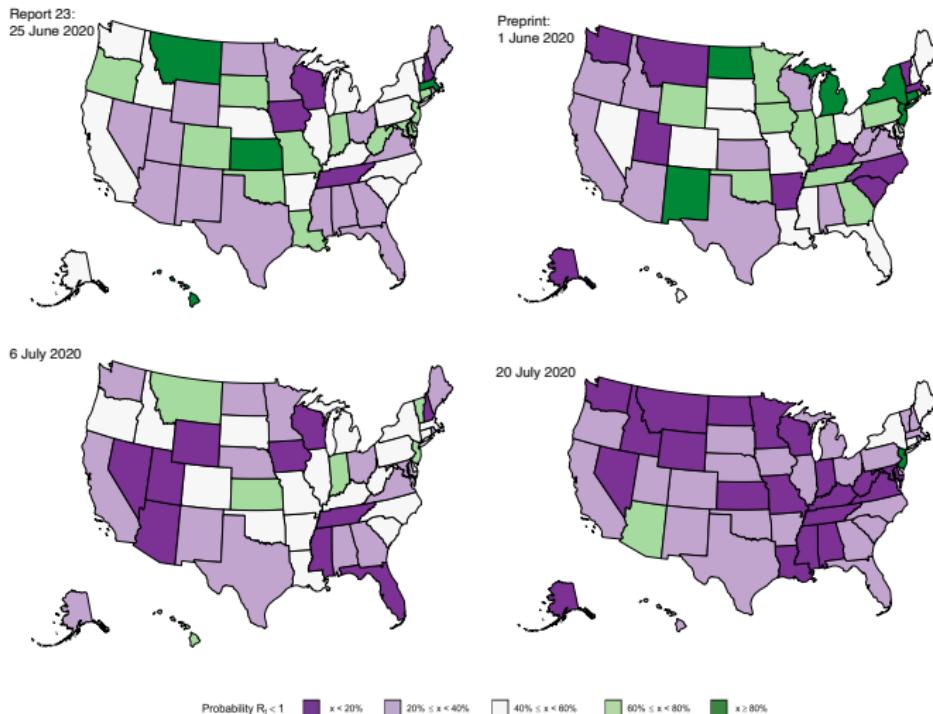
where  $\alpha_{\text{Brazil}}$  and  $\beta_{\text{Brazil}}$  denote the national level estimates, and

$$\sigma_1 \sim N^+(0, 1), \sigma_2 \sim N^+(0, 1). \quad (3)$$

Strong variation in fundamental epi parameter, time from onset of symptoms to death, across Brazil states:



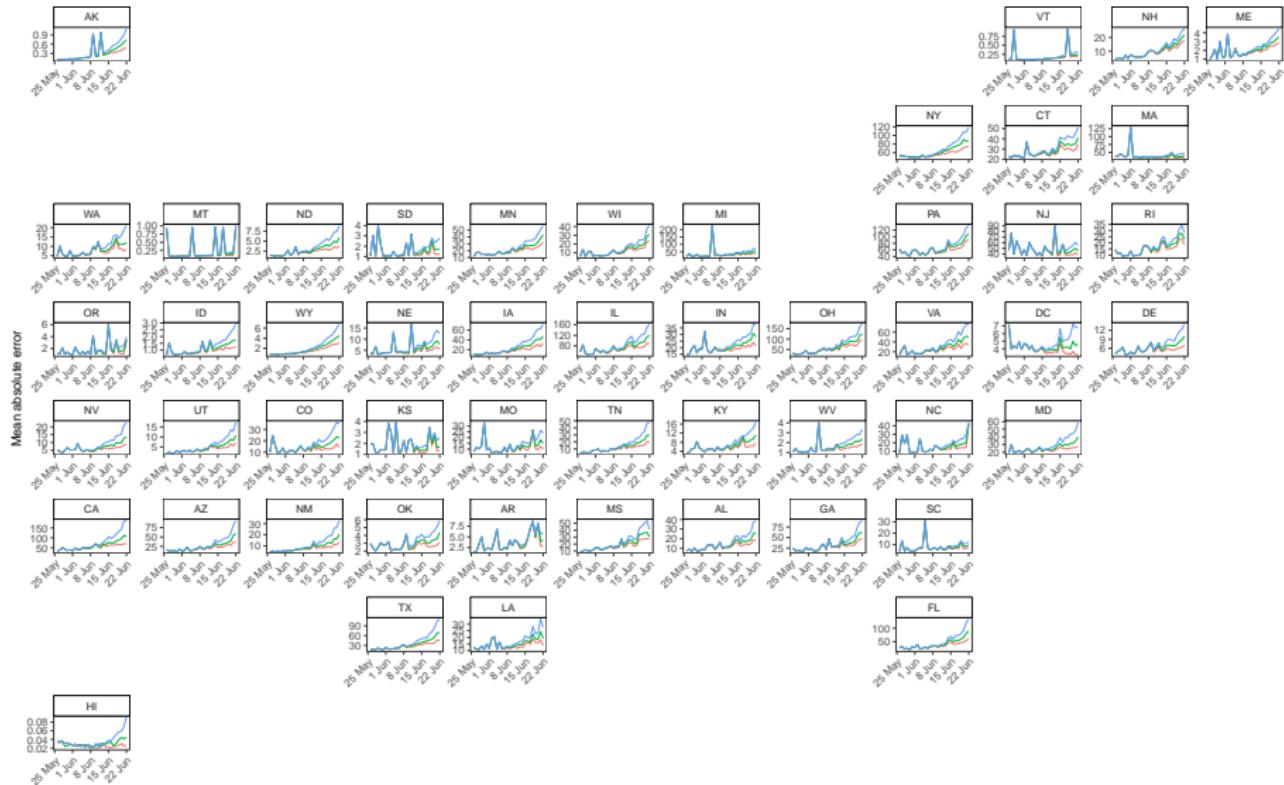
# USA: Transmission trends<sup>4</sup>



<sup>4</sup> Juliette Unwin: <https://vimeo.com/447237305>, <https://mrc-ide.github.io/covid19usa/>

## **USA: Error estimates**

mobility scenario — 0 — 20 — 40

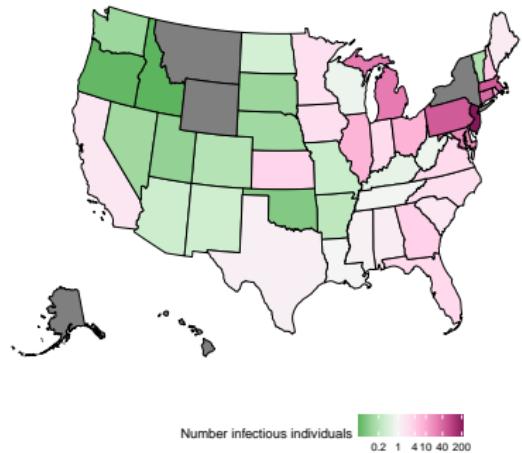


# New York State<sup>5</sup>

## 1. Reproduction number

## 2. Imported infections

Location	Rt	95% CI
Capital District	0.91	0.69-1.18
Central New York	0.77	0.51-1.10
Finger Lakes	0.93	0.63-1.23
Hudson Valley	1.07	0.82-1.37
Long Island	1.07	0.87-1.30
Mohawk Valley	0.95	0.73-1.20
New York City	0.89	0.63-1.18
North Country	0.74	0.45-1.08
Southern Tier	0.82	0.54-1.15
Western New York	1.040	0.762-1.33



1. Estimates given for 02/08/2020 using data from the New York State Governor's Office.
2. Produced using Emodo data from June 2020

<sup>5</sup>Helen Coupland: <https://vimeo.com/447237957>

# *epidemia*: a package for flexible epidemic modelling

# epidemia: Flexible Epidemic Modelling<sup>7</sup>

- **epidemia**<sup>6</sup> is an R-package for fitting hierarchical semi-mechanistic models similar to those just described.
- Inspired by **rstanarm** and **lme4** packages.
- Flexible models:
  - `R(country, date) ~ lockdown`
  - `R(country, date) ~ public_events + (lockdown|country)`
  - `R(country, date) ~ lockdown + rw(time=week, gr=country)`
- Choose models and priors for  $R_t$  and all observations.

epidemia 0.6.0



Reference

Articles ▾

Changelog



## epidemia

The `epidemia` package allows researchers to flexibly specify and fit Bayesian epidemiological models in the style of Flaxman et al. (2020). The package leverages R's formula interface to parameterize the time-varying reproduction rate as a function of covariates. Multiple populations can be modeled simultaneously with hierarchical models. The design of the package has been inspired by, and has borrowed from, `rstanarm` (Goodrich et al. 2020). `epidemia` uses `rstan` (Stan Development Team 2020) as the backend for fitting models.

### Links

Browse source code at  
[https://github.com/  
ImperialCollegeLondon/epidemia/](https://github.com/ImperialCollegeLondon/epidemia/)

### License

GPL-3

<sup>6</sup><https://imperialcollegelondon.github.io/epidemia>

<sup>7</sup>James Scott: <https://vimeo.com/447175473>

# epidemia: An Example

```
library(epidemia)
data(EuropeCovid)

# Collect args for epim
args <- EuropeCovid
args$group_subset <- c("Germany", "United_Kingdom")
args$algorithm <- "sampling"
args$sampling_args <- list(iter=1e3, seed=12345)

# model the reproduction number
args$rt <- epirt(
  formula = R(country, date) ~ public_events + (lockdown |
country),
  prior = rstanarm::normal(location=0, scale=0.5),
  prior_intercept = rstanarm::normal(scale=0.5),
  prior_covariance = rstanarm::decov(scale=0.5)
)
fit <- do.call(epim, args)
```

# epidemia: Example Output

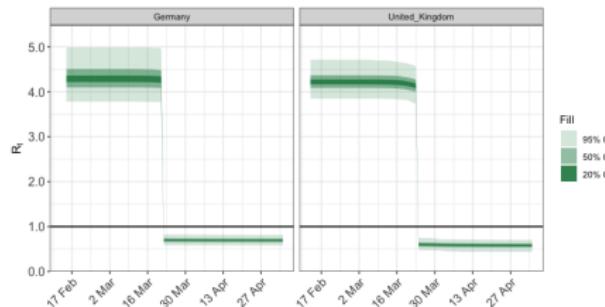
`print( fit )`

```
Rt regression parameters:
-----
                                         Median MAD_SD
R|(Intercept)                      0.3   0.4
R|public_events                     -0.5   0.5
R|b[(Intercept) country:Germany]    0.3   0.5
R|b[lockdown country:Germany]       -2.3   0.6
R|b[(Intercept) country:United_Kingdom] 0.3   0.5
R|b[lockdown country:United_Kingdom] -2.4   0.6
R|Sigma[country:(Intercept),(Intercept)] 0.3   0.4
R|Sigma[country:lockdown,(Intercept)] -0.2   0.6
R|Sigma[country:lockdown,lockdown]    2.8   2.0

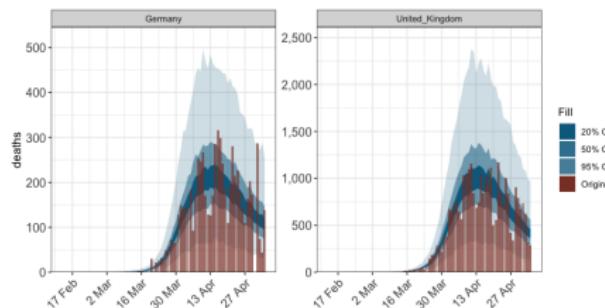
deaths regression parameters:
-----
                                         Median MAD_SD
deaths|(Intercept)                  0.0   0.0
deaths|reciprocal dispersion        4.8   0.8

Other model parameters:
-----
                                         Median MAD_SD
seeds[Germany]                     17.4   7.3
seeds[United_Kingdom]               35.5  12.1
tau                                31.6  18.8
```

`plot_rt( fit )`



`plot_obs(fit , type = "deaths")`



# Age structure and contact patterns

# Age structure and contact patterns (1)<sup>8</sup>

**Objective:** Extend the model to include age by integrating highly structured human contact.

## Data:

gPolymod contact survey [7] → Estimate baseline contact matrix in every U.S. State. Mobility data by age from Foursquare location technology Death by age data from DoH websites

## Model's updates:

- 1 Time-varying contact patterns  
→ Baseline contact matrix weighted by the mobility trends.
- 2 Susceptibility by age  
→ Probability of transmission given a contact is proportional to susceptibility.
- 3 Infection fatality rate by age

---

<sup>8</sup> Mélodie Monod: <https://vimeo.com/446922651>

# Age structure and contact patterns (2)

## Computationally:

- 1 Use of CmdStan's `reduce_sum` functionality to parallelize across states.
- 2 ~ 24 hours to run 8 chains with 1,500 iterations (1,000 warmup) in parallel.

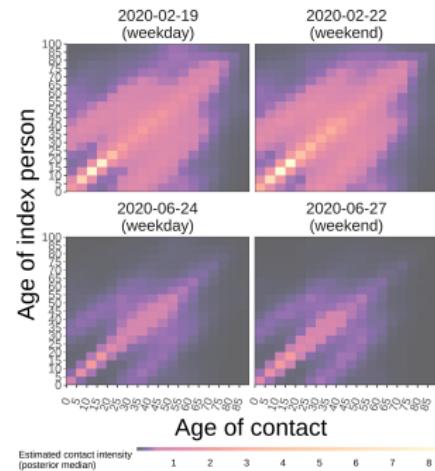
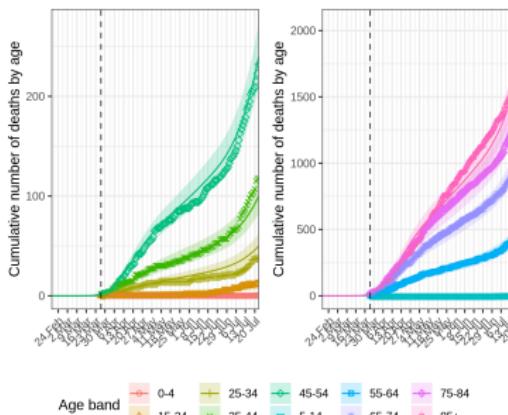


Figure: Death by age fit, and time-varying contact matrix in Florida

# Where do we go next?

# Where do we go next?

- Lockdown works, but it should never have to be imposed nationally again
- If it does, it's an abject failure of control
- We need a playbook of interventions to impose on a local scale, lockdown included
- Fine scale spatio-temporal models
- Causal, not associational analyses
- Fine scale compartments such as schools, stages of secondary care
- Improving epi parameters

# It's a team effort, we're hiring!<sup>9</sup>

- (alphabetical) Tresnia Berah, Samir Bhatt, Valerie Bradley, Helen Coupland, Seth Flaxman, Axel Gandy, Iwona Hawryluk, Henrique Hoeltgebaum, Jonathan Ish-Horowicz, Thomas Mellan, Swapnil Mishra, Mélodie Monod, Oliver Ratmann, Jamie Scott, Ricardo Schnekenberg, Juliette Unwin, Michaela Vollmer, Charlie Whittaker, Harrison Zhu.
- Come join us! we're fun to work with

---

<sup>9</sup> postdocs—full or part-time—and PhD students; email [Seth Flaxman](#) or [Samir Bhatt](#)

# References

# References I

- Flaxman, Mishra, Gandy et al, “Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe”. *Nature*. 2020.
- Unwin, Mishra, Bradley, Gandy et al, “State-level tracking of COVID-19 in the United States” *medrXiv*. 2020.
- Mellan, Henrique, Mishra, Whittaker, Schnakenberg et al, “Estimating COVID-19 cases and reproduction number in Brazil”. *medrXiv*. 2020.
- Mishra, Berah, Mellan et al, “On the derivation of the renewal equation from an age-dependent branching process: an epidemic modelling perspective”. *arXiv*. 2020.

## References II

-  Hawryluk, Mellan, Hoeltgebaum et al, "Inference of COVID-19 epidemiological distributions from Brazilian hospital data". *arXiv*. 2020.
-  Candido, Claro, de Jesus, Souza, Moreira, Dellicour, Mellan et al, "Evolution and epidemic spread of SARS-CoV-2 in Brazil". *Science*. 2020.
-  Mossong et al.  
*Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases*.  
Plos Medicine

# Acknowledgements

This research was partly funded by the The Imperial College COVID-19 Research Fund and supported by Centre funding from the UK Medical Research Council under a concordat with the UK Department for International Development, the NIHR Health Protection Research Unit in Modelling Methodology and Community Jameel. S.B. acknowledges the NIHR BRC Imperial College NHS Trust Infection and COVID themes, the Academy of Medical Sciences Springboard award and the Bill and Melinda Gates Foundation. S.F. acknowledges funding provided by the EPSRC. We thank F. Valka for creating our website, and A. Gelman and the Stan team for helpful discussions. We acknowledge the resources provided by Cirrus UK National Tier-2 HPC Service at EPCC (<http://www.cirrus.ac.uk>) funded by the University of Edinburgh and EPSRC (EP/P020267/1), and cloud compute time donated by Microsoft and Amazon. We also thank the New York State Governor's Office and Emodo, Inc.