

Supplementary appendix

Mortality from COVID-19 in 12 countries and 6 states of the United States

Country or region	Deaths (1)	Cases (000s)	Serious cases (3)	Ratio of (3)/(1)	Deaths / 1M pop.	First case reported	Population ≥20 years (millions)	IHME median estimate as of April 11	Ratio of IHME median to GACM median
USA	20,568	529	11,471	0.56	83	20-Jan	249	61,545	0.62
Italy	19,468	152	3,381	0.17	389	29-Jan	50	20,333	0.78
Spain	16,972	166	7,371	0.43	447	30-Jan	38	18,363	0.74
France	13,832	94	6,883	0.50	277	23-Jan	50	15,741	0.50
UK	9,875	79	1,559	0.16	190	30-Jan	52	37,494	0.74
Iran	4,474	72	3,930	0.88	77	18-Feb	58	na	na
Belgium	3,600	30	1,223	0.34	400	03-Feb	9	4,291	0.48
Hubei, China	3,219	68	134	0.04	64	10-Jan	50	na	na
Germany	2,940	126	4,895	1.66	43	26-Jan	68	7,080	0.58
Switzerland	1,081	25	386	0.36	154	24-Feb	7	4,893	2.40
Netherlands	2,737	26	1,384	0.51	211	26-Feb	13	18,067	3.64
Canada	653	23	557	0.85	22	24-Jan	30	na	na
Sub-Total	99,419	1,390	43,174	0.43	147	10-Jan	674	187,807	0.67
New York	8,650	182	5,009	0.58	577	01-Mar	15	13,463	0.61
New Jersey	2,183	58	1,914	0.88	312	04-Mar	7	3,915	0.50
Michigan	1,392	24	1,582	1.14	174	10-Mar	8	1,977	0.48
Louisiana	806	20	470	0.58	269	09-Mar	3	1,125	0.61
California	634	22	1,565	2.47	22	26-Jan	29	1,616	0.46
Washington state	496	10	191	0.39	83	20-Jan	6	842	0.68
Sub-Total	14,161	317	10,731	0.76	208	20-Jan	68	22,938	0.57
World total	110,108	1,757	50,525	0.46	14	10-Jan	7,795	na	na

Supplementary Table 1: Additional country-level information

	Unadjusted			Correction	Adjusted		
	Est	2.5	97.5		Est	2.5	97.5
Italy	25967	23236	30520	50%	38950	34854	45780
France	25896	20733	36989	10%	28486	22806	40688
UK	46277	24634	109040	70%	78671	41878	185368
Netherl'd	4756	3870	6498	20%	5707	4644	7798

Supplementary Table 2: Total deaths forecast and deaths adjusted using the specified correction factor to account for underreporting.

1 Data preparation

With the exception of data related to France, all data were obtained using the API for the site `coronavirus.app`. French data were taken from the site `dashboard.covid19.data.gouv.fr`, which provides data assembled by Santé publique France. Deaths occurring in both hospitals and long-term care facilities, and counts from the latter are available only from 31 March and are added to the database in bursts. No deaths from long term care facilities were recorded on 7 April, but over 800 deaths were registered on both 6 and 8 April. We made a decision to trust the cumulative death count from non-hospital facilities at the end of our study period (11 April), but not the daily reported values. The deaths outside hospitals, roughly one third of the total, were distributed across days in proportions observed for hospital deaths since the beginning of the data series (1 March).

It is not unambiguous when the virus was introduced in each country or region, and consequently when the probability of deaths first becomes positive and the model should be applied. Counts of zero deaths in each region prior to the first death being registered were excluded from the analysis, as was data from the first day deaths were recorded. Including deaths on the first day in the analysis would bias results towards predicting higher deaths in the early stages since the first observation would necessarily be non-zero.

2 Prior distributions

The observed deaths $Y_i(t)$ in region i on day t follows a Negative Binomial distribution with mean $\lambda_i(t)$ and overdispersion parameter τ , or more specifically

$$\begin{aligned}
 Y_i(t) &\sim \text{Neg Binom}[\lambda_i(t), \tau] \\
 \lambda_i(t) &= E_i C_i f(t; A_i, B_i, K_i) + D_i \\
 f(t; \xi, \omega, \alpha) &= \frac{2}{\omega\sqrt{2\pi}} e^{-\frac{(x-\xi)^2}{2\omega^2}} \int_{-\infty}^{\alpha\left(\frac{x-\xi}{\omega}\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \frac{2}{\omega\sqrt{2\pi}} e^{-\frac{(x-\xi)^2}{2\omega^2}} \int_{-\infty}^{\alpha\left(\frac{x-\xi}{\omega}\right)} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt
 \end{aligned}$$

The E_i is an age and sex standardized death count, obtained from population data for each country and age-specific mortality risk calculated from Italian data on 29 March 2020 (EpiCentro 2020). The model parameters which determine the shape of the epidemic curve are: C_i , the expected total deaths relative to the Italian rates on 29 March; A_i , the location of the peak of the epidemic; B_i , the duration of the epidemic (roughly the standard deviation of the skew-Normal); K_i , the skewness parameter specifying the steepness of the growth phase of the epidemic relative to the decline phase; and D_i , a ‘spark’ term allowing for occasional deaths outside the range of the main epidemic.

parameter	prior	95pct interval
Location A_i	$N(29 \text{ Mar}, 100^2)$	Apr 2019, Jan 2022
Horizontal scale B_i	$\text{Gamma}(60, 3)$	15, 25
Vertical scale C_i	$\text{Gamma}(1.5, 1/3)$	0.32, 14
Skewness K_i	$\text{Gamma}(3, 0.5)$	1.2, 14
Spark term D_i	$\text{Gamma}(0.5, 1)$	0.00049, 2.5
Overdispersion τ	$\text{Exp}(1)$	0.025, 3.7

Supplementary Table 3: Prior distributions for model parameters.

Table 3 shows the prior distributions for each of the model parameters, the densities for several parameters are displayed in Figure 1. Figures 1 d and 1 e show the course of an epidemic for a population of 1 million people with the age distribution of Italy, and the distribution of the final number of total cases appears in 1 f. The prior median and 95% prior interval are shown above the horizontal axis of 1 f.

The duration parameter B_i has a 95% prior interval spanning 15 to 25 days, most of a country’s deaths would be expected to occur within a span of four times this parameter. The intensity parameter C_i has a prior which expects the mortality rate in a country with Italy’s age structure to be between 100 and 2,000 per million. The number of deaths per day could be between 10 and 200 per million at its peak. For each posterior sample’ daily counts, we computed $\lambda_i(t)$, and simulated potential future death counts from the negative binomial distribution.

3 Validation

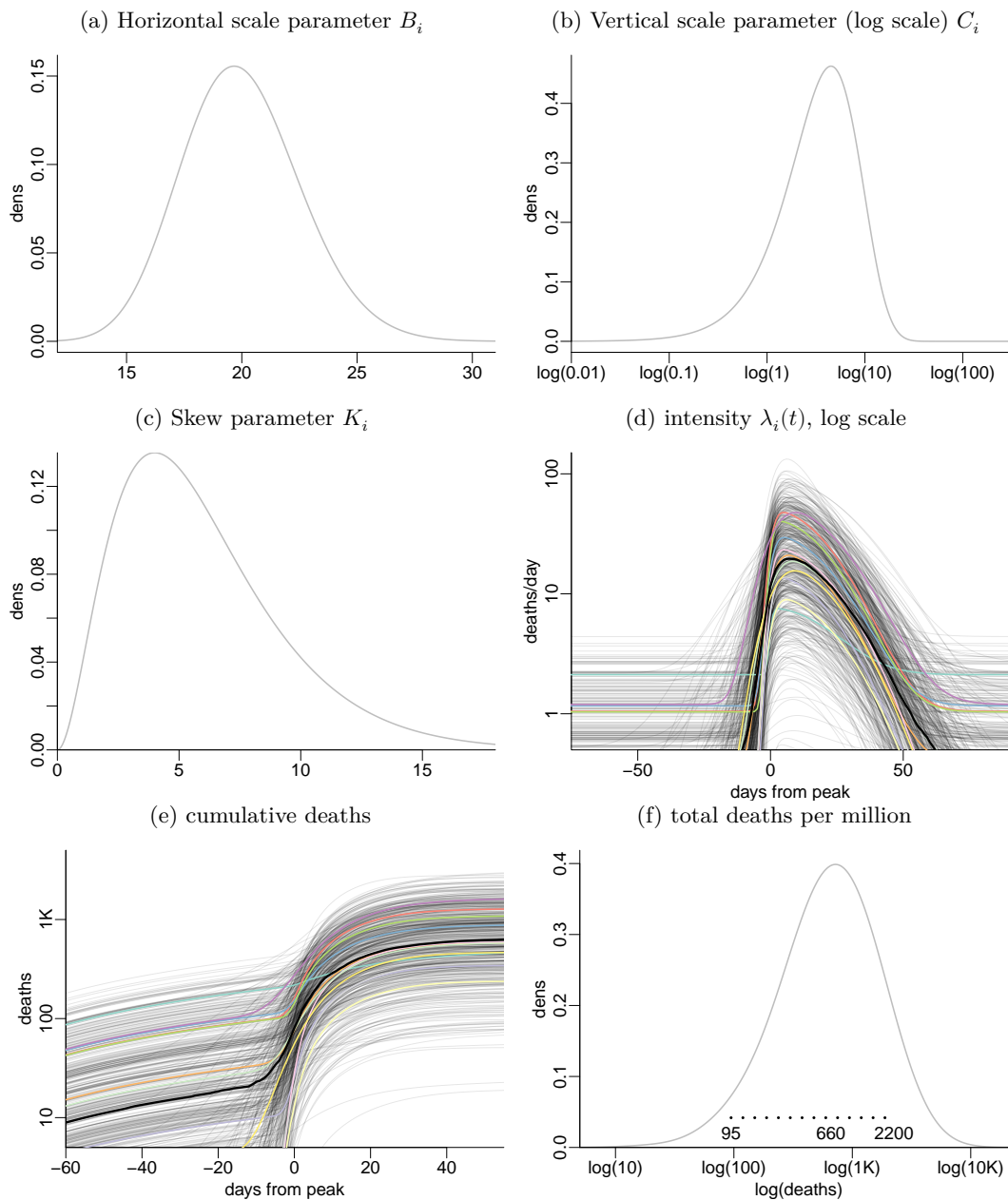
As a validation exercise, the model was trained on deaths up to 20 March and used to predict deaths to 14 April. Figure 2 shows predictions made from training data an the subsequent true data. Notable is the predictions made for New York which were made when daily deaths were 20. The median prediction for New York greatly underestimates the eventual trajectory, although the 95% global envelope contains the observed data.

Figure 3 shows the predictions made for deaths up to 14 April divided by the cumulative deaths observed to that date, for reporting areas with more than 20 cases on 20 March. The eventual count is in the 50% prediction interval for roughly half of the predictions, only the data from New York and Belgium lie outside th 95% interval.

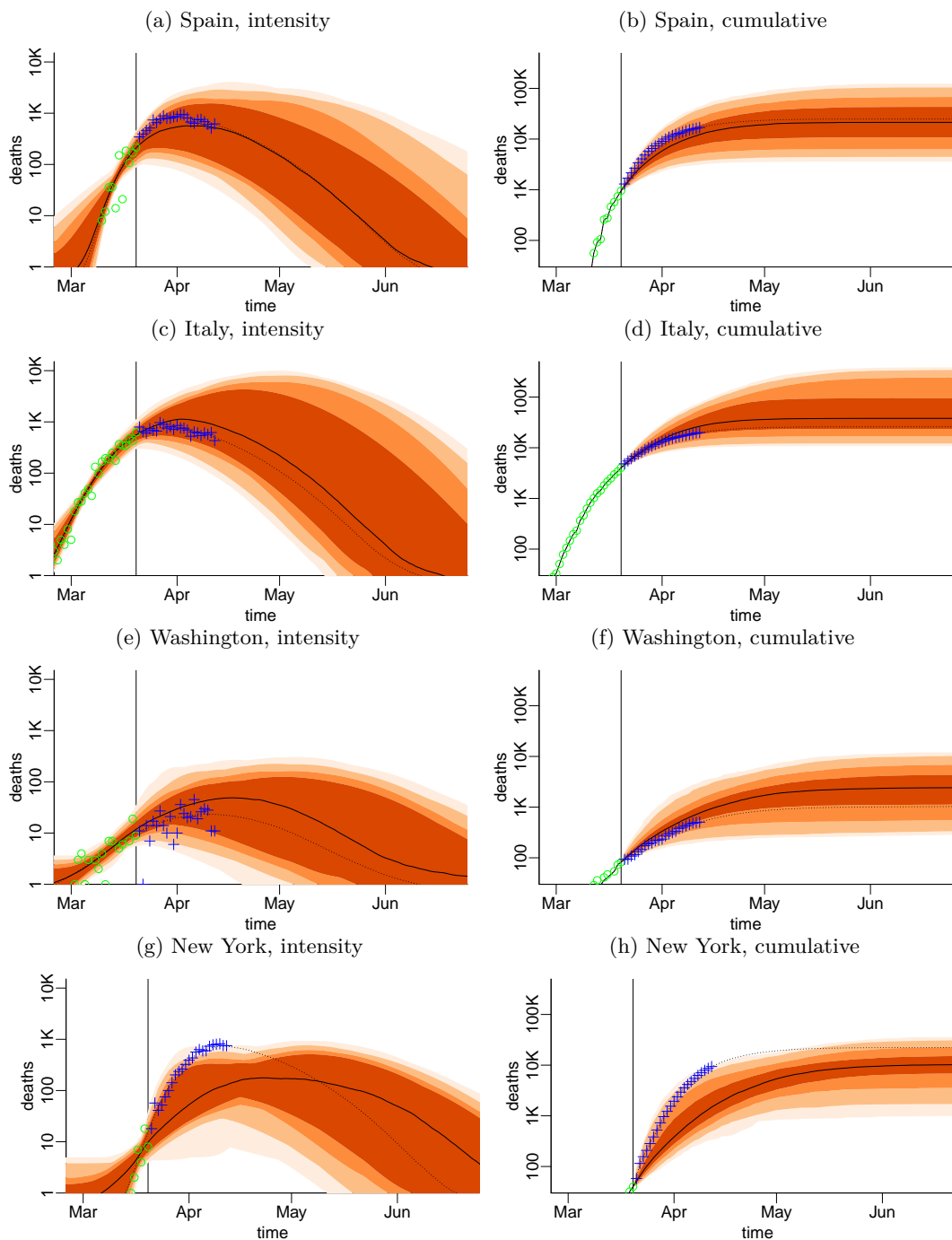
References

EpiCentro (2020). *Integrated surveillance of COVID-19 in Italy*. URL: https://www.epicentro.iss.it/coronavirus/bollettino/Infografica_29marzo%20ENG.pdf.

Supplementary Figure 1: Prior distributions and epidemic trajectories simulated from the priors.



Supplementary Figure 2: Validation, training data on or before 19 March (○); validation data (+); posterior median from training data (—); and posterior median from full data (⋯).



Supplementary Figure 3: Validation errors for predicting deaths to 14 April using data up to 20 March.

