Effects of non-pharmaceutical interventions on COVID-19: A Tale of Three Models

Vincent Chin[1,2], John P.A. Ioannidis[4,5,6,7,8], Martin A. Tanner[3], and Sally Cripps[1,2,*]

[1]ARC Centre for Data Analytics for Resources and Environments, Australia
[2]School of Mathematics and Statistics, The University of Sydney, Australia
[3]Department of Statistics, Northwestern University, USA
[4]Stanford Prevention Research Center, Department of Medicine, Stanford University, USA
[5]Department of Epidemiology and Population Health, Stanford University, USA
[6]Department of Biomedical Data Sciences, Stanford University, USA
[7]Department of Statistics, Stanford University, USA
[8]Meta-Research Innovation Center at Stanford (METRICS), Stanford University, USA
[*]Corresponding author: sally.cripps@sydney.edu.au +61 425-276-967

September 14, 2020

First Submission: July 22, 2020

Abstract

In this paper, we compare the inference regarding the effectiveness of the various non-pharmaceutical interventions (NPIs) for COVID-19 obtained from three SIR models, all developed by the Imperial College COVID-19 Response Team. One model was applied to European countries and published in *Nature*[1] (model 1), concluding that complete lockdown was by far the most effective measure, responsible for 80% of the reduction in R_t, and 3 million deaths were avoided in the examined countries. The Imperial College team applied a different model to the USA states[2] (model 2), and in response to our original submission, the Imperial team has proposed in a referee report a third model which is a hybrid of the first two models (model 3). We demonstrate that inference is highly nonrobust to model specification. In particular, inference regarding the relative effectiveness of NPIs changes substantially with the model and decision makers who are unaware of, or ignore, model uncertainty are underestimating the risk attached to any decisions based on that model. Our primary observation is that by applying to European countries the model that the Imperial College team used for the USA states (model 2), complete lockdown has no or little effect, since it was introduced typically at a point when R_t was already very low. Moreover, using several state-of-the-art metrics for Bayesian model comparison, we demonstrate that model 2 (when applied to the European data) is better supported by the data than the model published in Nature[1]. In particular, serious doubt is cast on the conclusions in Flaxman et al.[1], whether we examine the data up to May 5th (as in Flaxman et al.[1]) or beyond the point when NPIs began to be lifted. Only by objectively considering a wide variety of models in a statistically principled manner, can one begin to address the effectiveness of NPIs such as lockdown. The approach outlined in this paper provides one such path.

Keywords: COVID-19; Non-pharmaceutical Interventions; Bayesian SIR Models

1 A Tale of Three Models

The different models (Flaxman et al.[1] and Unwin et al.[2]) produced by the Imperial College COVID-19 Response Team aim to explain the evolution of R_t. We will refer to these models as model 1 (the model applied to European countries in the *Nature* publication[1]), model 2 (the model applied to the USA states[2]), and model 3 (the model that Flaxman et al.[1] proposed in their referee report to our original submission to Matters Arising), respectively. The first two models for the R_t are fundamentally different. In model 1, the proportional variation of R_t from the initial R_0 is modelled as a step function and only allowed to change in response to an intervention. Therefore, any decrease in R_t (even if this decrease is a result of the increasing proportion of the population who are infected, to changes in human behaviour, clustered contact structures and/or pre-existing immunity[3]) must, by the model construction, be attributed to interventions and the impact is immediate with no time lag or gradual change, when a new intervention is adopted. In model 2, the proportional variation of R_t from R_0 has a different functional form and is allowed to vary with mobility indicators for various activities. These mobility indicators are proxies for changes in human behaviour, whether that change is due to one or more centrally imposed interventions or whether it is the product of individuals responding to the epidemic on their own initiative independently of centrally imposed interventions. Model 2 also does not presume step functions and is therefore capable of capturing more gradual changes over time. Model 1 is deliberately simple and designed to test the impact of interventions without the confounders of mobility data,

which themselves will reflect the impact of the interventions. Model 2 leaves out interventions and uses mobility data as predictors in the evolution of R_t. The advantage of model 2 is that it gives a more flexible estimate of R_t, by allowing it to change with mobility trends. Although there is no explicit causal structure in model 2, the time sequence of the data and events makes inference around the impact of interventions possible by observing if the change in R_t precedes a given intervention or interventions or not. Here, we apply both models 1 and 2 to the European data to compare the results and inferences they obtain. See supplementary methods for details on methods and data. In response to our original submission, the Imperial team proposed yet a third model which includes a single NPI, lockdown, as well as mobility, and we refer to this as model 3. We extend Imperial's model 3 to include all the NPIs that were used their original model 1, and discuss inferences regarding this model.

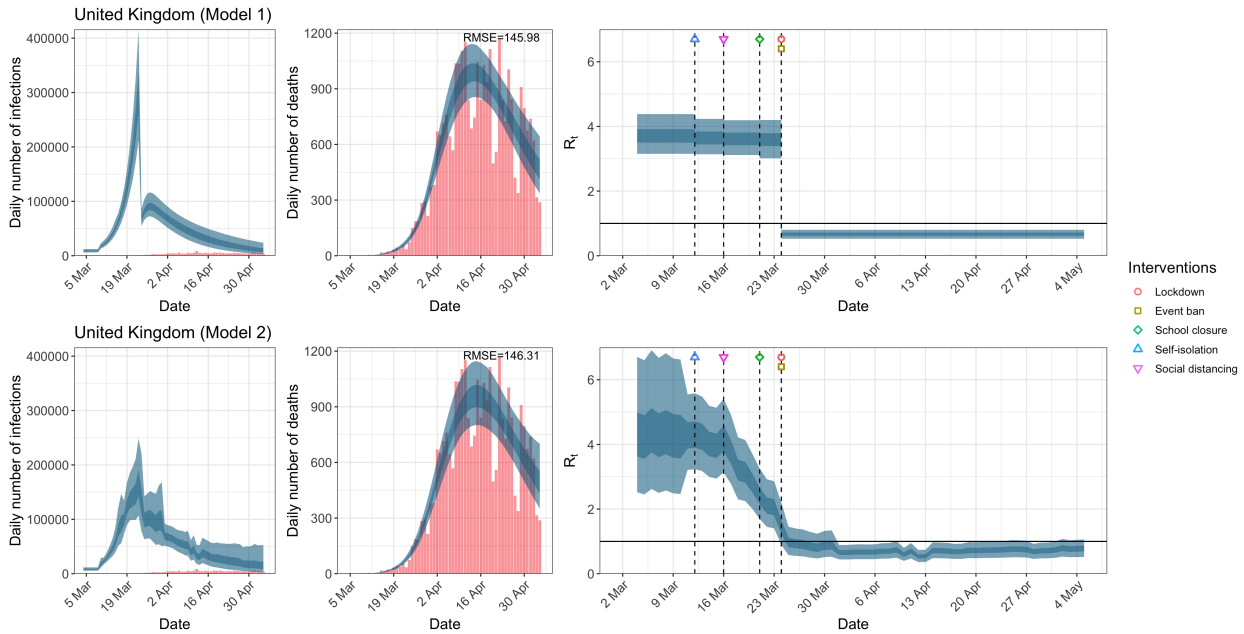
2 Results

There are two main observations when comparing the modeling results in Figure 1a, as well as the corresponding Extended Data Figures 1a–1e. First, and most notable, is that while the models give very different trajectories of R_t, both models produce visually similar fits to the observed daily death counts, although we note that the root mean square error (RMSE) for model 2 is less than or equal to the RMSE for model 1 in 9 out of the 11 countries. Additionally, and most importantly, the predictive density of model 2 provides a better fit to the data than model 1 as measured by two state-of-the-art criteria for Bayesian model comparison, the Deviance information criteria (DIC)[4] and the Watanabe-Akaike information criteria (WAIC)[5]. Table 2 and Table 3 provide details. The second observation is that inference regarding the impact of interventions varies significantly between the two models. The inference from model 1 indicates that lockdown had the biggest impact of all the interventions in all countries. Indeed, Flaxman et al.[1] states that *Lockdown has an identifiable large impact on transmission (81% [75%- 87%] reduction)*.

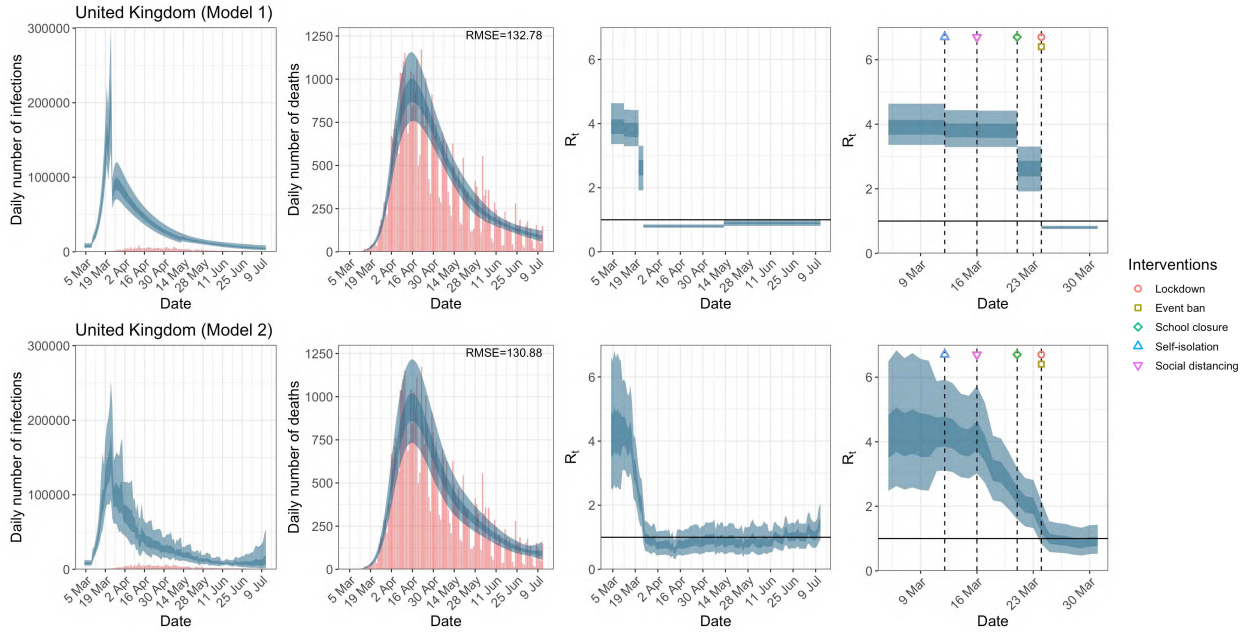
In contrast, model 2 shows clearly that R_t was falling well before lockdown, which occurred after the sharp decline in R_t. With the exception of Belgium (and to a lesser degree France), this is consistently seen across the European countries (see Extended Data Figures 1a–1e) and is in line with the mobility data, see Figure 2. At the time lockdown was adopted in the UK, R_t had already decreased to 1.46 (95% CI, 0.89 to 2.20) from an initial R_0 of 4.46 (95% CI, 2.64 to 7.20) according to model 2, and the largest drop in R_t occurred after the implementation of self-isolation measures and encouraging social distancing, see Table 1.

We also evaluated the effect of extending the time horizon of analysis until July 12th where restrictions had been lifted in some of these countries. At the suggestion of Flaxman et al.[1] in their referee report, we allow for the effect of the NPIs to be asymmetric. That is, we allow the impact of lifting of an NPI on R_t to be different in magnitude from the impact of imposing that NPI. To make the models 1 and 2 comparable we start both models on March 4th. We also show the results for model 1 with a start date of February 13th, the date used by[1], as well as a March 4th start date, in Extended Table 3. As shown by Extended Table 3, inference regarding the impact of an NPI is not highly sensitive to the start date.

We also include an additional three countries, i.e. Greece, the Netherlands and Portugal, in the analysis. Extended Data Table 1 shows the end dates of school closures, ban on public events and lockdown used in the analysis. Figure 1b presents a comparison of the results obtained from both models for the UK for the extended data, while Extended Data Figures 2a–2g show similar results



(a) Daily infections, daily deaths and R_t until May 5th.



(b) Daily infections, daily deaths and R_t until July 12th in columns 1–3. Column 4 is a magnification of column 3 around the period of the NPIs.

Figure 1: United Kingdom. The start time for the plots is 10 days before 10 deaths are recorded. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.

for the remaining 13 countries. Extended Data Table 2 reports the RMSE while Tables 2 and 3 report measures of fit, effective sample sizes and two information criteria; Deviance information

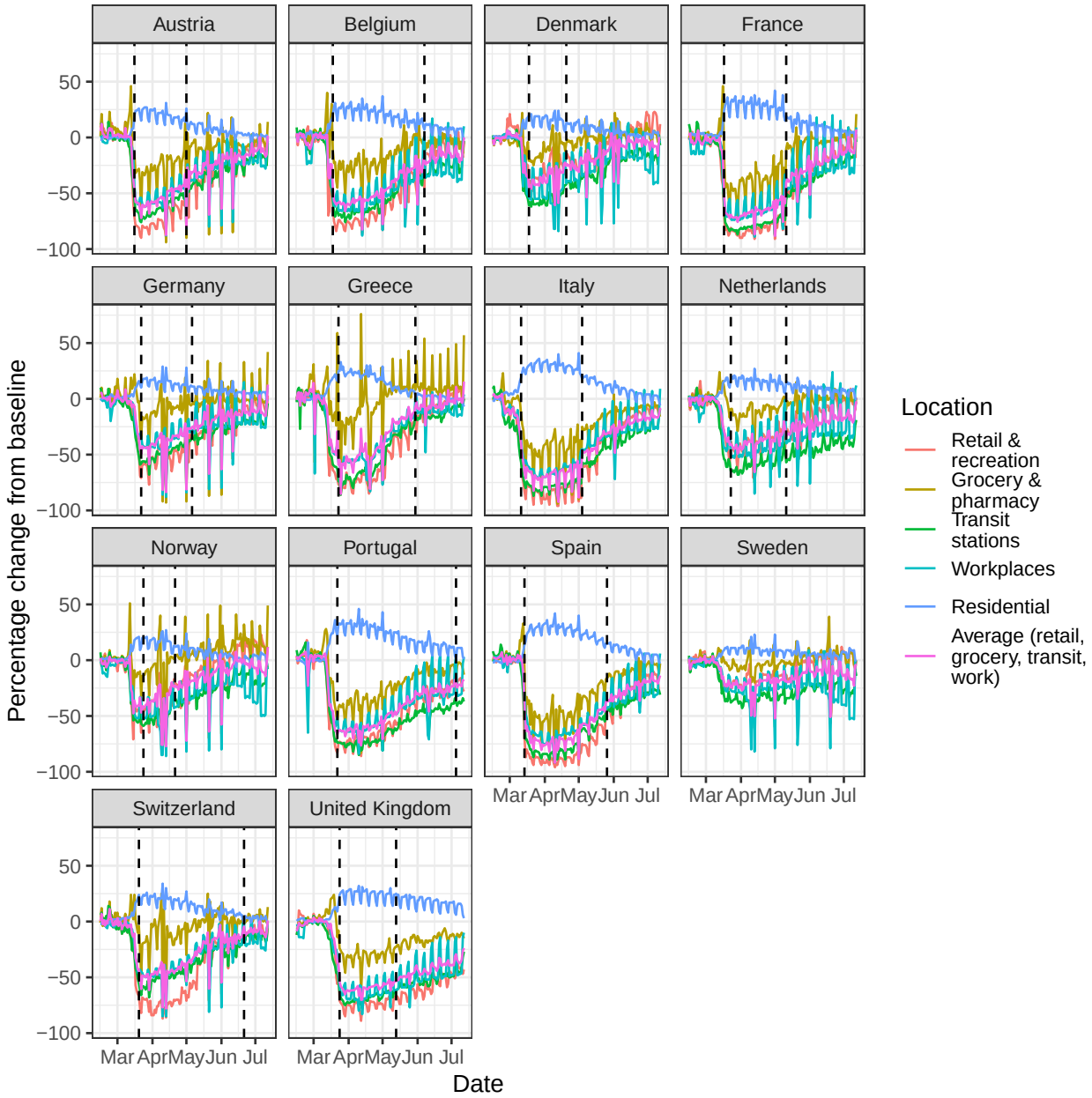


Figure 2: Percentage change in mobility from baseline level from February 15th to July 12th, by locations in each of the European countries examined in Flaxman et al.[1], as well as an additional three countries consisting of Greece, the Netherlands and Portugal. Average mobility is computed based on the trends in retailers and recreation venues, grocery markets and pharmacy, transit stations and workplaces. Black dashed lines in each plot indicate the lockdown start and end dates.

criterion (DIC)[4] and Watanabe-Akaike information criterion (WAIC).

Figure 1b, Extended Data Figures 2a–2g, and Tables 2 and 3 present a number of features. Model 2 provides better fits to the data than model 1, as evidenced by the RMSE, which is lower for 11 out of the 14 countries. Importantly, the impact of lockdown in model 1 is significantly reduced; the CI for the coefficient of lockdown is (0.73,1.33), for the period up to July 12th, while

Country	R_0	R_t immediately at NPIs introduction				
		SI	SD	SC	EB	LD
UK	3.72 (3.15,4.38)	3.65 (3.14,4.23)	3.62 (3.12,4.19)	3.59 (3.01,4.20)	0.67 (0.53,0.80)	0.67 (0.53,0.80)
Austria*	3.38 (2.10,4.95)	0.54 (0.42,0.66)	0.54 (0.42,0.66)	3.28 (2.01,4.79)	–	0.54 (0.42,0.66)
Belgium	5.63 (3.86,7.43)	5.54 (3.78,7.47)	5.36 (3.65,7.32)	5.36 (3.65,7.32)	5.43 (3.72,7.33)	0.92 (0.80,1.05)
Denmark	3.76 (2.43,5.56)	3.64 (2.30,5.56)	3.57 (2.30,5.32)	3.57 (2.30,5.32)	3.64 (2.30,5.56)	0.69 (0.59,0.79)
France	4.70 (4.07,5.42)	4.47 (3.56,5.26)	4.47 (3.56,5.26)	4.55 (3.73,5.25)	4.58 (3.81,5.30)	0.69 (0.60,0.78)
Germany	4.08 (3.30,5.06)	4.01 (3.20,5.08)	3.97 (3.15,4.95)	3.94 (3.15,4.92)	0.71 (0.58,0.83)	0.71 (0.58,0.83)
Italy	3.26 (2.91,3.66)	3.10 (2.51,3.60)	3.10 (2.51,3.60)	3.22 (2.82,3.61)	3.10 (2.51,3.60)	0.69 (0.62,0.76)
Norway*	2.61 (1.44,4.00)	2.51 (1.38,3.90)	2.48 (1.38,3.89)	–	–	0.40 (0.26,0.58)
Spain	5.27 (4.34,6.37)	0.66 (0.57,0.73)	5.18 (4.20,6.26)	5.15 (4.16,6.22)	0.66 (0.57,0.73)	0.66 (0.57,0.73)
Sweden	2.12 (1.57,2.75)	2.08 (1.54,2.69)	2.06 (1.55,2.62)	–	0.85 (0.64,1.08)	–
Switzerland*	3.00 (2.17,3.93)	–	2.84 (2.09,3.69)	2.87 (2.13,3.72)	2.89 (2.13,3.81)	0.54 (0.43,0.66)

(a) Model 1.

Country	R_0	R_t immediately at NPIs introduction				
		SI	SD	SC	EB	LD
UK	4.46 (2.64, 7.20)	4.33 (3.23, 5.58)	4.20 (3.13, 5.41)	2.41 (1.69, 3.25)	1.46 (0.89, 2.20)	1.46 (0.89, 2.20)
Austria*	2.69 (1.06, 5.71)	0.87 (0.34, 1.78)	0.87 (0.34, 1.78)	1.43 (0.56, 2.91)	–	0.87 (0.34, 1.78)
Belgium	4.36 (2.38, 7.17)	4.19 (2.27, 6.72)	4.29 (2.33, 6.89)	4.29 (2.33, 6.89)	4.49 (2.47, 7.18)	4.46 (3.09, 6.09)
Denmark	2.39 (1.08, 5.13)	1.29 (0.56, 2.51)	0.89 (0.38, 1.75)	0.89 (0.38, 1.75)	1.29 (0.56, 2.51)	0.69 (0.28, 1.40)
France	4.05 (2.46, 6.49)	3.62 (2.68, 4.62)	3.62 (2.68, 4.62)	4.02 (3.01, 5.14)	4.53 (3.38, 5.88)	2.15 (1.35, 3.10)
Germany	4.49 (2.42, 7.22)	4.18 (2.25, 6.78)	4.15 (2.24, 6.74)	4.04 (2.84, 5.55)	1.24 (0.74, 1.84)	1.24 (0.74, 1.84)
Italy	4.62 (2.71, 7.29)	1.74 (1.05, 2.50)	1.74 (1.05, 2.50)	2.87 (2.02, 3.75)	1.74 (1.05, 2.50)	1.22 (0.74, 1.80)
Norway*	2.22 (1.04, 4.95)	0.65 (0.24, 1.37)	0.82 (0.32, 1.67)	–	–	0.60 (0.31, 1.02)
Spain	5.25 (3.28, 7.83)	1.64 (1.00, 2.39)	5.31 (4.12, 6.71)	4.30 (3.25, 5.52)	1.64 (1.00, 2.39)	1.64 (1.00, 2.39)
Sweden	3.79 (1.85, 6.81)	3.24 (1.64, 5.62)	2.77 (1.84, 3.97)	–	1.59 (1.03, 2.23)	–
Switzerland*	3.45 (1.54, 6.40)	–	2.46 (1.58, 3.56)	2.43 (1.56, 3.52)	2.80 (1.77, 4.13)	0.99 (0.60, 1.45)

(b) Model 2.

Table 1: Basic reproduction number R_0 and time-varying reproduction number R_t immediately at time of introduction of NPIs given by both models using data up to May 5th for all eleven countries analysed in Flaxman et al.[1]. These NPIs are self-isolation (SI), social distancing (SD), school closure (SC), event ban (EB) and lockdown (LD). 95% credible intervals are given in parentheses below the corresponding point estimates. Countries where the seeding of new infections occur after the introduction of NPIs are denoted with an asterisk.

Model	Time period	lppd	p_{WAIC1}	p_{WAIC2}	p_{DIC}	$-2\log p(\mathbf{y}\|\hat{\theta}_{Bayes})$
1	Up to May 5th	-2799.25	28.24	30.24	26.55	5600.18
2	Up to May 5th	-2777.50	33.28	36.10	33.68	5554.60
1	Up to July 12th	-6634.85	69.17	97.56	42.93	13295.94
2	Up to July 12th	-6594.78	89.34	122.59	68.31	13210.60

Table 2: Estimates of log pointwise predictive density (lppd), effective number of parameters (WAIC refers to Watanabe-Akaike information criterion; DIC refers to Deviance information criterion) and log predictive density $\log p(\mathbf{y}|\hat{\theta}_{Bayes})$ evaluated at posterior mean $\hat{\theta}_{Bayes}$.

Model	Time period	WAIC1	WAIC2	DIC
1	Up to May 5th	5654.98	5658.98	5653.28
2	Up to May 5th	**5621.56**	**5627.20**	**5621.96**
1	Up to July 12th	13408.04	13464.82	13381.80
2	Up to July 12th	**13368.24**	**13434.74**	**13347.22**

Table 3: Estimates of various information criteria; the Watanabe-Akaike information criterion, $WAIC1 = -2\text{lppd} + 2p_{WAIC1}$ and $WAIC2 = -2\text{lppd} + 2p_{WAIC2}$ which uses lppd as a measure of fit with p_{WAIC1} and p_{WAIC2} as the effective number of parameters to penalize the fit respectively; the Deviance information criterion DIC $= -2\log p(\mathbf{y}|\hat{\theta}_{Bayes}) + 2p_{DIC}$ which uses $\log p(\mathbf{y}|\hat{\theta}_{Bayes})$, as the measure of fit, and p_{DIC} as the penalty. Note that a lower value implies a better predictive density, and the preferred model for each criteria and time period is shown in bold.

the CI for the coefficient based on data up to May 5th is (1.37, 2.05). The CI for the coefficient of lifting lockdown is (0.00,0.15), suggesting that it is significant.

The authors[1], in their referee report for our submission, proposed a third model – model 3, in which they included mobility and lockdown, as the only NPI, and found that lockdown was still significant in reducing R_t. We question why they chose only to include lockdown, and not any other of the NPIs, either jointly or singly, given that their initial goal was to quantify the relative contributions of several NPIs to the reduction in R_t. In response, we construct a similar model but, for thoroughness and in the spirit of the original Nature paper, we include include all of the NPIs, as well as mobility. Figure 3 plots the posterior distribution for the regression coefficients, and shows that lockdown and event ban had similar effect sizes on the reduction of R_t. This paints a very different picture from the Nature paper which claimed that lockdown was responsible for 80% of the drop in R_t. Moreover, we question the usefulness of including NPIs as well as mobility data in the same model, because of confounding effects, but we include this analysis only in response to the referee report from Imperial. NPIs reduce R_t by reducing contact among individuals. An indirect measure of the reduction in individual contact is the mobility data, and so these data will be highly correlated with NPIs, making inference difficult.

Finally, we should point out that Figure 1b shows that the gradual increase in R_t following the minimum value commenced well before the lifting of lockdown. This suggests that as people become less frightened by the prospect of a catastrophe, mobility (see Figure 2), and hence the time-varying reproduction number, increases.

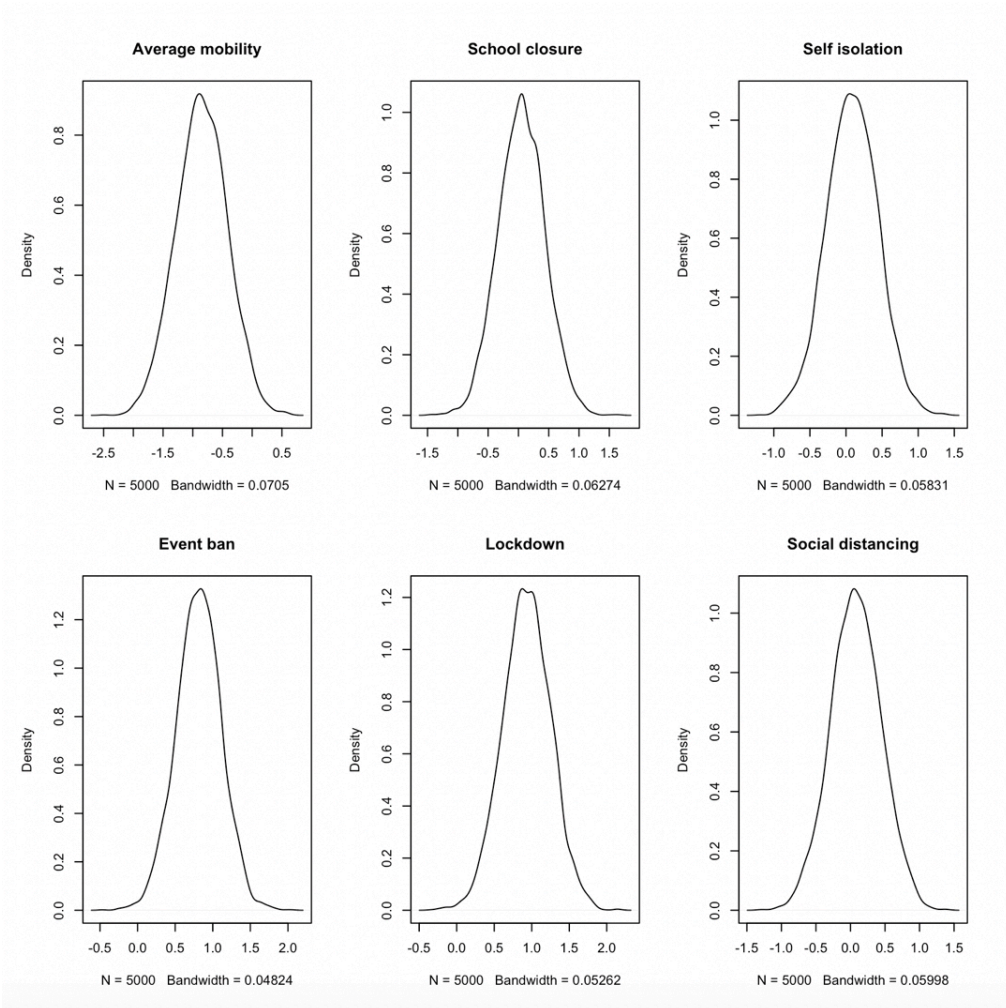


Figure 3: Density estimates of the posterior distributions of regression coefficients for the covariates in model 3, for the data up to May 5th.

3 Conclusion

The purpose of this paper is to demonstrate that there is substantial model uncertainty which is ignored by the Nature paper[1], and that inference around the effectiveness of NPIs is highly model dependent. Failing to report this uncertainty to those who make policy decisions will ultimately undermine the public's trust in the value of decisions based on statistical modeling. In attempting to answer the question 'which results should be used to guide policy making in lifting restrictions?', we must be transparent about what we do not know. Flaxman et al.[1] make the statement *We find that, across 11 countries, since the beginning of the epidemic, 3,100,000 [2,800,000 - 3,500,000] deaths have been averted due to interventions*. However, we have shown that their model 2, which is better supported by the data than model 1, provides very different inferences concerning the effectiveness of intervention strategies for the period March 4th to May 5th (as well as up to July 12th) in the UK. Moreover when we include all the NPIs in their model 3 we find that event ban had a similar effect on the reduction of R_t as did lockdown- a very different result from the Nature paper[1]. Although it is tempting to congratulate ourselves on our decision to implement lockdown,

citing the number of lives that were saved, we should resist this temptation, and examine other possible explanations in an objective and statistically principled manner. The approach outlined in this paper provides one such path. Failure to do this and therefore mis-attribute causation could mean we fail to find the optimal solution to this very challenging and complex problem, given that complete lockdown can also have many adverse consequences[6].

We do not want to reach the opposite extreme of claiming with certainty that lockdown definitely had no impact. Other investigators using a different analytical approach have suggested also benefits from lockdown, but of much smaller magnitude (13% relative risk reduction[7]) that might not necessarily match complete lockdown-induced harms in a careful decision analysis. Another modeling approach has found that benefits can be reaped by simple self-imposed interventions such as washing hands, wearing masks, and some social distancing[8]. Observational data need to be dissected very carefully and substantial uncertainty may remain even with the best modelling[9]. Regardless, causal interpretations from models that are not robust should be avoided. Given the analyses that we have performed using the three models that the Imperial College team has developed, one cannot exclude that the attribution of benefit to complete lockdown is a modelling artefact.

Acknowledgement

We congratulate the Imperial College Response Team for sharing openly the code for their models and for the overall transparency of their work that has allowed performing these analyses. We thank Hadi Ashfar for his suggestions to improve the computational efficiency of the HMC scheme. We also thank Jack Wood for his help in the construction of Extended Data Table 1.

Author Contributions

All authors contributed equally to this work. VC performed all the computations and produced all the graphics. SC wrote the initial draft. JI and MT wrote subsequent drafts. All authors discussed the results and implications and commented on the manuscript at all stages.

Conflicts of Interest

None.

Code Availability

All source code for the replication of our results is available from the Imperial College COVID-19 Response Team's Github repository: `https://github.com/ImperialCollegeLondon/covid19model`. Daily confirmed cases and deaths data are publicly available from the European Centre of Disease Control's website.

References

[1] Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* (2020). Doi: https://doi.org/10.1038/s41586-020-2405-7.

[2] Unwin, H. J. T. *et al.* State-level tracking of COVID-19 in the United States. *medRxiv* (2020). Doi: https://doi.org/10.1101/2020.07.13.20152355.

[3] Grifoni, A. *et al.* Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501 (2020). Doi: https://doi.org/10.1016/j.cell.2020.05.015.

[4] Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 583–639 (2002).

[5] Watanabe, S. A widely applicable bayesian information criterion. *Journal of Machine Learning Research* **14**, 867–897 (2013).

[6] Melnick, E. R. & Ioannidis, J. P. A. Should governments continue lockdown to slow the spread of COVID-19? *BMJ* **369** (2020). Doi: https://doi.org/10.1136/bmj.m1924.

[7] Islam, N. *et al.* Physical distancing interventions and incidence of coronavirus disease 2019: Natural experiment in 149 countries. *BMJ* **370** (2020). Doi: https://doi.org/10.1136/bmj.m2743.

[8] Teslya, A. *et al.* Impact of self-imposed prevention measures and short-term government-imposed social distancing on mitigating and delaying a COVID-19 epidemic: A modelling study. *PLoS Medicine* (2020). Doi: https://doi.org/10.1371/journal.pmed.1003166.

[9] Ioannidis, J. P. A., Cripps, S. & Tanner, M. A. Forecasting for COVID-19 has failed. *International Journal of Forecasting* (2020). To Appear.

[10] Google LLC. Google COVID-19 Community Mobility Reports. Retrieved from: https://www.google.com/covid19/mobility/ (2020). Last accessed: July 15, 2020.

[11] Liu, Y., Gayle, A. A., Wilder-Smith, A. & Rocklöv, J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine* **27** (2020). Doi: https://doi.org/10.1093/jtm/taaa021.

[12] Hale, T., Webster, S., Petherick, A., Phillips, T. & Kira, B. Oxford COVID-19 Government Response Tracker. Retrieved from: https://github.com/OxCGRT/covid-policy-tracker (2020). Last accessed: July 15, 2020.

[13] Our World in Data. Policy Responses to the Coronavirus Pandemic. Retrieved from: https://ourworldindata.org/policy-responses-covid (2020). Last accessed: July 15, 2020.

[14] SBS News. Denmark reports no spike in coronavirus cases since lifting lockdown. Retrieved from: https://www.sbs.com.au/news/denmark-reports-no-spike-in-coronavirus-cases-since-lifting-lockdown (2020). Last accessed: July 15, 2020.

[15] The Local. AFTER LOCKDOWN: Are Denmark's and Norway's restrictions now like Sweden's? Retrieved from: https://www.thelocal.com/20200421/explained-are-denmark-and-norways-restrictions-still-tougher-than-swedens (2020). Last accessed: July 15, 2020.

Supplementary Methods

Model 2 assumes that R_t is a function of mobility and allows this impact to vary regionally by the use of regional specific random effects terms. The data used by Unwin et al.[2] to estimate R_t is the Google's COVID-19 Community Mobility Report. We adapt this methodology to the European context by modelling potential idiosyncrasies in mobility trends across countries using country-specific random effects. To complete the specification of model 2 for the European data, we use Google's COVID-19 Community Mobility Report[10] which provides data measuring the percentage change in mobility compared to a baseline level for visits to a number of location categories; retailers and recreation venues, grocery markets and pharmacies, parks, transit stations, workplaces and residential places. We use the average change in mobility across all location categories, excluding residential places and parks, as a measure of the reduction in mobility, so that the reproduction number $R_{t,m}$ is given by:

$$R_{t,m} = R_{0,m} \cdot f(-X_{t,m}\alpha - \beta_{1,m} - X_{t,m}\beta_{2,m} - \epsilon_{m,w(m)(t)}), \tag{1}$$

where $X_{t,m}$ is the average mobility at time t for country m and $\epsilon_{m,w(m)(t)}$ is a weekly AR(2) process centred around 0.

The seeding of new infections in models 1 and 2 is chosen to be 10 days before the day a country has cumulatively observed 10 deaths so that mobility data are available for all the countries examined. In the UK, this corresponds to March 4th, which is later than the infection start date of February 13th used by Flaxman et al.[1].

For posterior inference of model 2, we use the same prior distributions as in Unwin et al.[2] except for R_0, where a weakly informative prior of a normal distribution truncated below at 1 with mean 3.28 and standard deviation 2 is used. This prior is chosen so that approximately 95% of the prior density is between 1 and 7[11], and that R_0 is above the critical value of 1 at the start of the epidemic. For model 1, we use the same priors as in Flaxman et al.[1] for the analysis up to May 5th and July 12th.

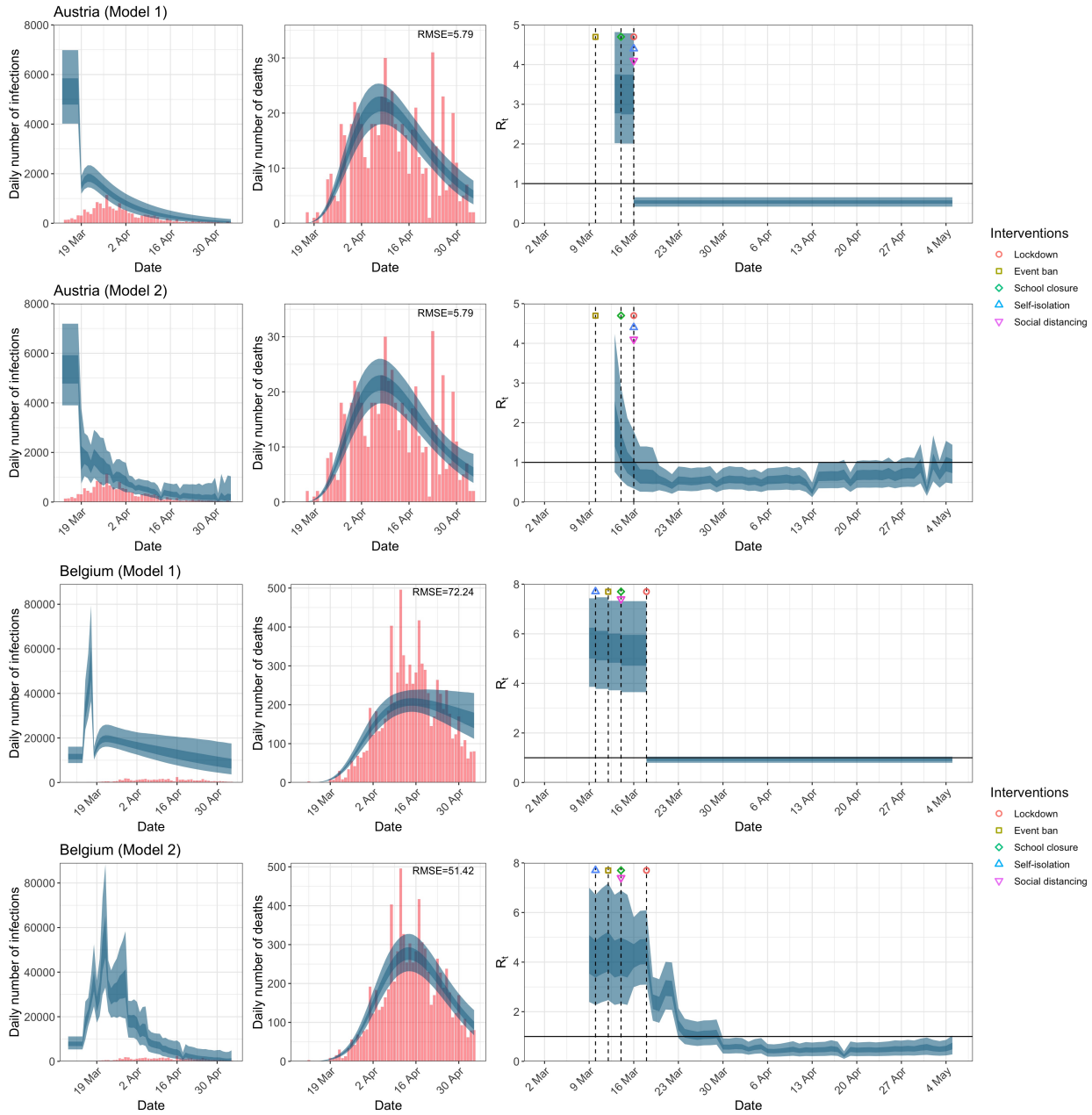
The functional form of $R_{t,m}$ based on model 3 suggested by Flaxman et al.[1] analysed in our paper is:

$$R_{t,m} = R_{0,m} \cdot f(-X_{t,m}\alpha_0 - \sum_{k=1}^{5} I_{k,t,m}\alpha_k - \beta_{1,m} - X_{t,m}\beta_{2,m} - \epsilon_{m,w(m)(t)}), \tag{2}$$

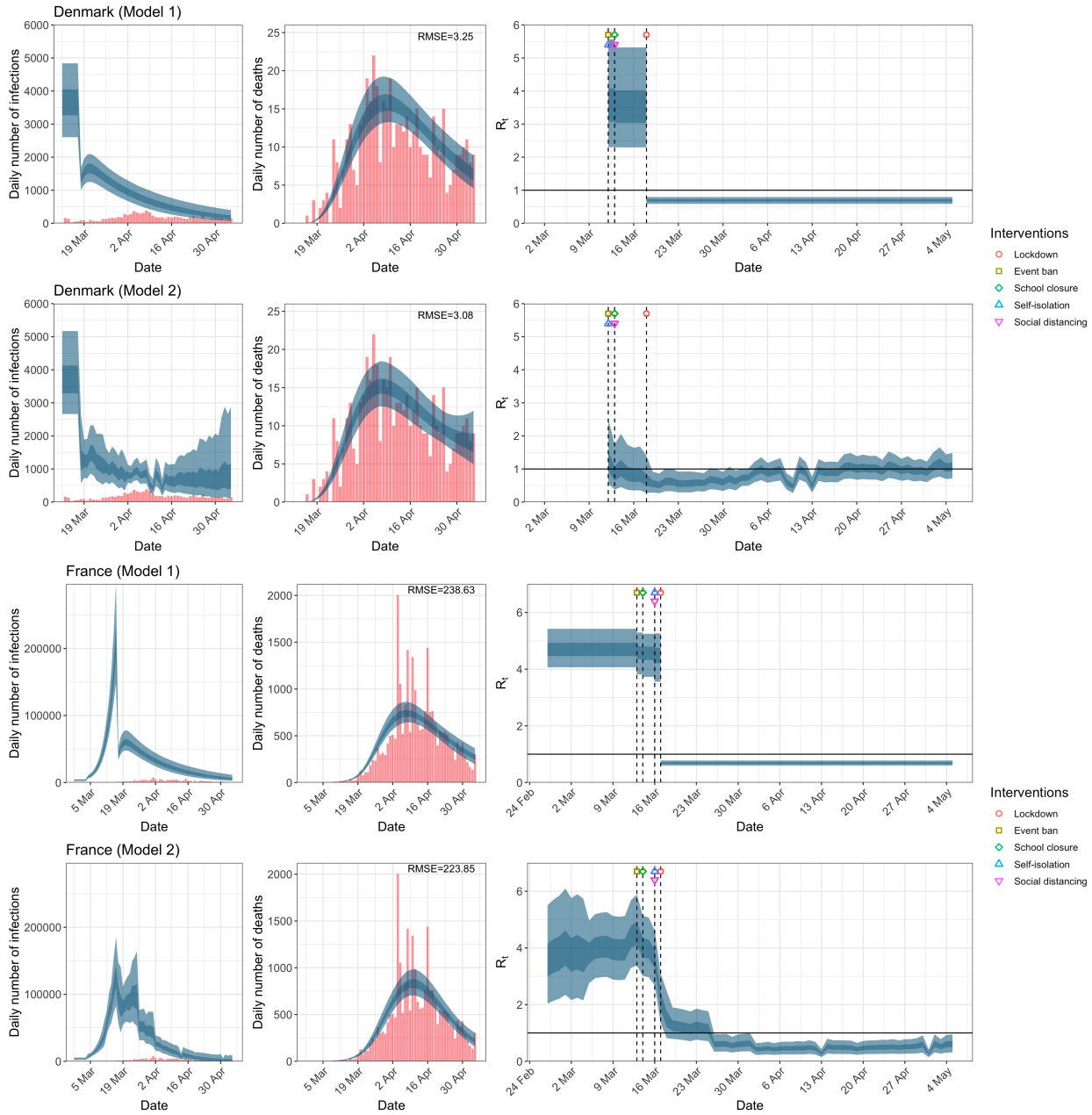
where the subscript k refers to the imposition of various NPIs given in Table 4.

k	NPIs
1	school closure
2	event ban
3	lockdown
4	self-isolation
5	social distancing

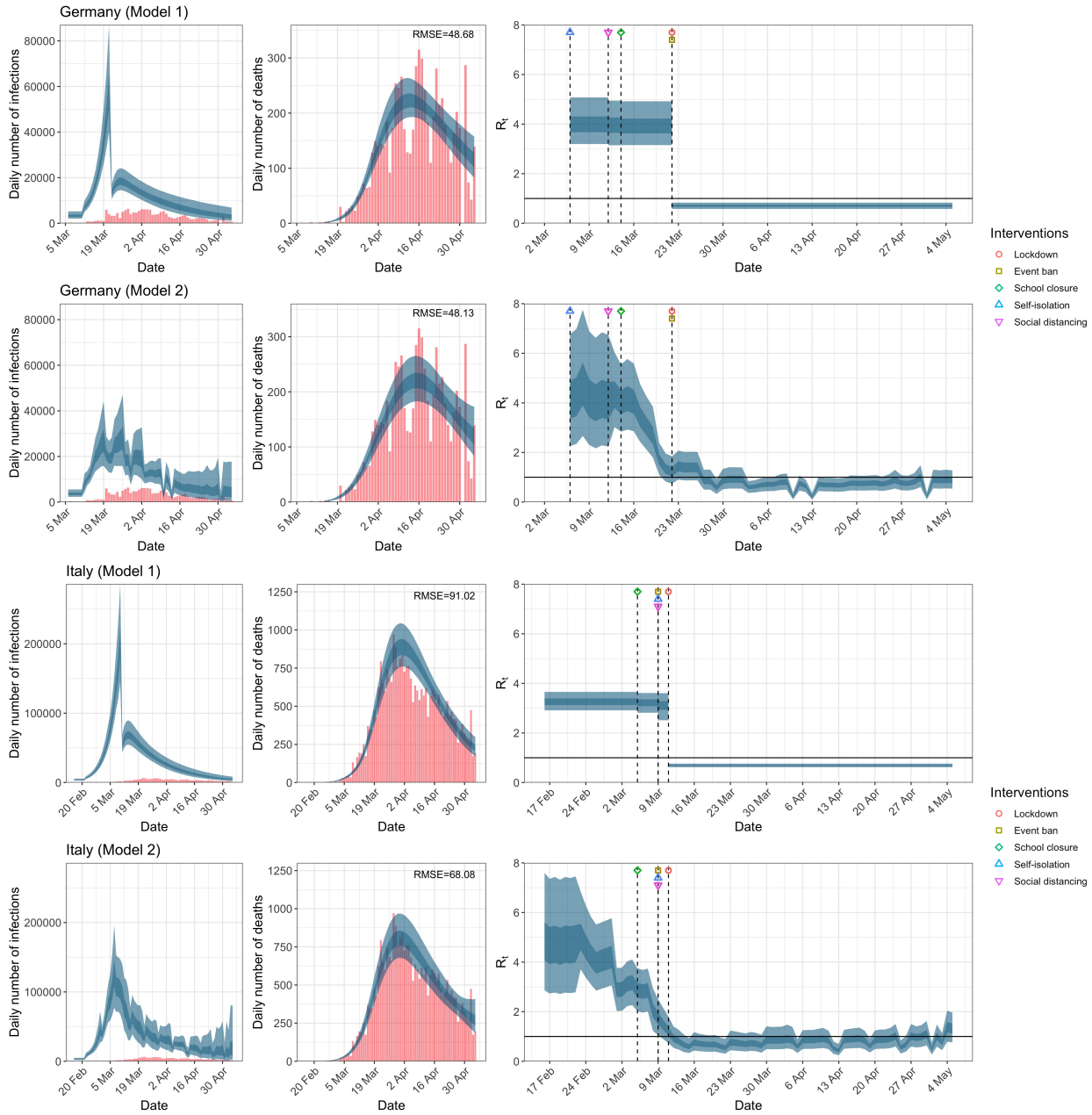
Table 4: Correspondence of subscripts for k to each NPI.



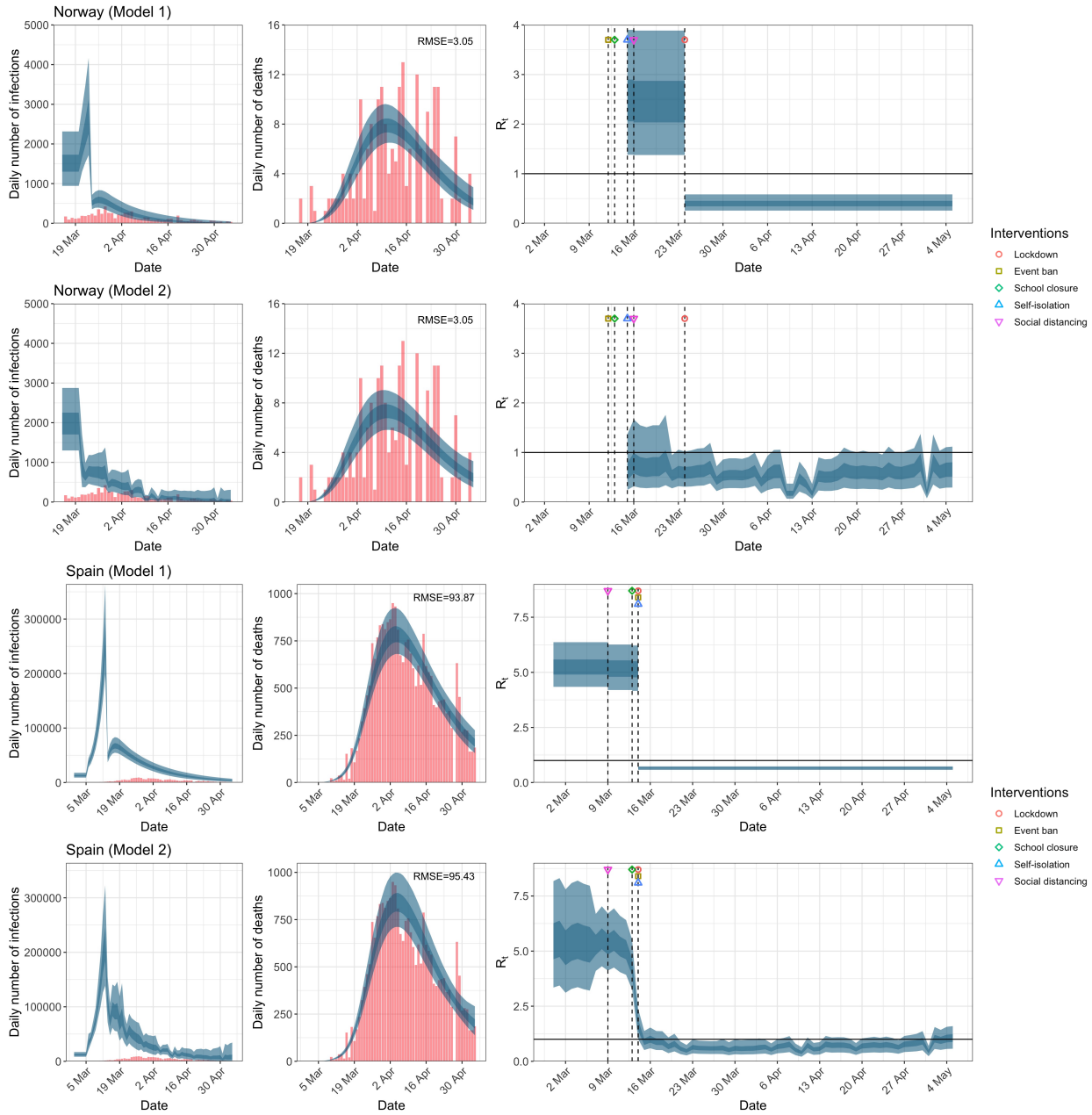
Extended Data Figure 1a: Austria and Belgium. Plots of daily infections, daily deaths and time-varying reproduction number R_t until May 5th. The start time for the plots is 10 days before 10 deaths are recorded. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.



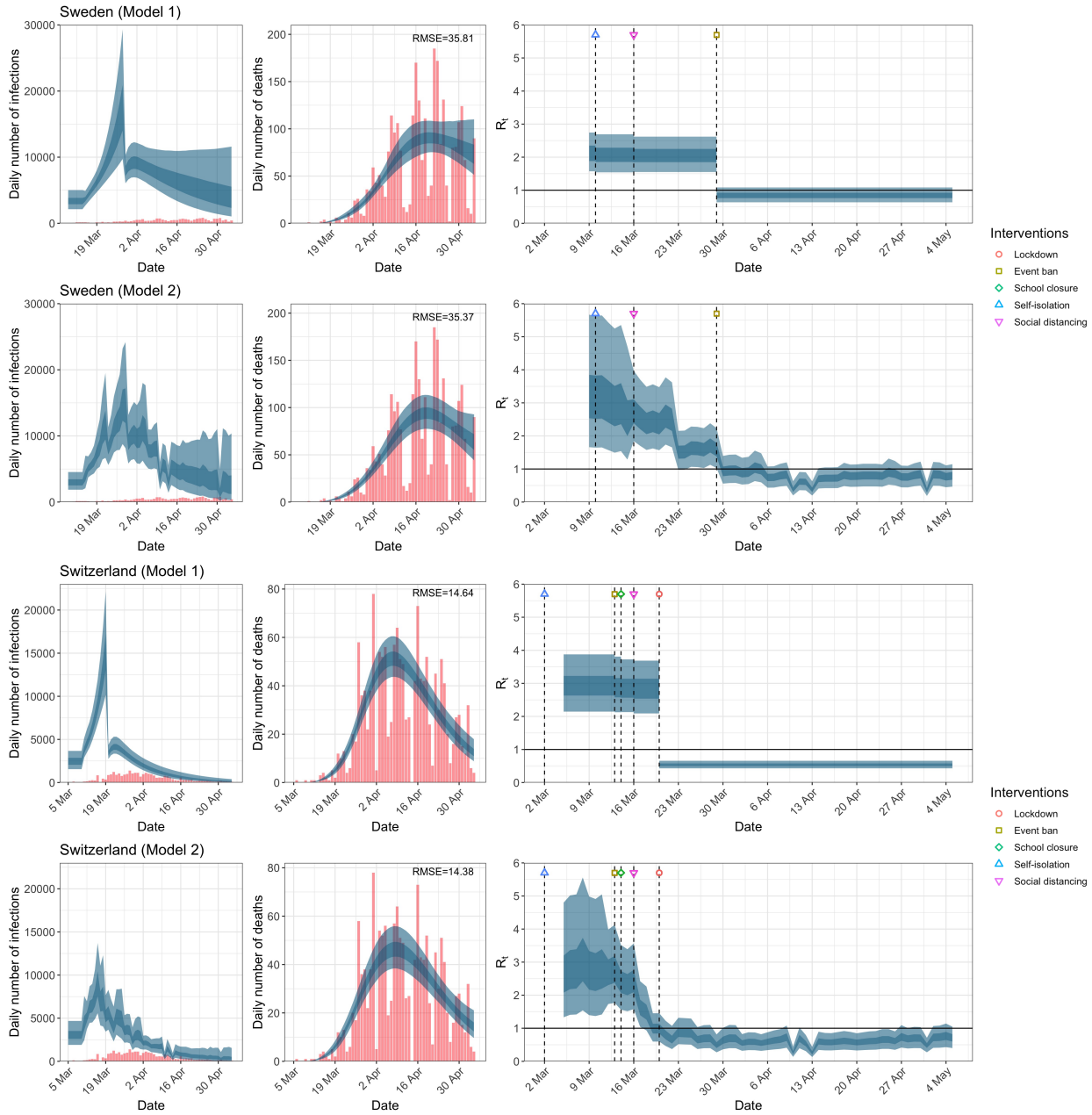
Extended Data Figure 1b: Denmark and France. Plots of daily infections, daily deaths and time-varying reproduction number R_t until May 5th. The start time for the plots is 10 days before 10 deaths are recorded. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.



Extended Data Figure 1c: Germany and Italy. Plots of daily infections, daily deaths and time-varying reproduction number R_t until May 5th. The start time for the plots is 10 days before 10 deaths are recorded. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.



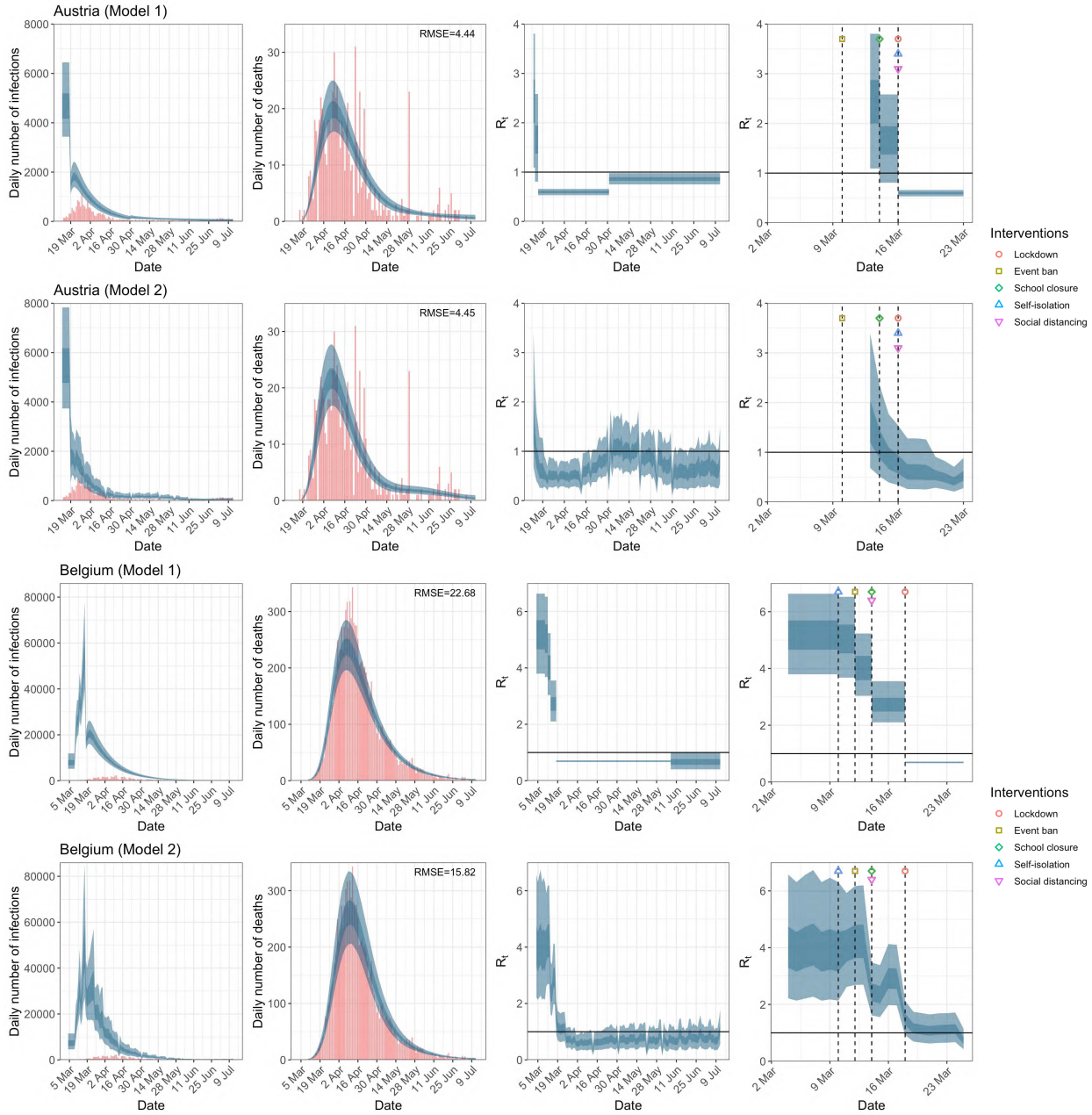
Extended Data Figure 1d: Norway and Spain. Plots of daily infections, daily deaths and time-varying reproduction number R_t until May 5th. The start time for the plots is 10 days before 10 deaths are recorded. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.



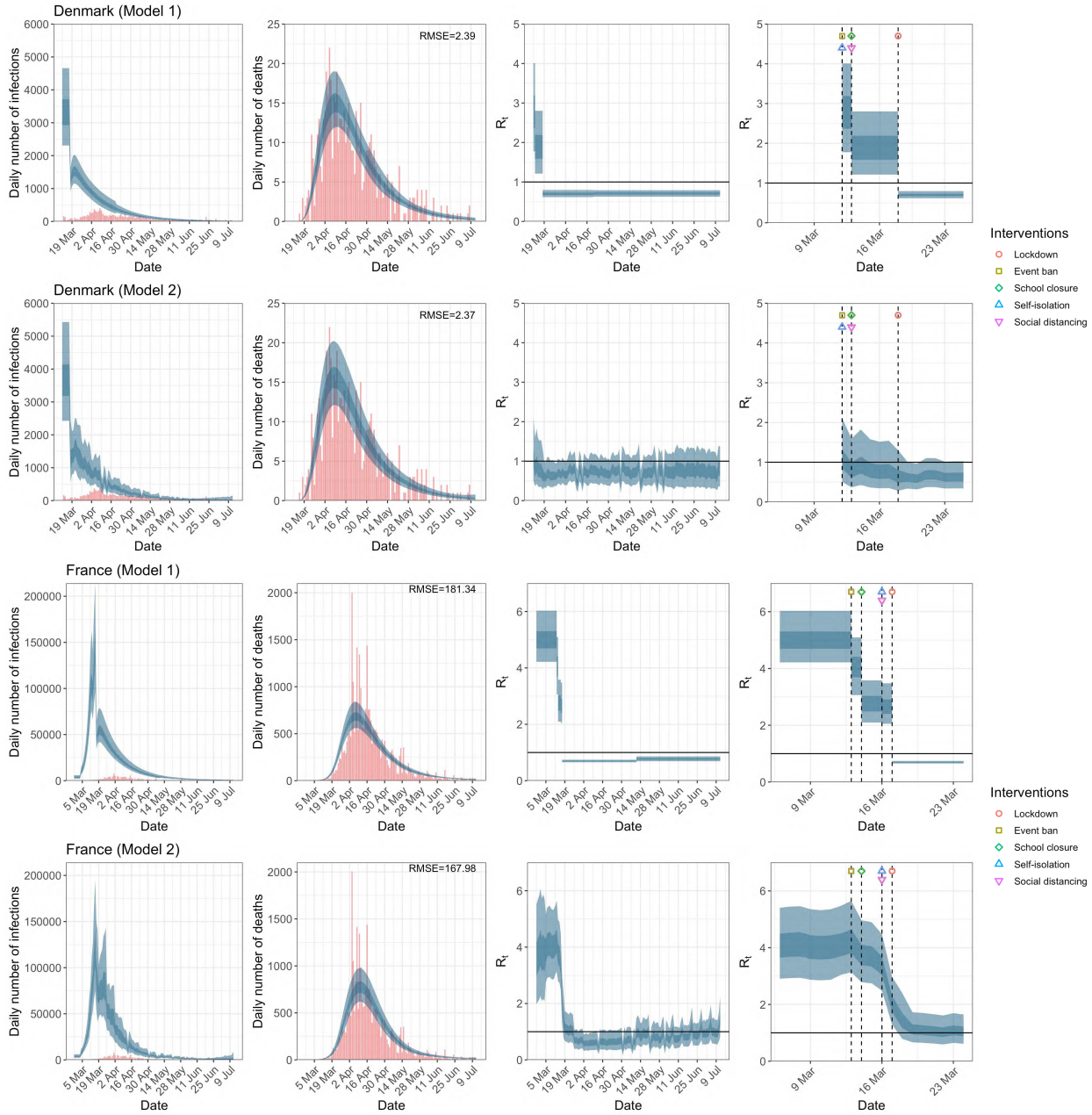
Extended Data Figure 1e: Sweden and Switzerland. Plots of daily infections, daily deaths and time-varying reproduction number R_t until May 5th. The start time for the plots is 10 days before 10 deaths are recorded. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.

Country	School closures	Public events ban	Lockdown
UK	✓	✓	May 13th
Austria	May 18th	✓	May 1st
Belgium	July 1st	✓	June 7th
Denmark	✓	✓	April 20th
France	June 22th	✓	May 11th
Germany	July 7th	✓	May 6th
Greece	June 1st	June 15th	May 30th
Italy	✓	✓	May 4th
Netherlands	June 15th	July 1st	May 11th
Norway	May 11th	June 2nd	April 21st
Portugal	✓	✓	July 5th
Spain	✓	✓	May 26th
Sweden	✗	✓	✗
Switzerland	June 6th	✓	June 21st

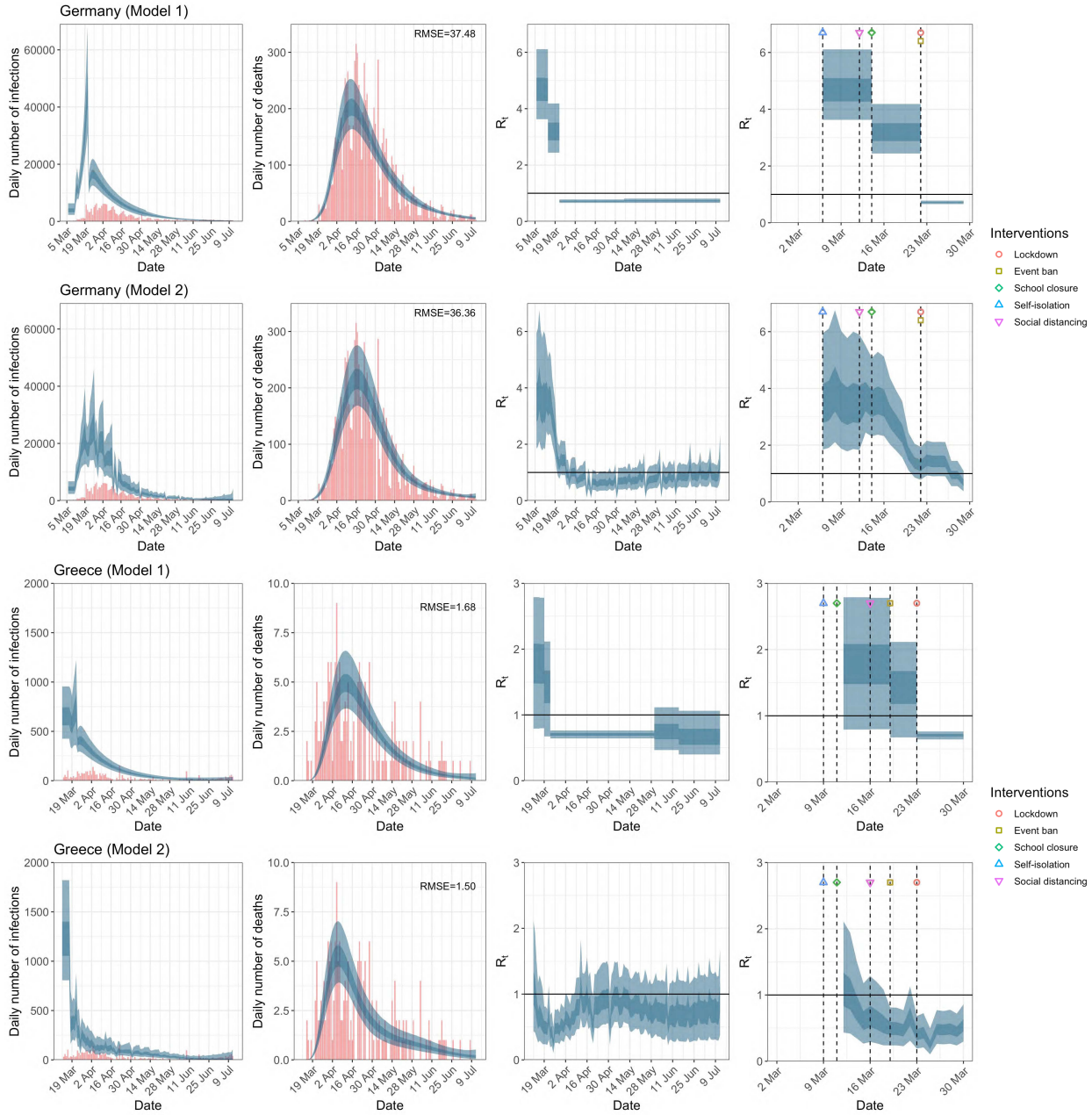
Extended Data Table 1: End dates for school closures[12], public events ban[12] and lockdown in each country[13;14;15]. NPIs that are still in place are shown in ✓, while NPIs that were not implemented are shown in ✗.



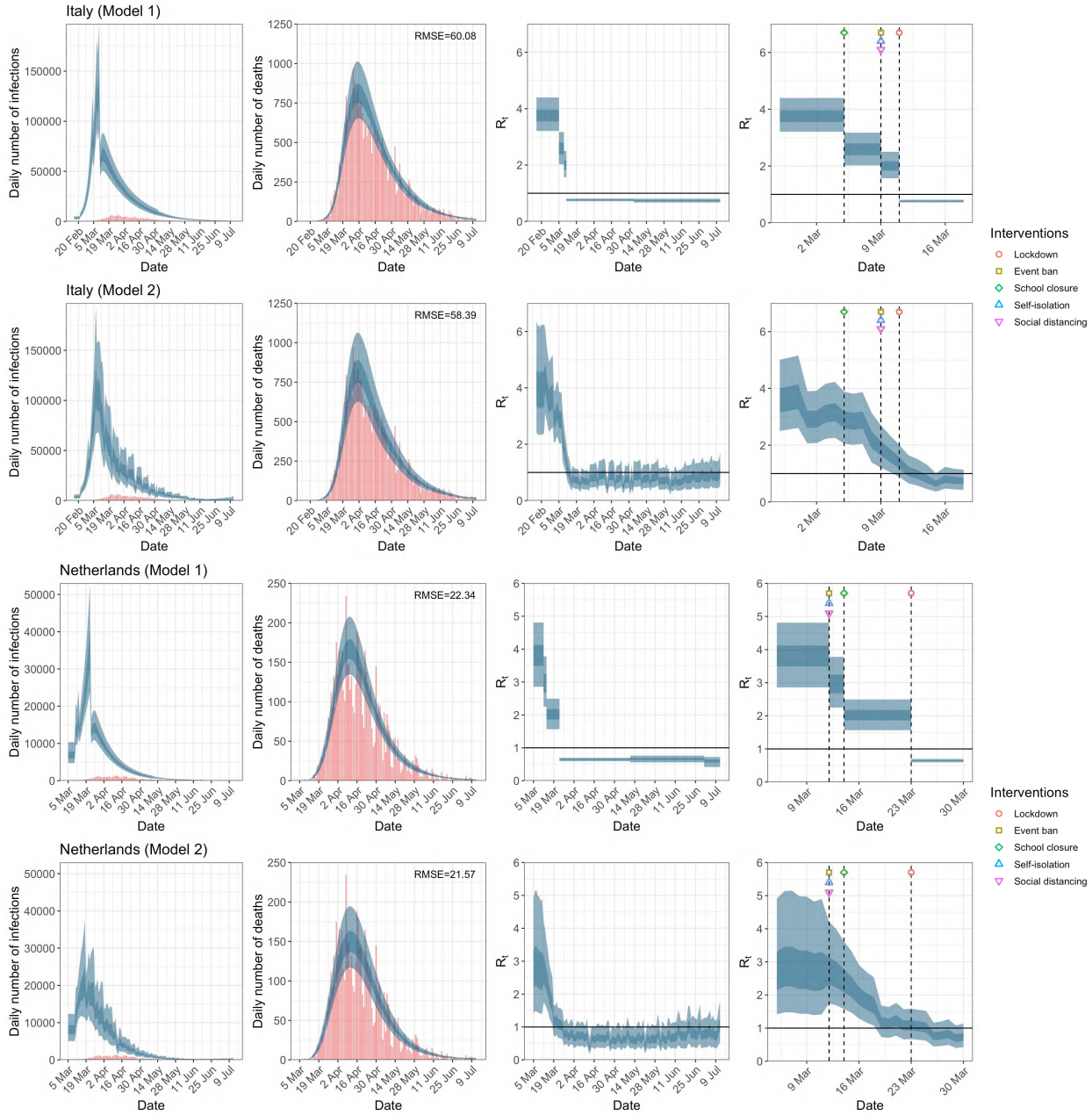
Extended Data Figure 2a: Austria and Belgium. Plots of daily infections, daily deaths and time-varying reproduction number R_t in columns 1–3 until July 12th. The start time for the plots is 10 days before 10 deaths are recorded. Column 4 is a magnification of column 3 around the period of the NPIs. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.



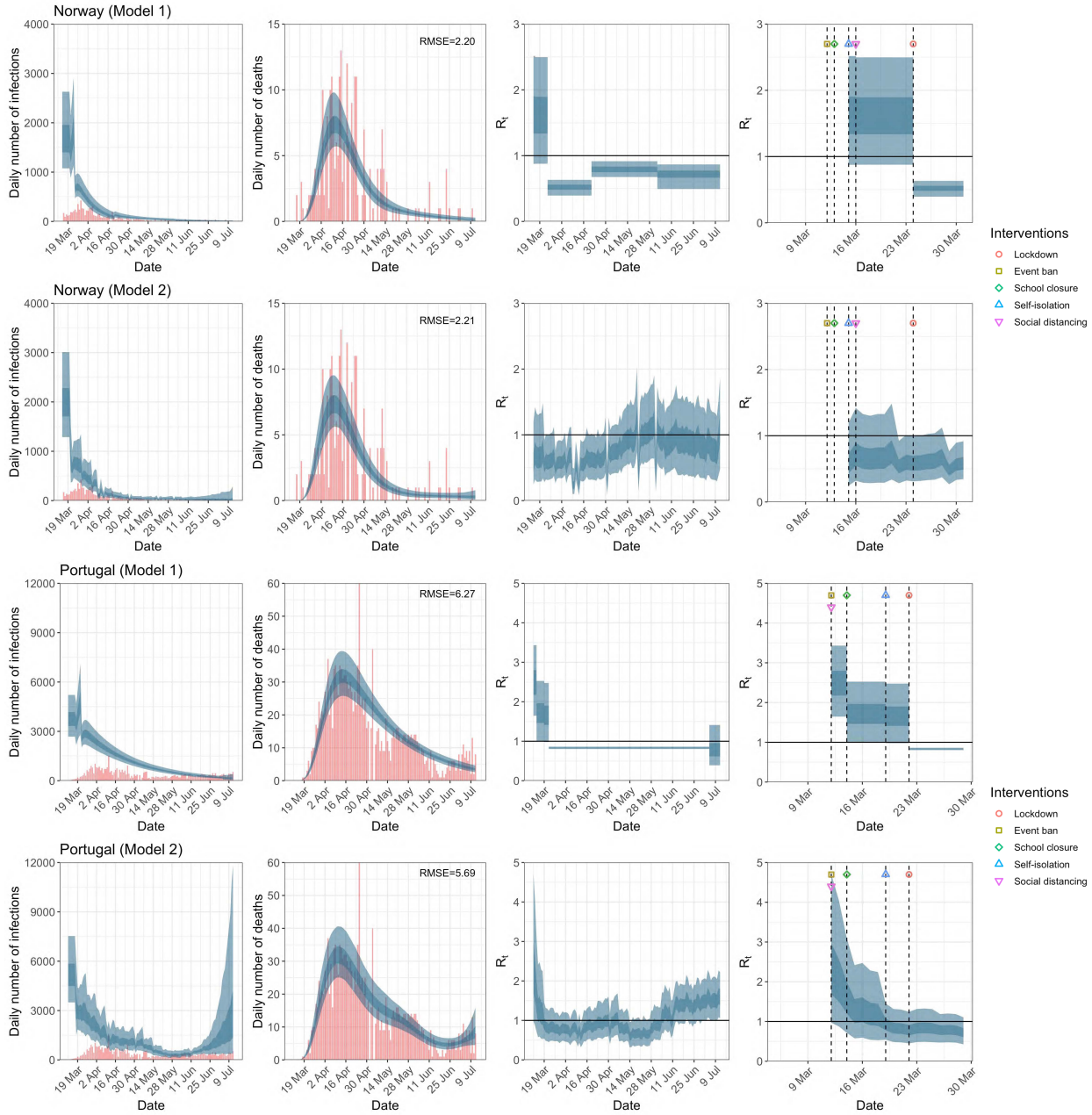
Extended Data Figure 2b: Denmark and France. Plots of daily infections, daily deaths and time-varying reproduction number R_t in columns 1–3 until July 12th. The start time for the plots is 10 days before 10 deaths are recorded. Column 4 is a magnification of column 3 around the period of the NPIs. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.



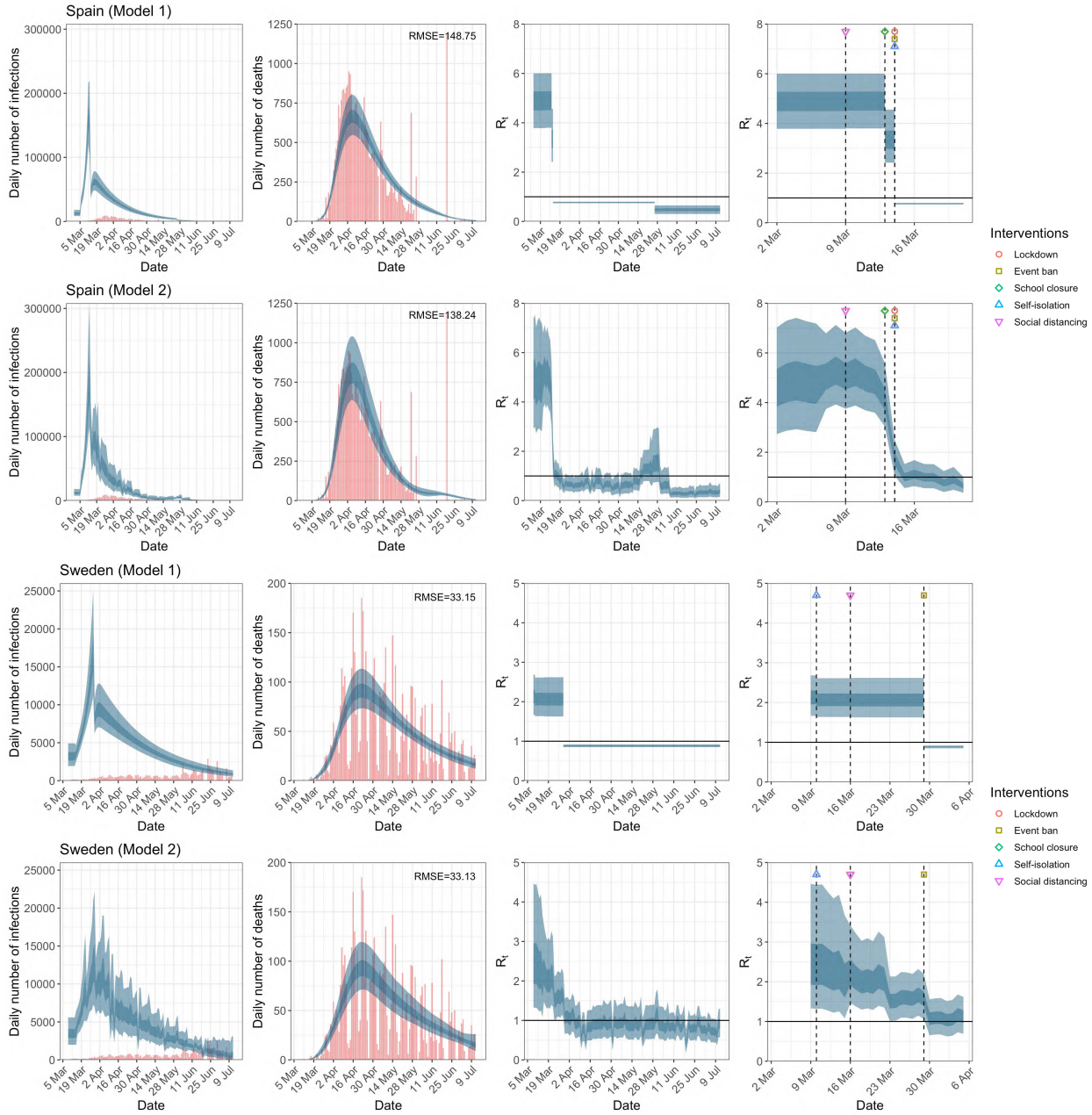
Extended Data Figure 2c: Germany and Greece. Plots of daily infections, daily deaths and time-varying reproduction number R_t in columns 1–3 until July 12th. The start time for the plots is 10 days before 10 deaths are recorded. Column 4 is a magnification of column 3 around the period of the NPIs. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.



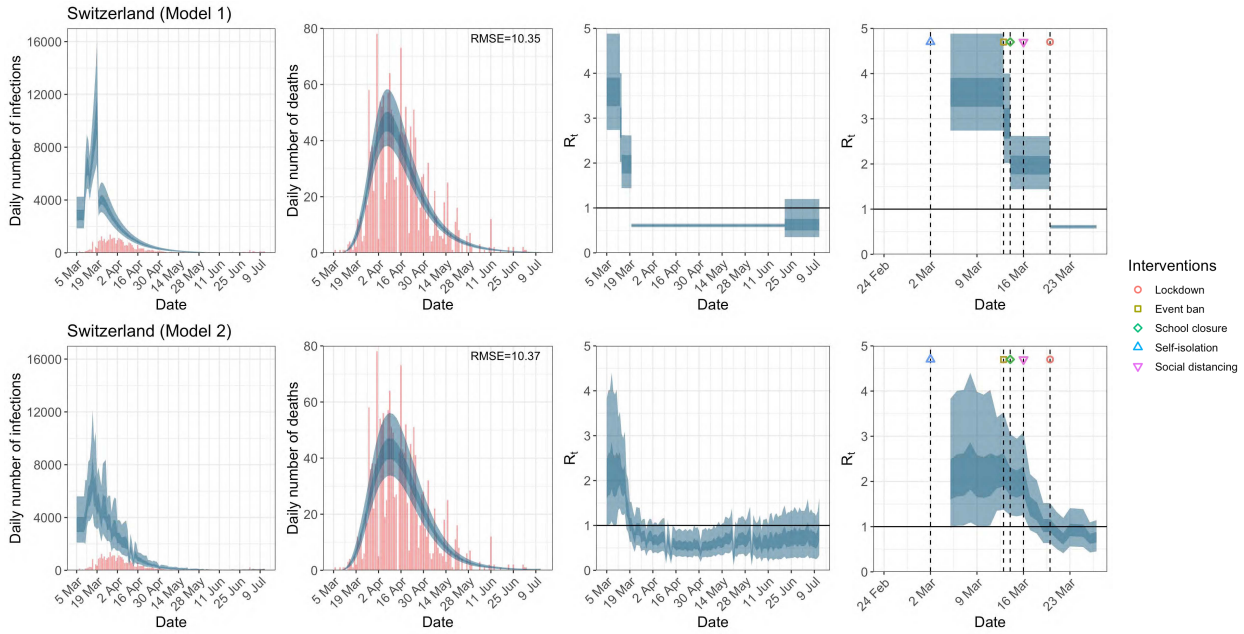
Extended Data Figure 2d: Italy and the Netherlands. Plots of daily infections, daily deaths and time-varying reproduction number R_t in columns 1–3 until July 12th. The start time for the plots is 10 days before 10 deaths are recorded. Column 4 is a magnification of column 3 around the period of the NPIs. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.



Extended Data Figure 2e: Norway and Portugal. Plots of daily infections, daily deaths and time-varying reproduction number R_t in columns 1–3 until July 12th. The start time for the plots is 10 days before 10 deaths are recorded. Column 4 is a magnification of column 3 around the period of the NPIs. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.



Extended Data Figure 2f: Spain and Sweden. Plots of daily infections, daily deaths and time-varying reproduction number R_t in columns 1–3 until July 12th. The start time for the plots is 10 days before 10 deaths are recorded. Column 4 is a magnification of column 3 around the period of the NPIs. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.



Extended Data Figure 2g: Switzerland. Plots of daily infections, daily deaths and time-varying reproduction number R_t in columns 1–3 until July 12th. The start time for the plots is 10 days before 10 deaths are recorded. Column 4 is a magnification of column 3 around the period of the NPIs. Observed counts of daily infections and daily deaths are shown in red, and their corresponding 50% and 95% CIs are shown in dark blue and light blue respectively.

Country	Up to May 5th		Up to July 12th	
	Model 1	Model 2	Model 1	Model 2
UK	**145.98**	146.31	132.78	**130.88**
Austria	**5.79**	**5.79**	**4.44**	4.45
Belgium	72.24	**51.42**	22.68	**15.82**
Denmark	3.25	**3.08**	2.39	**2.37**
France	238.63	**223.85**	181.34	**167.98**
Germany	48.68	**48.13**	37.48	**36.36**
Italy	91.02	**68.08**	60.08	**58.39**
Norway	**3.05**	**3.05**	**2.20**	2.21
Spain	**93.87**	95.43	148.75	**138.24**
Sweden	35.81	**35.37**	33.15	**33.13**
Switzerland	14.64	**14.38**	**10.35**	10.37
Greece			1.68	**1.50**
Netherlands			22.34	**21.57**
Portugal			6.27	**5.69**

Extended Data Table 2: RMSE of daily death counts for Models 1 and 2 for the time period up to May 5th and July 12th.

Time period	Lockdown		Event ban		School closure	
	Imposition	Lifting	Imposition	Lifting	Imposition	Lifting
From February 13th	(0.87, 1.45)	(−0.01,0.08)	(0.00, 0.61)	(−0.01, 0.58)	(−0.01, 0.24)	(−0.01, 0.11)
From March 4th	(0.73, 1.33)	(0.00, 0.15)	(0.08, 0.45)	(0.00, 0.45)	(0.18, 0.64)	–

Extended Data Table 3: 95% credible intervals for the coefficients of imposition and lifting lockdown, event ban and school closure in model 1 up to July 12th for 2 different infection start dates in the UK. For the start date from March 4th, the lifting of school closure is excluded from the model due to non-convergence.