EpiNow2

Estimate real-time case counts and time-varying epidemiological parameters

Sam Abbott, EpiForecasts, 22 September 2021
@seabbs

cmmid
centre for the mathematical modelling of infectious diseases

LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

There are many like it, but this one is mine (ours)

Some slides “borrowed” from Sebastian Funk and Nikos Bosse

Authors

Sam Abbott. Author, maintainer.

Joel Hellewell. Author.

Katharine Sherratt. Author.

Katelyn Gostic. Author.

Joe Hickson. Author.

Hamada S. Badr. Author.

Michael DeWitt. Author.

Robin Thompson. Contributor.

Sophie Meakin. Contributor.

James Munday. Contributor.

Nikos Bosse. Contributor.

Paul Mee. Contributor.

Peter Ellis. Contributor.

Pietro Monticone. Contributor.

EpiForecasts. Author.

Sebastian Funk. Author.

Talk outline

1. Aims
2. Background
3. Case Model
4. Death model
5. Data requirements and processing
6. Implementation
7. Performance
8. Future work
9. Questions

Aim

Develop a toolkit for retrospective and real-time tracking of infectious disease outbreaks that utilises the benefits (if any) of a hybrid statistical and mechanistic approach.

Background

1. Models developed as needed from February 2020 with several versions

2. Nowcasting and effective reproduction number estimates published each day in over 1000 locations: epiforecasts.io/covid

3. Forecasts submitted to SPI-M (UK government advisory committee), the CDC, German and Poland and ECDC forecasting hubs

4. Estimates used in numerous downstream analyses to draw inferences about the dynamics of the pandemic

5. Part of a project to evaluate methods for forecasting infectious diseases and to develop tooling to do so

Case model

Objectives

1. Develop a model that can be used for real-time surveillance, nowcasting, and short-term forecasting.

2. Include known epidemiological structure of the infection and reporting process.

3. The model should include a parameter that is referenced to the infection process and that can be used to compare disparate surveillance data sources (here the effective reproduction number).

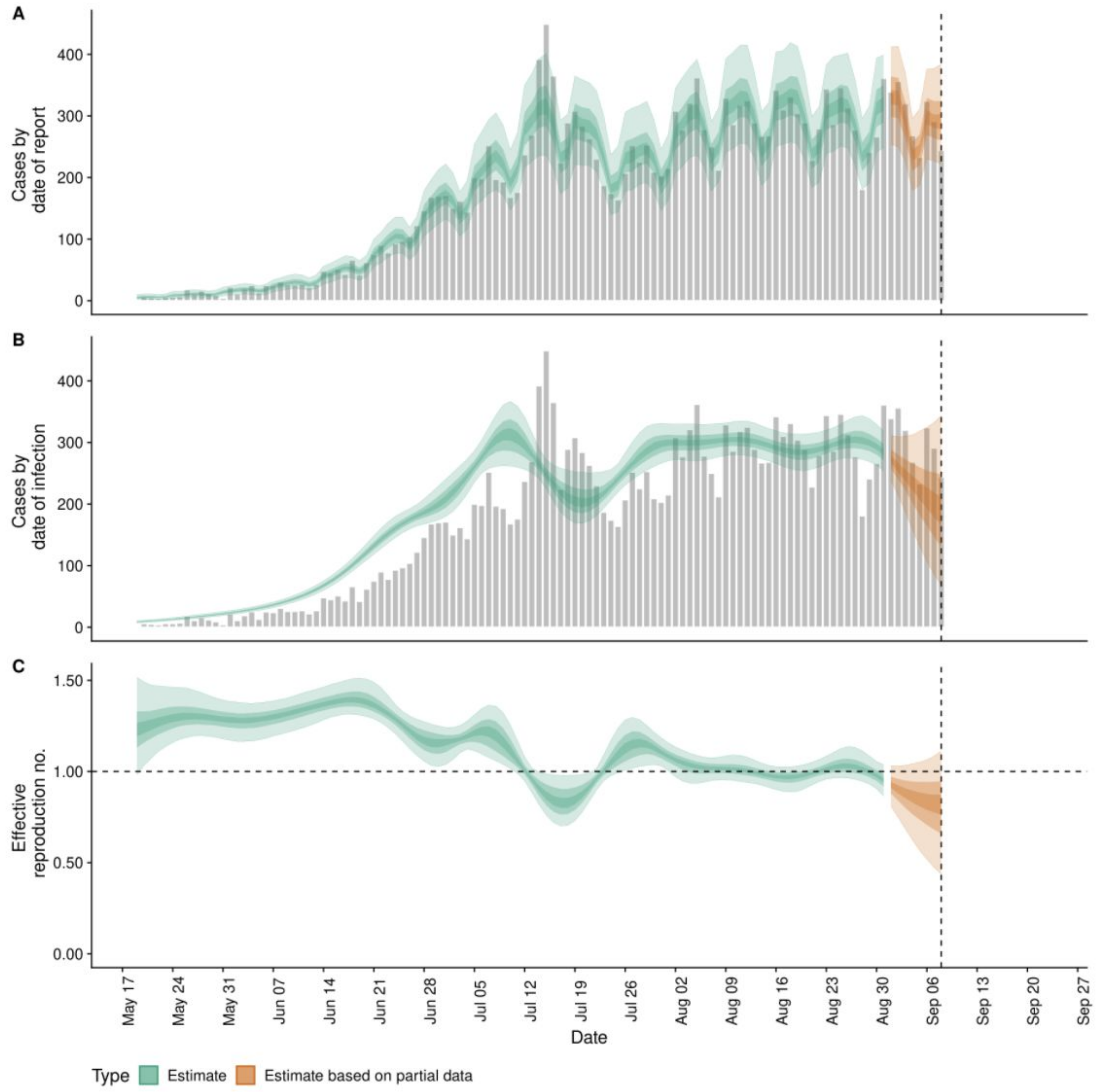
4. The model should ideally capture changes in trend as quickly as possible.

Sherratt, K. et al 2021. Philosophical Transactions of the Royal Society B: Biological Sciences.
https://doi.org/10.1098/rstb.2020.0283

Our approach

- Bayesian approach combining nowcasting and *R* estimation
- Uncertain generation interval estimates from Singapore allowing for negative serial intervals
- Latent process for estimating I_t
- Negative binomial reporting with multiplicative day-of-the-week effect
- R_t estimates with correlation between R_{t+1} and R_t based on Gaussian Process prior
- All implemented in Stan and as open-source R package https://epiforecasts.io/EpiNow2/

Confirmed cases, their estimated date of report, date of infection, and time-varying reproduction number estimates



Abbott et al., 2020, *Wellcome Open Res*

https://epiforecasts.io/covid/posts/national/united-kingdom/

The model

$$\log R_t = \log R_{t-1} + \mathrm{GP}_t$$

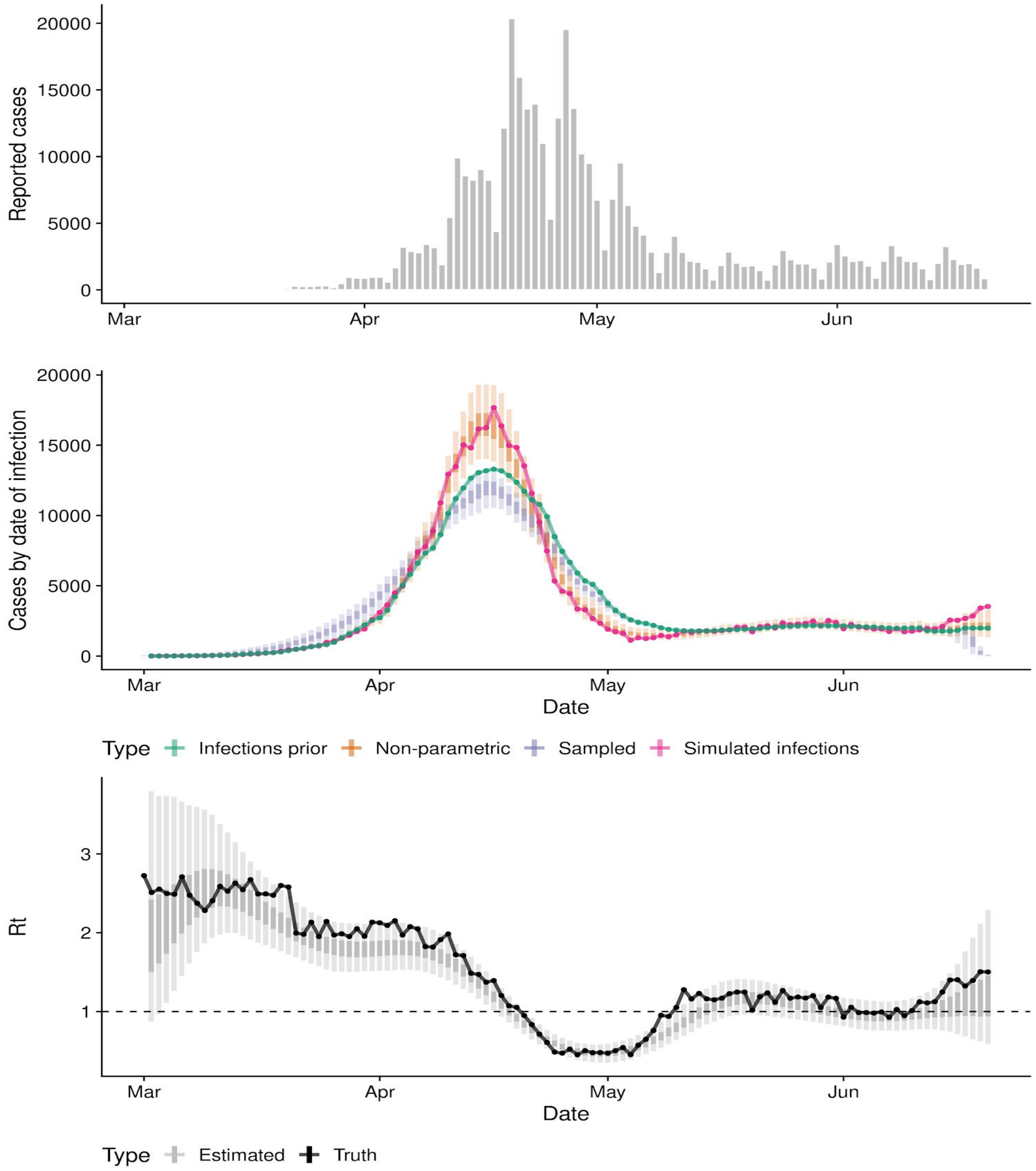
$$I_t = R_t \sum_{\tau=1}^{15} w(\tau|\mu_w, \sigma_w) I_{t-\tau}$$

$$O_t = \sum_{\tau=0}^{15} \xi_O(\tau|\mu_{\xi_O}, \sigma_{\xi_O}) I_{t-\tau}$$

$$D_t = \alpha \sum_{\tau=0}^{15} \xi_D(\tau|\mu_{\xi_D}, \sigma_{\xi_D}) O_{t-\tau}$$

$$C_t \sim \mathrm{NB}\left(\omega_{(t \mod 7)} D_t, \phi\right)$$

Method: https://doi.org/10.12688/wellcomeopenres.16006.2
Stan code: https://git.io/JUxRt

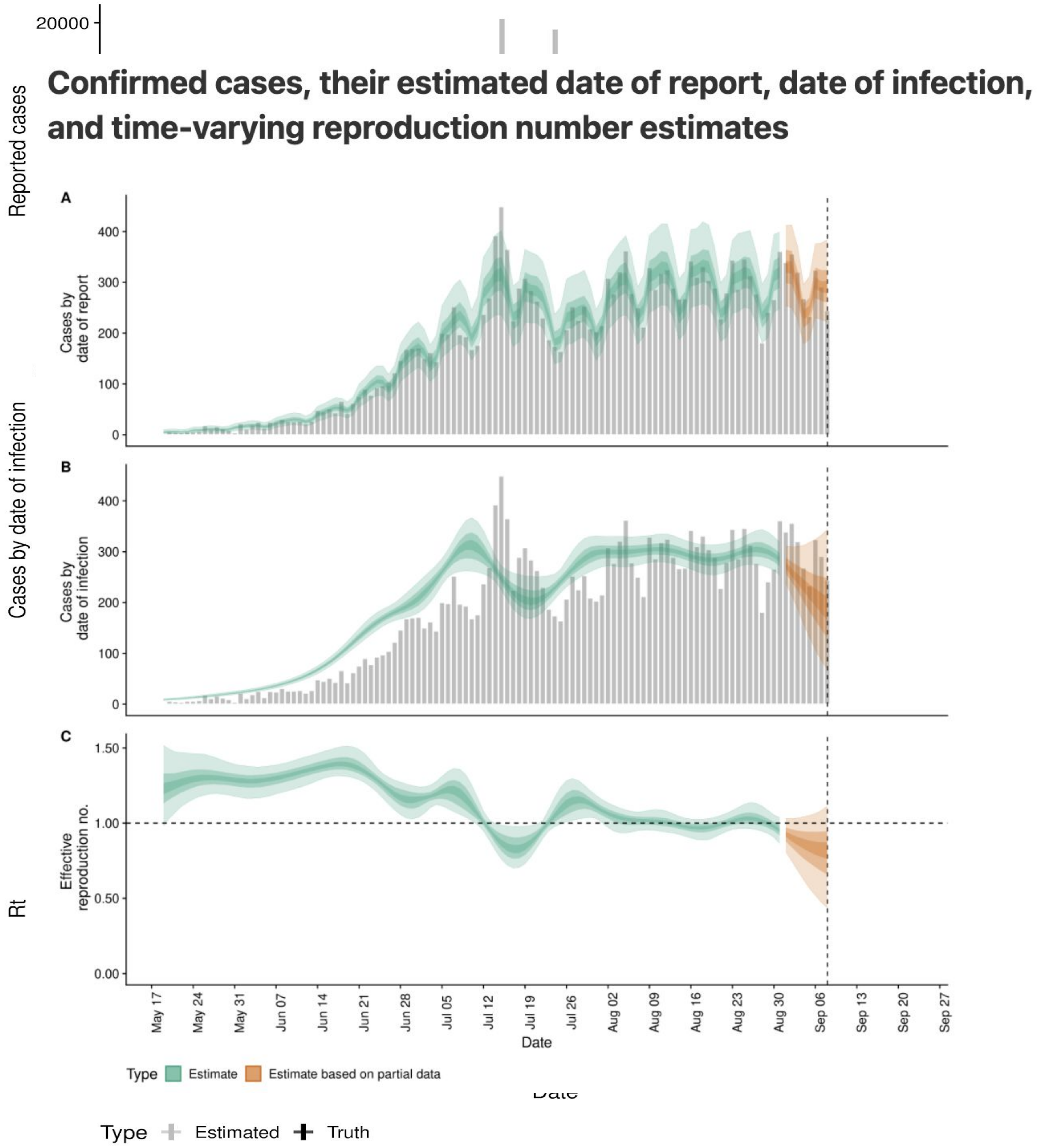


https://github.com/epiforecasts/backcalc/blob/master/report.md

Where,

$$I_t = I_0 \exp(rt)$$
$$I_0 \sim \mathcal{LN}(\log I_{obs}, 0.2)$$
$$r \sim \mathcal{LN}(r_{obs}, 0.2)$$

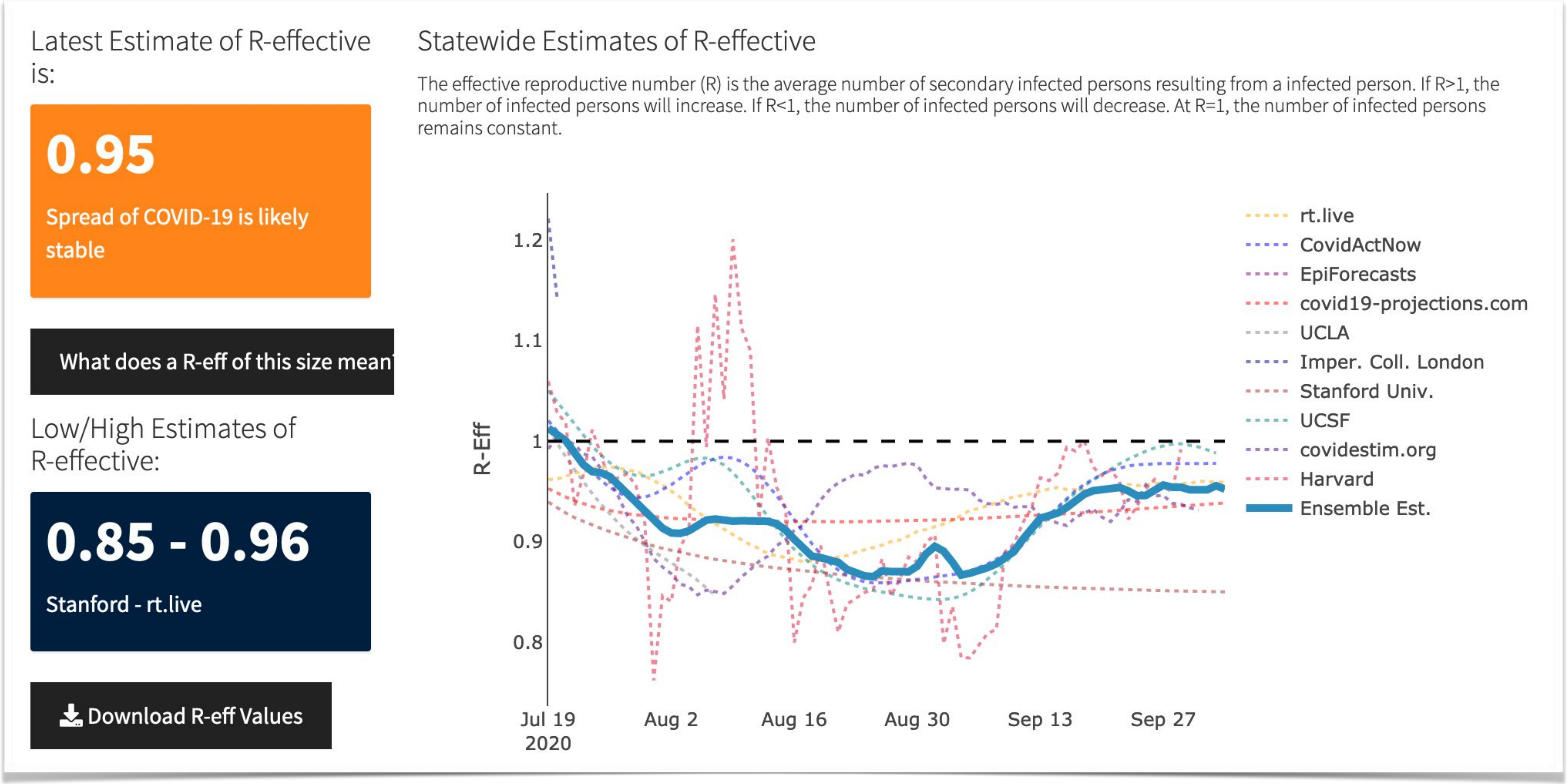
$$w \sim \mathcal{G}(\mu_w, \sigma_w)$$
$$\xi_O \sim \mathcal{LN}(\mu_{\xi_O}, \sigma_{\xi_O})$$
$$\xi_D \sim \mathcal{LN}(\mu_{\xi_D}, \sigma_{\xi_D})$$



Method: https://doi.org/10.12688/wellcomeopenres.16006.2
Stan code: https://git.io/JUxRt

https://epiforecasts.io/covid/posts/national/united-kingdom/

How can we tell which approach works best?



https://calcat.covid19.ca.gov/cacovidmodels/

Forecasting based on reproduction numbers

1. Forecasting a useful aim in and of itself but also a potential method for choosing the optimal real-time surveillance tool

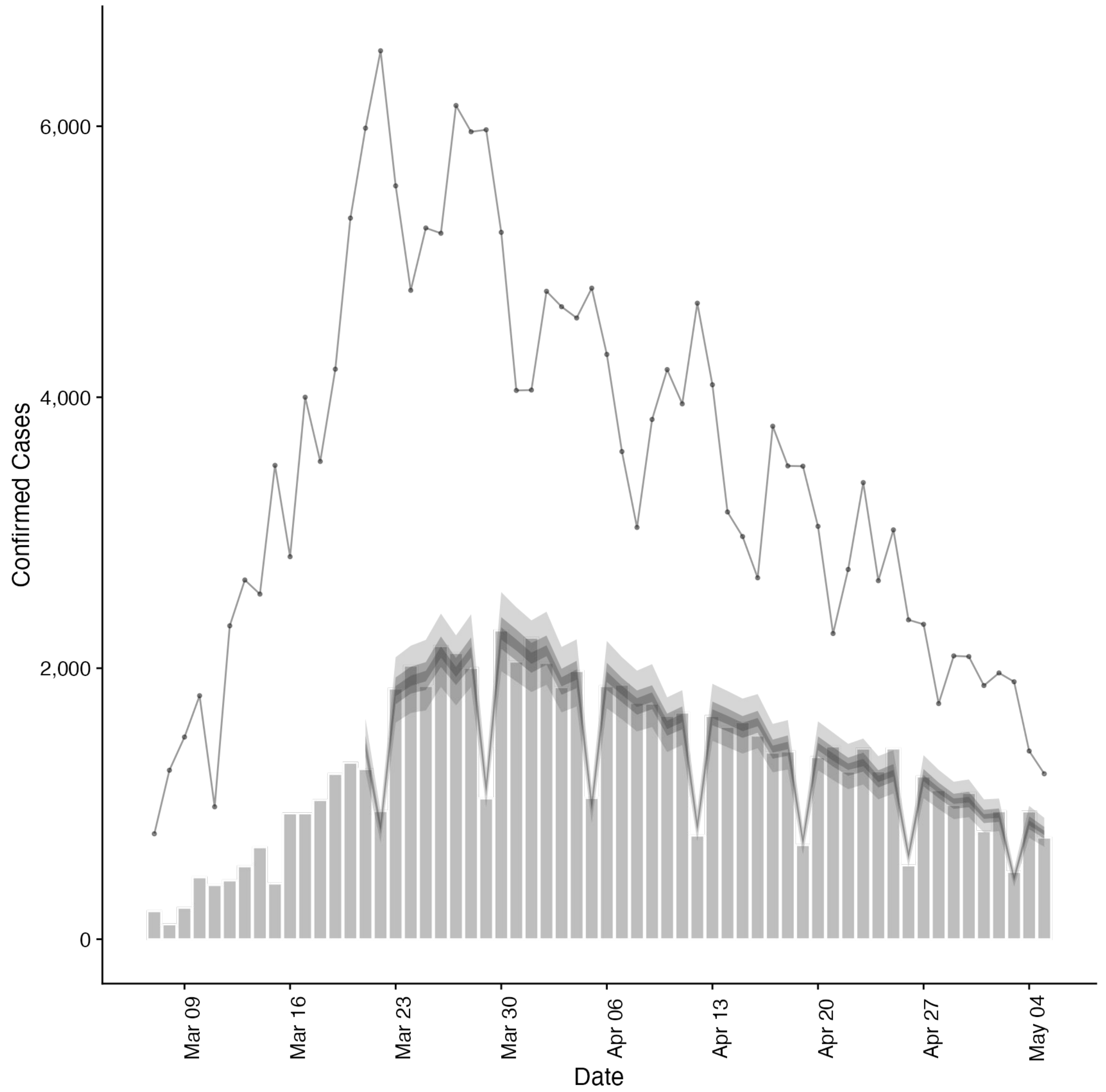
2. Simplistic assumption of no change in the reproduction number beyond the forecast horizon

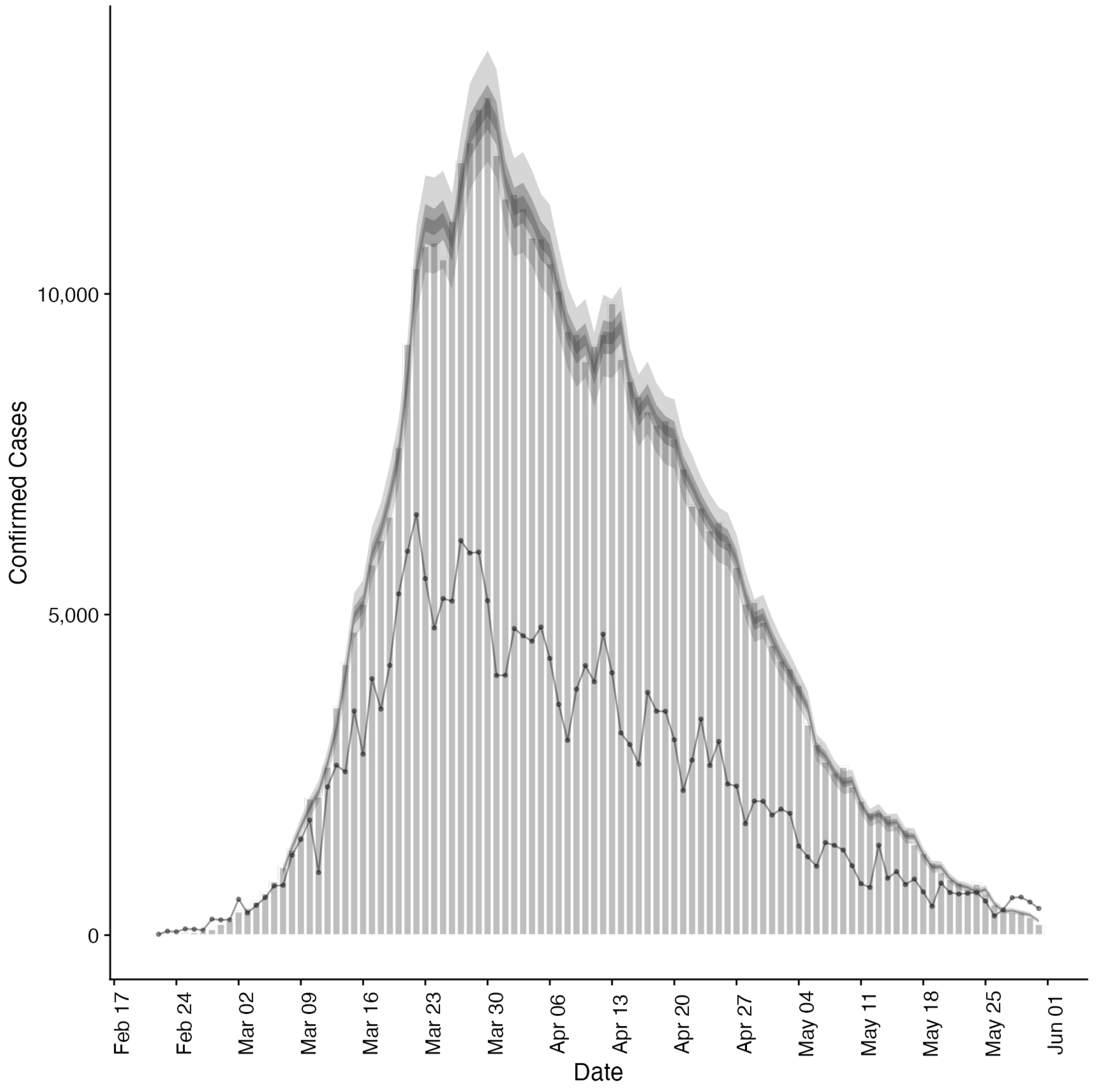
3. Modification of effective reproduction based on total susceptible population to prevent implausible forecasts

Death forecast

Our approach

- Bayesian
- Prior specifying the log normal delay from the primary observation to the secondary observation
- Prior specifying the scaling of the primary to secondary observation
- Options for non-cumulative/cumulative relationship
- Negative binomial reporting with multiplicative day-of-the-week effect
- All implemented in Stan and as open-source R package https://epiforecasts.io/EpiNow2/





https://epiforecasts.io/EpiNow2/reference/estimate_secondary.html

The model

$$D_t \sim \mathrm{NB}\left(\omega_{(t \mod 7)} \alpha \sum_{\tau=0}^{30} \xi(\tau|\mu,\sigma) C_{t-\tau}, \phi\right)$$

$$\frac{\omega}{7} \sim \mathrm{Dirichlet}(1,1,1,1,1,1,1)$$

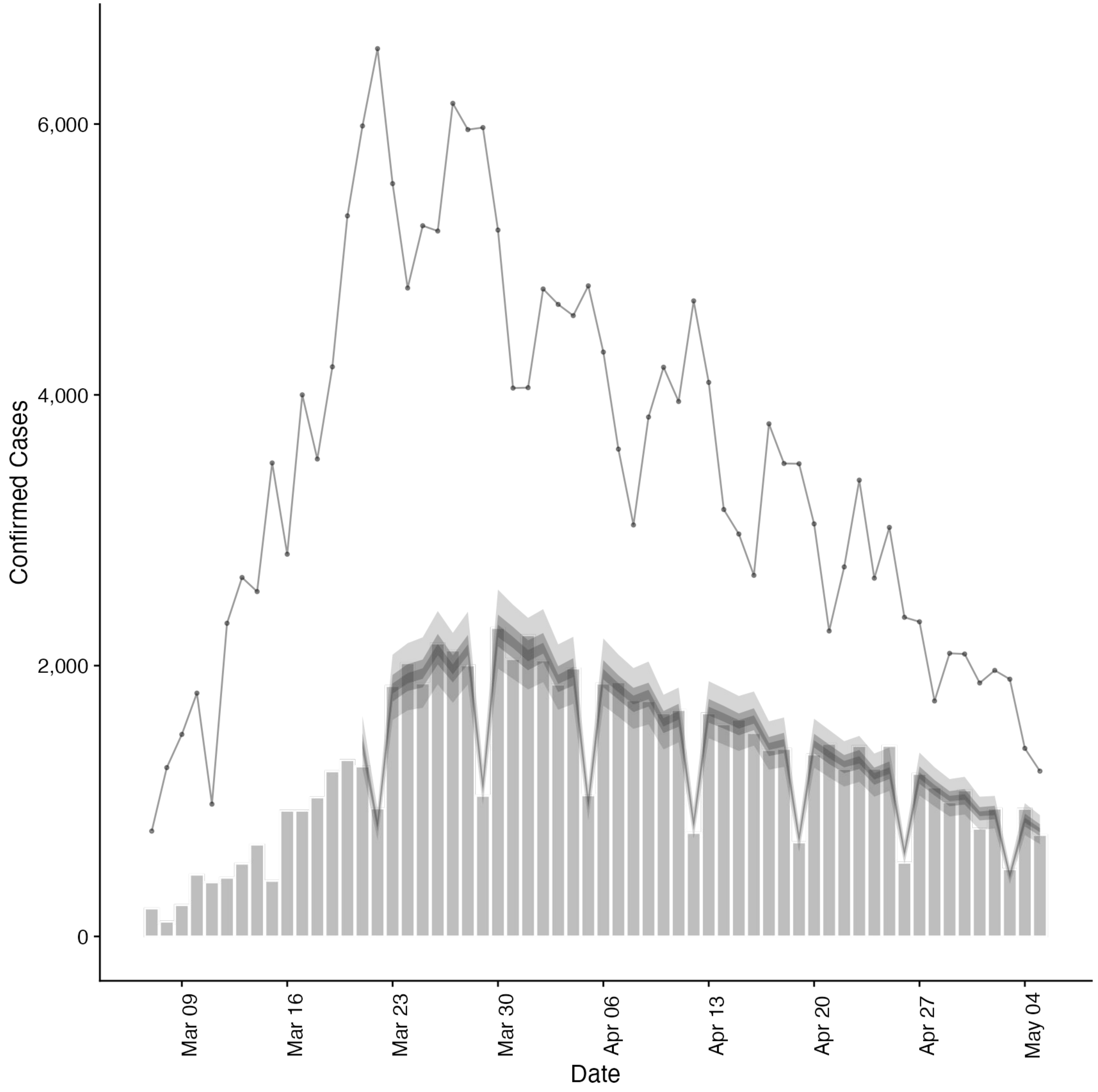
$$\alpha \sim \mathcal{N}(0.01, 0.02)$$

$$\xi \sim \mathcal{LN}(\mu, \sigma)$$

$$\mu \sim \mathcal{N}(2.5, 0.5)$$

$$\sigma \sim \mathcal{N}(0.47, 0.2)$$

$$\phi \sim \frac{1}{\sqrt{\mathcal{N}(0,1)}}$$



Method: Bosse et al. (in pres)
Stan code: https://git.io/Jz04S

https://epiforecasts.io/EpiNow2/reference/estimate_secondary.html

Data requirements

Data requirements

Case forecast

1. Daily case notifications (though aggregations also supported)

2. An estimate of the generation time (optional)

3. An estimate of the incubation period (optional)

4. An estimate of the delay between onset and report (optional)

Data requirements

Death forecast

1. Daily case notifications

2. Posterior samples from a forecast of future case notifications

3. Daily death notifications (or an aggregation up to weekly)

4. An estimate of the delay between case and death notification (optional)

Data processing

Data processing

None!

1. No automated outlier handling

2. No manual outlier handling

3. No aggregation of the daily data

4. No seasonality adjustment

Performance

Forecast scores

Scores separated by target and forecast horizon. Only models with submissions in each of the last 4 weeks are shown.

Cases | Deaths

1 week ahead horizon | 2 weeks ahead horizon | 3 weeks ahead horizon | 4 weeks ahead horizon

CSV | Excel

Search:

model	n	n_loc	rel_wis	rel_ae	50% Cov.	95% Cov.	bias
MUNI-ARIMA	598	32	0.83	0.81	0.59	0.92	-0.03
LANL-GrowthRate	823	32	0.86	0.84	0.48	0.87	-0.22
EuroCOVIDhub-ensemble	853	32	0.9	0.91	0.47	0.88	0.05
epiforecasts-EpiExpert_direct	93	18	0.94	0.88	0.28	0.62	0.29
EuroCOVIDhub-baseline	853	32	1	1	0.39	0.89	0.05
RobertWalraven-ESG	853	32	1.02	0.93	0.32	0.63	-0.24
IEM_Health-CovidProject	839	32	1.15	1.12	0.29	0.7	0.14
ILM-EKF	852	32	1.47	1.36	0.4	0.83	0.04
epiforecasts-EpiNow2	851	32	1.49	1.39	0.44	0.84	-0.02
USC-SIkJalpha	853	32	1.56	1.24	0.26	0.4	-0.01
Karlen-pypm	268	29	2	1.78	0.37	0.79	0.14
UVA-Ensemble	825	32		1.83	0.39	0.7	-0.02

Forecast scores

Scores separated by target and forecast horizon. Only models with submissions in each of the last 4 weeks are shown.

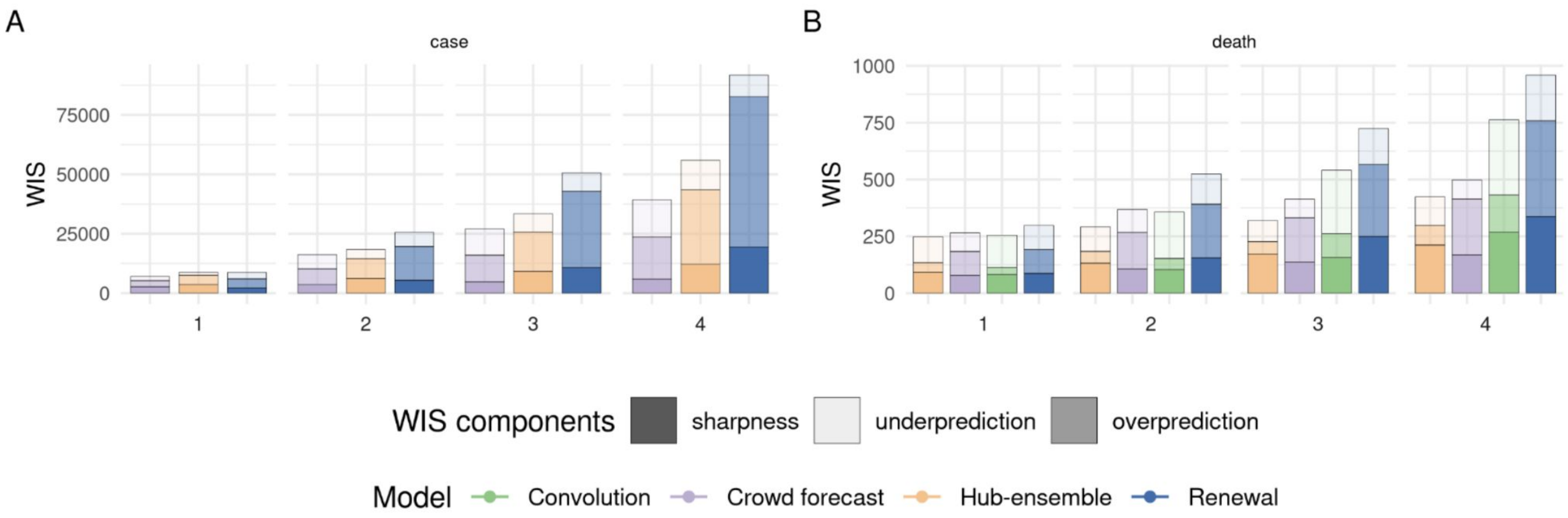
Cases Deaths

1 week ahead horizon 2 weeks ahead horizon 3 weeks ahead horizon 4 weeks ahead horizon

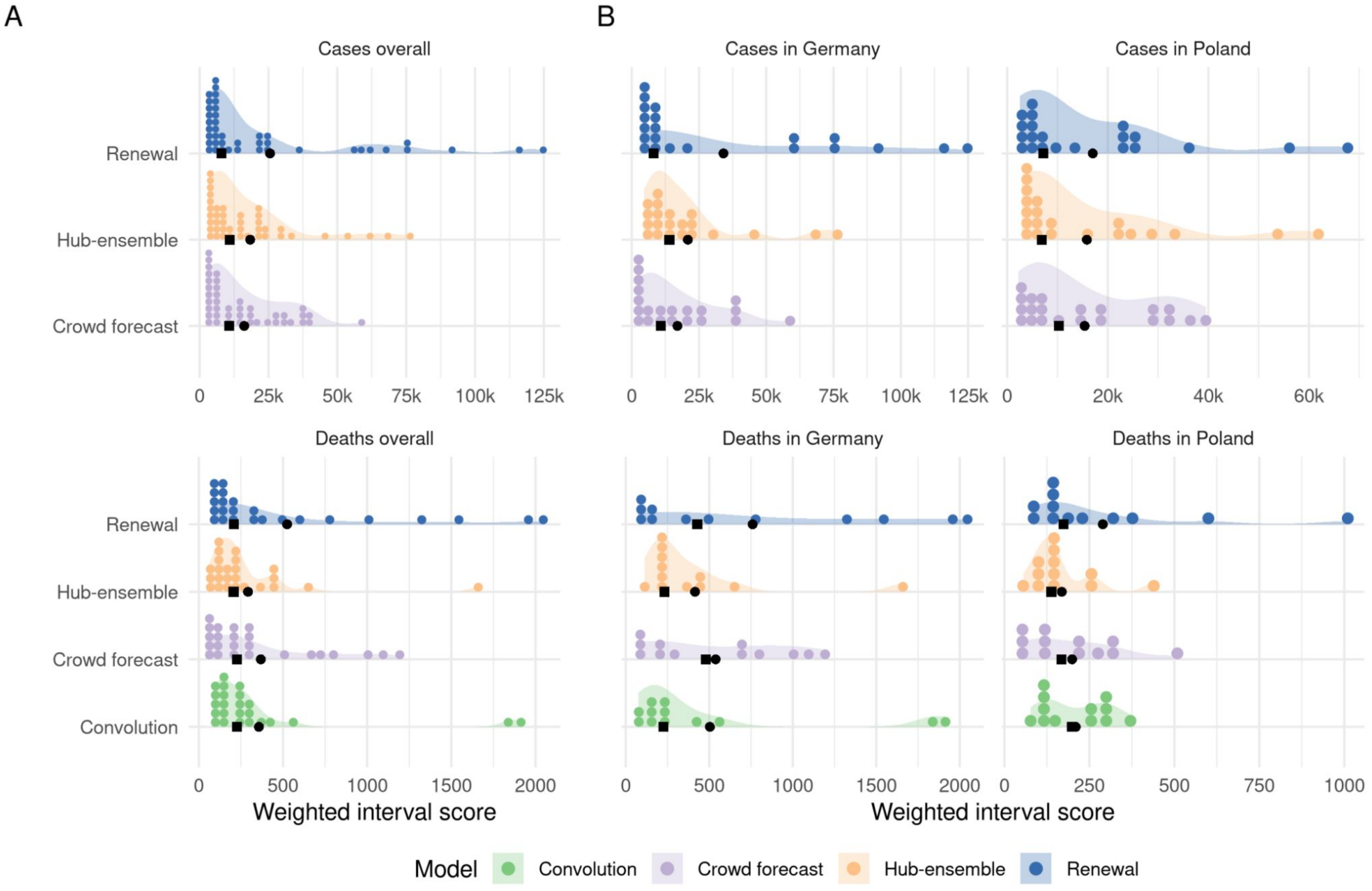
CSV Excel

Search:

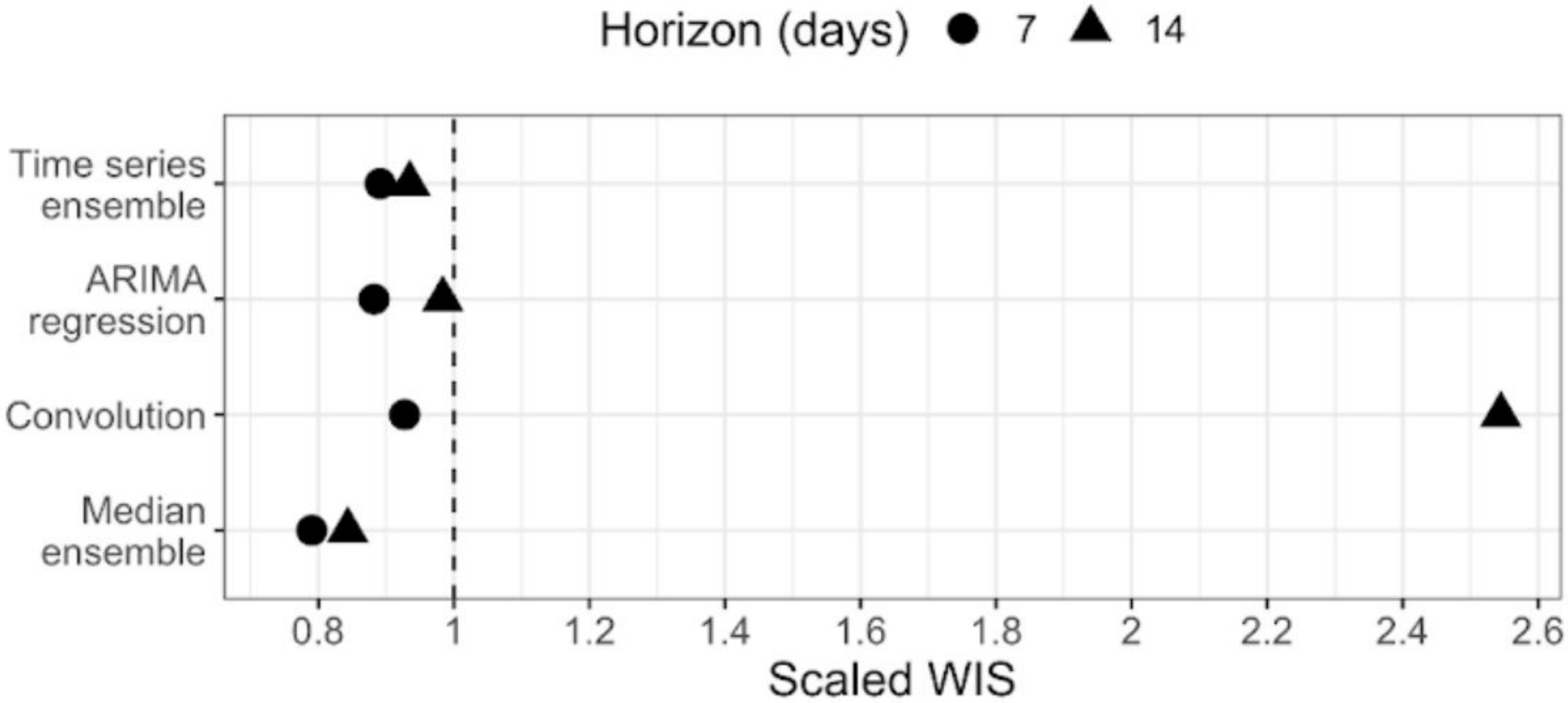
model	n	n_loc	rel_wis	rel_ae	50% Cov.	95% Cov.	bias
EuroCOVIDhub-ensemble	858	32	0.47	0.5	0.66	0.97	0.1
LANL-GrowthRate	827	32	0.59	0.62	0.38	0.79	-0.07
Karlen-pypm	268	29	0.62	0.68	0.45	0.94	-0.12
UMass-MechBayes	758	32	0.64	0.71	0.5	0.92	0.08
ILM-EKF	857	32	0.7	0.67	0.81	0.98	0.12
RobertWalraven-ESG	858	32	0.72	0.71	0.58	0.83	-0.11
USC-SIkJalpha	858	32	0.73	0.67	0.36	0.5	0.17
MUNI-ARIMA	602	32	0.73	0.81	0.63	0.96	0.01
epiforecasts-EpiNow2	831	32	0.8	0.85	0.49	0.92	0.14
IEM_Health-CovidProject	842	32	0.84	0.92	0.47	0.84	0.07
EuroCOVIDhub-baseline	858	32	1	1	0.65	0.95	0.08
USyd-OneModelMan	762	32		0.76			



Bosse et al., in prep



Bosse et al., in prep



Meakin et al., in prep

Summary

Summary

- EpiNow2 is a toolkit for the surveillance of infectious disease outbreaks.
- Contains multiple models for estimating the reproduction number, case fatality ratios, and dealing with data truncation.
- These models can be used for short-term forecasts.
- They sometimes perform well (I promise) but recently have been doing poorly.
- Performance is better at short vs long time horizons and generally has a skewed distribution.
- A potential explanation is assuming everything is static beyond the forecasting horizon and lack of outlier handling
- EpiNow2 is under active development with UI improvements, increased model flexibility (more and better gaussian processes, static parameters, more time-varying parameter options) in the works.

Open questions

- Model structure
    a. Does adding mechanistic understanding improve forecasting models?
    b. What is the impact of misspecifying mechanisms?
    c. Have we captured the temporal evolution of Rt?
    d. What other structure should we include (space, variants, age etc.)?
- Role of data handling vs model structure
    a. Daily vs weekly data?
    b. Outlier handling?
    c. Are we forecasting NPIs?
- Computational efficiency
    a. Have we chosen the right level of complexity?
    b. Other / online methods?
- How do we create robust, easy-to-use and general tools?