

# Inferring the effective start dates of non-pharmaceutical interventions during COVID-19 outbreaks

Ilia Kohanovski<sup>1</sup>, Uri Obolski<sup>2,3</sup>, and Yoav Ram<sup>1,4,a</sup>

<sup>1</sup>School of Computer Science, Interdisciplinary Center Herzliya, Herzliya 4610101, Israel

<sup>2</sup>School of Public Health, Faculty of Medicine, Tel Aviv University, Tel Aviv 6997801, Israel

<sup>3</sup>Porter School of the Environment and Earth Sciences, Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 6997801, Israel

<sup>4</sup>School of Zoology, Faculty of Life Sciences, Tel Aviv University, Tel Aviv 6997801, Israel

<sup>a</sup>Corresponding author: yoav@yoavram.com, ORCID 0000-0002-9653-4458

January 21, 2021

## Abstract

During Feb-Apr 2020, many countries implemented non-pharmaceutical interventions, such as school closures and lockdowns, with variable schedules, to control the COVID-19 pandemic caused by the SARS-CoV-2 virus. Overall, these interventions seem to have reduced the spread of the pandemic. We hypothesize that the official and effective start date of such interventions can be noticeably different, for example due to slow adoption by the population, or because the authorities and the public are unprepared. We fit an SEIR model to case data from 12 regions to infer the effective start dates of interventions and contrast them with the official dates. We find mostly late, but also early effects of interventions. For example, Italy implemented a nationwide lockdown on Mar 11, but we infer the effective date on Mar 17 ( $^{+3.05}_{-2.01}$  days 95% CI). In contrast, Germany announced a lockdown on Mar 22, but we infer an effective start date on Mar 19 ( $^{+0.92}_{-0.99}$  days 95% CI). We demonstrate that differences between the official and effective start of NPIs can lead to under-estimating their impact, and discuss potential causes and consequences of our results.

Keywords: SEIR, COVID-19, public health, epidemic, infectious disease, NPI

## 26 Introduction

The COVID-19 pandemic has resulted in implementation of extreme non-pharmaceutical interventions (NPIs) in many affected countries. These interventions, from social distancing to lockdowns, are applied in a rapid and widespread fashion. NPIs are designed and assessed using epidemiological models, which follow the dynamics of infection to forecast the effect of different mitigation and suppression strategies on the levels of infection, hospitalization, and fatality. These epidemiological models usually assume that the effect of NPIs on infection dynamics begins at the officially declared date<sup>9,11,17</sup>.

Adoption of public-health recommendations is often critical for effective response to infectious diseases, and has been studied in the context of HIV<sup>15</sup> and vaccination<sup>6,22</sup>, for example. However, behavioural and social change does not occur immediately, but rather requires time to diffuse in the population through media, social networks, and social interactions. Moreover, compliance to NPIs may differ between different interventions and between people with different backgrounds. For example, in a survey of 2,108 adults in the UK during Mar 2020, Atchison et al.<sup>3</sup> found that those over 70 years old were more likely to adopt social distancing than young adults (18-34 years old), and that those with lower income were less likely to be able to work from home and to self-isolate. Similarly, compliance to NPIs may be impacted by personal experiences. Smith et al.<sup>19</sup> have surveyed 6,149 UK adults in late Apr 2020 and found that people who believe they have already had COVID-19 are more likely to think they are immune, and less likely to comply with social distancing guidelines. Compliance may also depend on risk perception as perceived by the the number of domestic cases or even by reported cases in other regions and countries. Interestingly, the perceived risk of COVID-19 infection has likely caused a reduction in the number of influenza-like illness cases in the US starting from mid-Feb 2020<sup>23</sup>.

Here, we hypothesise that there is a significant difference between the official start of NPIs and their effective adoption by the public and therefore their effect on infection dynamics. We use a *Susceptible-Exposed-Infected-Recovered* (SEIR) model and a *Markov Chain Monte Carlo* (MCMC) parameter estimation framework to infer the effective start date of NPIs from publicly available COVID-19 case data in 12 geographical regions. We compare these estimates to the official dates, and find that they

Country	First	Last
Austria	Mar 10 2020	Mar 16 2020
Belgium	Mar 12 2020	Mar 18 2020
Denmark	Mar 12 2020	Mar 18 2020
France	Mar 13 2020	Mar 17 2020
Germany	Mar 12 2020	Mar 22 2020
Italy	Mar 5 2020	Mar 11 2020
Norway	Mar 12 2020	Mar 24 2020
Spain	Mar 9 2020	Mar 14 2020
Sweden	Mar 12 2020	Mar 18 2020
Switzerland	Mar 13 2020	Mar 20 2020
United Kingdom	Mar 16 2020	Mar 24 2020
Wuhan	Jan 23 2020	Jan 23 2020

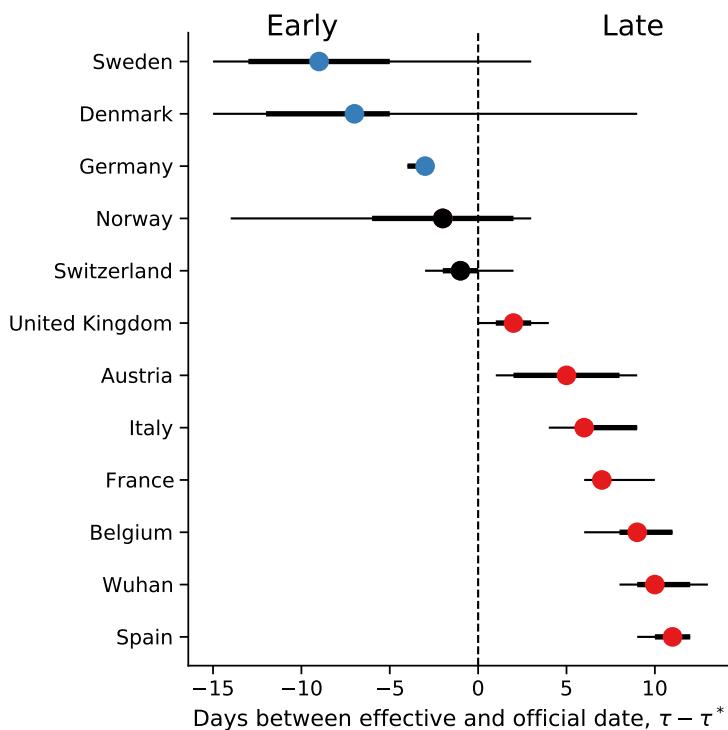
**Table 1: Official start of non-pharmaceutical interventions.** The date of the first intervention is for a ban of public events, or encouragement of social distancing, or for school closures. In all countries except Sweden, the date of the last intervention ( $\tau^*$ ) is for a lockdown. In Sweden, where a lockdown was not ordered during the studied dates, the last date is for school closures. Dates for European countries from Flaxman et al.<sup>9</sup>, date for Wuhan, China from Pei and Shaman<sup>18</sup>. See Figure S1 for a visual presentation.

54 include both late and early effects of NPIs on infection dynamics. We conclude by demonstrating  
 how differences between the official and effective start of NPIs can confound assessments of their  
 56 impacts.

## Results

58 Several studies have described the effects of non-pharmaceutical interventions in different geographical  
 regions<sup>9,11,17</sup>. Some of these studies have assumed that the parameters of the epidemiological model  
 60 change at a specific date (Eq. 6), and set the change date  $\tau$  to the official NPI date  $\tau^*$ , usually the  
 lockdown start date (Table 1). They then fit the model once for time  $t < \tau^*$  and once for time  $t \geq \tau^*$ .  
 62 For example, Li et al.<sup>17</sup> estimate the infection dynamics in China before and after  $\tau^*$ , which is set at  
 Jan 23, 2020. Thereby, they effectively estimate the transmission and reporting rates before and after  
 64  $\tau^*$  separately.

Here, we estimate the joint posterior distribution of the effective start date of NPIs,  $\tau$ , and the  
 66 transmission and reporting rates before and after  $\tau$  from the entire data, rather than splitting the data  
 at  $\tau$ . This is done under the simplifying assumption that all interventions start at a specific date, despite  
 68 the reality that the durations between the first and last NPIs were between 4 and 12 days (Table 1,  
 Figure S1). We then estimate  $\hat{\tau}$  as the median of the marginal posterior distribution of  $\tau$ . Credible

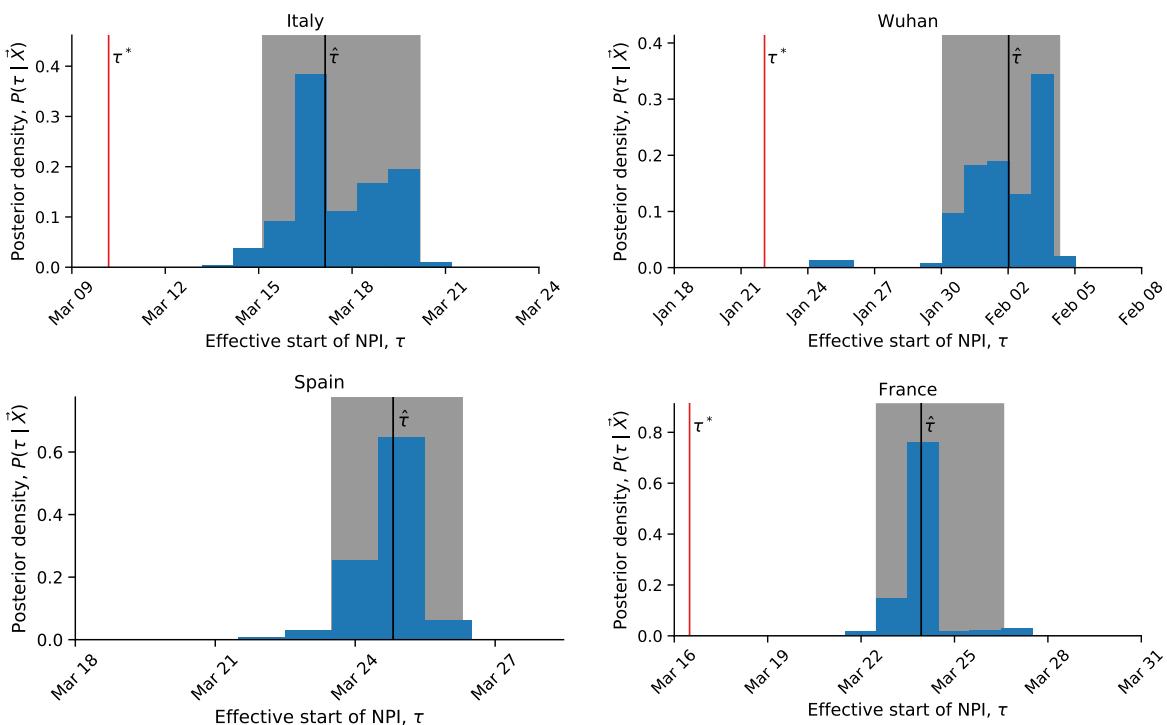


**Figure 1: Official vs. effective start of non-pharmaceutical interventions.** The difference between  $\tau$  the effective and  $\tau^*$  the official start of NPIs is shown for different regions. The effective date is delayed in UK, Austria, Italy, France, Belgium, Spain, and Wuhan, China, compared to the official date (red markers). In contrast, the estimated effective dates in Sweden, Denmark, and Germany are earlier than the official dates (blue markers), although uncertainty is low only for Germany (i.e., zero is not in 95% CI). The credible intervals for Sweden, Denmark, and Norway are especially wide, see text and Figure 3 for possible explanation. Here, the markers show  $\hat{\tau}$ , the marginal posterior median (Table 2).  $\tau^*$  is the last NPI date (a lockdown, except in Sweden; Table 1). Thin and bold lines show 95% and 75% credible intervals, respectively. Figure S2 shows a similar summary when estimating  $\hat{\tau}$  using case data up to Mar 28 rather than Apr 11.

70 intervals (CI) are calculated as the highest density intervals<sup>16</sup>, and their upper and lower boundaries  
 are reported as  $\begin{smallmatrix} +\text{upper} \\ -\text{lower} \end{smallmatrix}$  in days relative to  $\hat{\tau}$ .

72 We compare the posterior predictive plots of a model with a free  $\tau$  with those of a model with  $\tau$  fixed  
 at  $\tau^*$  and a model without  $\tau$  (i.e. transmission and reporting rates are constant). All three models  
 74 were fitted to case data up to Apr 11, used to predict out-of-sample case data up to Apr 24, and these  
 predictions were then compared to the real case data. The model with free  $\tau$  clearly produces better  
 76 and less variable predictions (Figure S3): In all 11 of the European countries, the expected posterior  
 RMSE (root mean squared error) of the out-of-sample predictions is lowest for the model with a free  $\tau$   
 78 (Table S2). When we compare the models using WAIC (Eq. 10), the model with a free  $\tau$  is strongly  
 preferred in 9 out of 12 countries, the exceptions being Norway (only slight preference), Sweden, and  
 80 Denmark (Table S1). Indeed, we estimate low effect of NPIs on transmission in Denmark and Sweden  
 (i.e.  $\lambda = 0.7$  and 0.74, respectively; see Table 2). This may interfere with the inference of  $\tau$  due to  
 82 unidentifiability. Notably, the data for Sweden and Denmark do not have a single "peak" during the  
 evaluated dates, possibly leading to wide credible intervals on  $\tau$  (Figure 1) and poor WAIC in the  
 84 model with free  $\tau$ , whereas the duration between the first and last interventions was especially long in  
 Norway (Table 1).

86 We compare the official  $\tau^*$  and effective  $\tau$  start of NPIs and find that in most regions the effective start  
 of NPIs differs from the official date: in 10 of 12 countries the 75% credible interval on  $\tau$  does not  
 88 include  $\tau^*$  (7 of 12 countries when considering a 95% interval; Figure 1). The exceptions are Norway  
 and Switzerland (see below). The former also has a wide credible interval, perhaps because it has the  
 90 longest duration between the first and last NPIs (Table 1).



**Figure 2: Late effect of non-pharmaceutical interventions.** Posterior distribution of  $\tau$ , the effective start date of NPI, is shown as a histogram of MCMC samples. Red line shows the official last NPI date  $\tau^*$ . Black line shows the estimated effective start date  $\hat{\tau}$ . Shaded area shows a 95% credible interval.

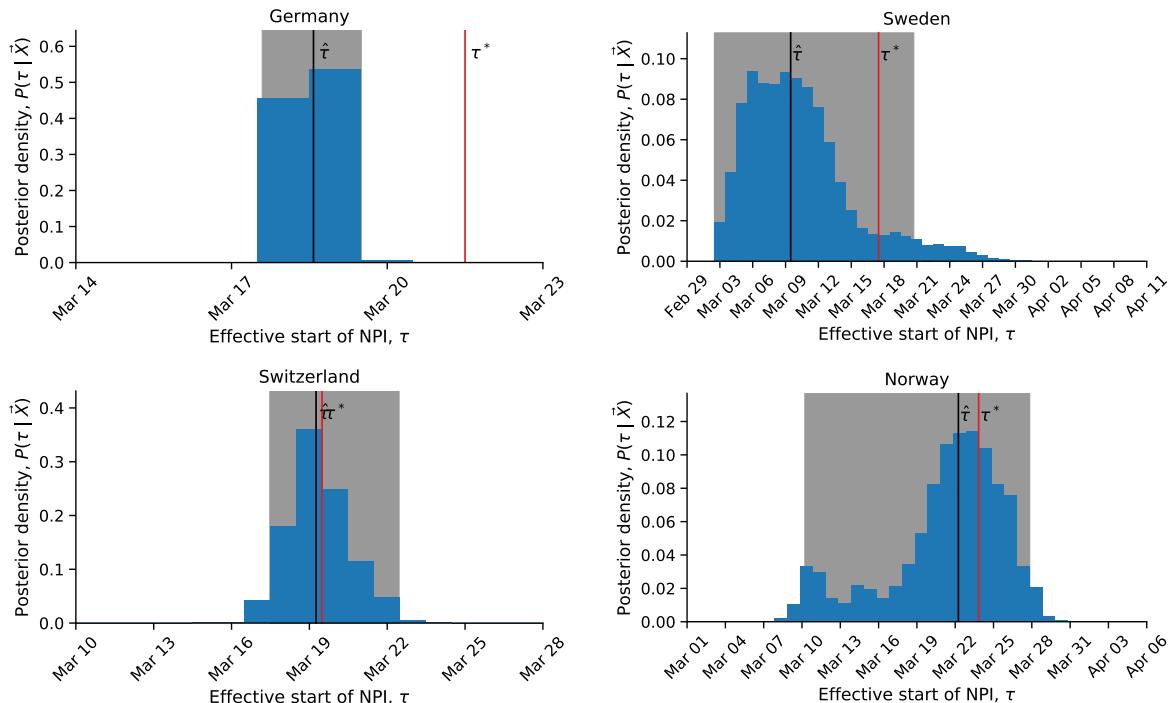
**Late effective start of NPIs.** In half of the examined regions, we estimate that the effective start of  
92 NPIs  $\tau$  is later than the official date  $\tau^*$ .

In Italy, the first case was officially confirmed on Feb 21. School closures were implemented on Mar 5<sup>9</sup>, a lockdown was declared in Northern Italy on Mar 8, with social distancing implemented in the rest of the country, and the lockdown was extended to the entire nation on Mar 11<sup>11</sup>. That is, the first and last official NPI dates are Mar 5 and Mar 11. However, we estimate the effective date  $\hat{\tau}$  six days after the lockdown, at Mar 17 ( $^{+3.05}_{-2.01}$  days 95% CI; Figure 2).

98 In Wuhan, China, a lockdown was ordered on Jan 23<sup>17</sup>, but we estimate the effective start of NPIs to  
 99 be ten days later, at Feb 2 ( $^{+2.29}_{-2.97}$  days 95% CI). Yet, there is low but noticeable posterior probability for  
 100 Jan 25 (Figure 2), for which the effect of NPIs on transmission is considerably lower (Figure S4).

In Spain, social distancing was encouraged starting on Mar 8<sup>9</sup>, but mass gatherings still occurred on Mar 8, including a march of 120,000 people for the [International Women's Day](#), and a football match between [Real Betis](#) and [Real Madrid](#) (final score 2–1) with a crowd of 50,965 in Seville. A national lockdown was only announced on Mar 14<sup>9</sup>. Nevertheless, we estimate the effective start of NPI eleven days later, at Mar 25(<sup>+1.70</sup><sub>-1.43</sub> days 95 %CI, Figure 2).

106 Similarly, in France we estimate the effective start of NPIs at Mar 24 ( $^{+2.65}_{-1.44}$  days 95% CI, Figure 2).  
107 This is a week later than the official lockdown, which started at Mar 17, and more than 10 days after  
108 the earliest NPI, banning of public events, which started on Mar 13<sup>9</sup>.



**Figure 3: Early and exact effect of non-pharmaceutical interventions.** Posterior distribution of  $\tau$ , the effective start date of NPI, is shown as a histogram of MCMC samples. Red line shows the official last NPI date  $\tau^*$ . Black line shows the estimated effective start date  $\hat{\tau}$ . Shaded area shows a 95% credible interval.

**Early effective start of NPIs.** In some regions we estimate an effective start of NPIs  $\tau$  that is *earlier* than the official date  $\tau^*$  (Figure 1). The only conclusive early case is Germany, in which we estimate the effective start of NPIs at Mar 19 ( $^{+0.92}_{-0.99}$  days 95 %CI; Figure 3). This estimate falls between the first and last official NPI dates, Mar 12 and Mar 22 (Table 1). Therefore, when we refer to this case as "early", we mean that the effective date (Mar 19) occurs *before* the official lockdown date (Mar 22), not that it occurs before all NPIs. Interestingly, Germany had the second longest duration between first and last NPIs after Norway (10 and 12 days respectively; Table 1). However, the credible interval for the effective start date in Germany is narrow (1.91 days), whereas it is very wide in Norway

(17.61 days). The significantly earlier estimate of  $\hat{\tau}$  relative to  $\tau^*$  can suggest that early NPIs in  
118 Germany more effectively reduced transmission rates compared to other countries. Another possible  
interpretation is that the German population anticipated the lockdown and began to act accordingly  
120 before it started.

We also estimate an early effective start of NPIs in Denmark, Norway, and Sweden. However, the  
122 credible intervals are quite wide (Figure 1), and in Denmark and Sweden the evidence did not support  
the model with free  $\tau$  over the model with  $\tau$  fixed at the official date (WAIC values in Table S1).  
124 Indeed, Denmark and Sweden had the smallest estimated effect of NPIs on transmission rates, which  
probably hinders our ability to estimate  $\tau$  in these countries (). Moreover, in Sweden the number of  
126 daily cases continued to grow up to Apr 11, rather than "peak" (Figure S3). In Denmark, the opposite  
occurred: there were seemingly two "peaks", on Mar 11 and on Apr 8 (Figure S3); the first "peak" may  
128 be a result of stochastic events, for example due to a large cluster of cases or an accumulation of tests.  
We suspect that these missing and additional "peaks" increase the uncertainty in our inference.

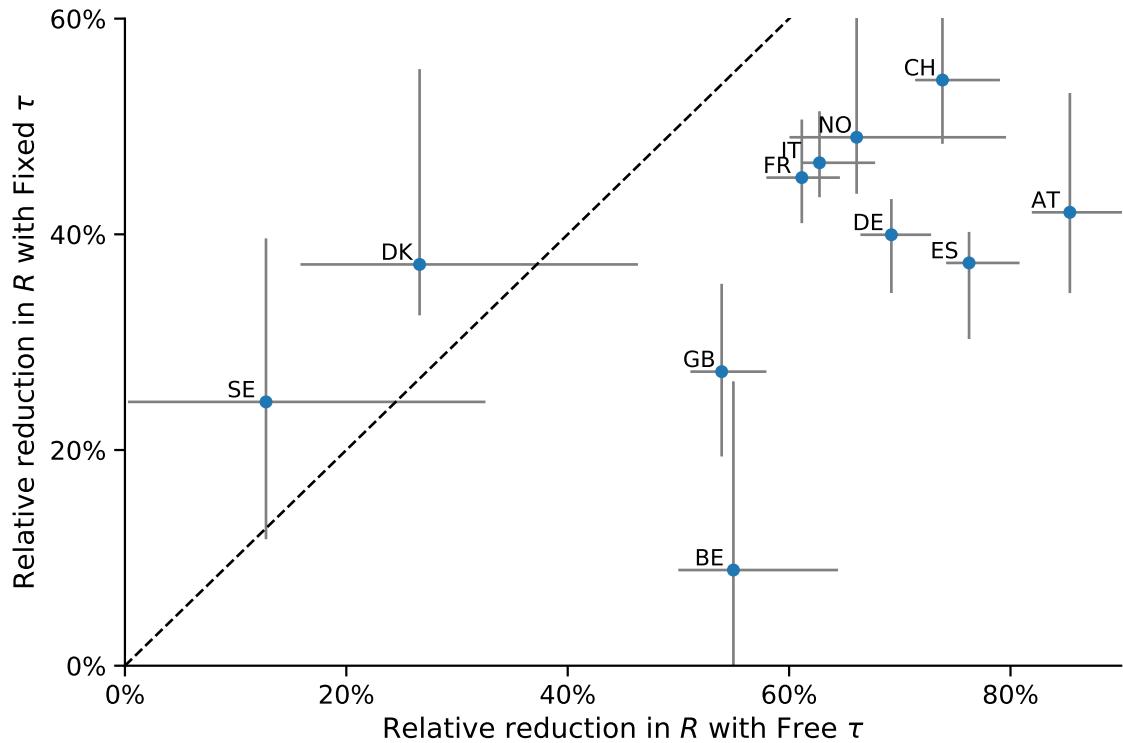
130 The estimated effective date in Norway is Mar 22, two days earlier than the official date of Mar 24.  
However, the posterior distribution is very wide ( $^{+5.59}_{-12.03}$  days 95% CI): it covers the range between  
132 Mar 10, two days before the first NPI, and Mar 27, three days after the last NPI (Table 1, Figure S1).  
The high uncertainty might be due to the long duration between the first and last NPIs; however,  
134 Germany had the second longest duration between first and last NPIs, and the corresponding posterior  
distribution is very narrow (Figure 3). There may also be an unidentifiability issues between  $\tau$  and the  
136 effect of NPIs on transmission (Figure S4).

**Exact effective start of NPIs.** We find one case in which the official and effective dates match and  
138 the credible interval is narrow. Switzerland ordered a national lockdown on Mar 20, after banning  
public events and closing schools on Mar 13 and 14<sup>9</sup>. Indeed, the posterior median  $\hat{\tau}$  is Mar 19  
140 ( $^{+3.2}_{-1.78}$  days 95% CI, see Figure 3). It's also worth mentioning that Switzerland was the first to mandate  
self isolation of confirmed cases<sup>9</sup>.

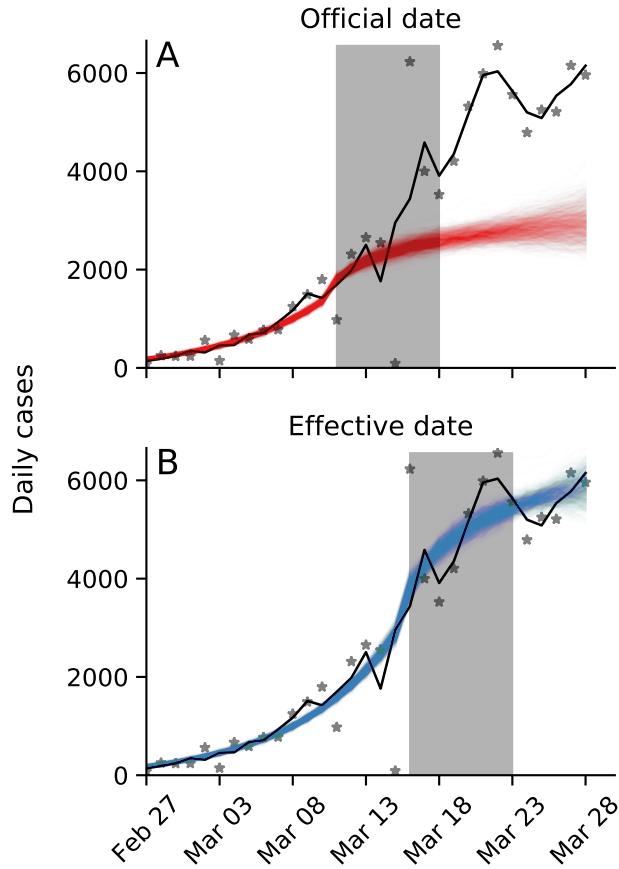
142 **Assessment of impact of NPIs.** The *effective reproduction number*  $R$  is the average number of  
secondary cases caused by an infected individual after an epidemic is already underway<sup>5</sup>. We infer  
144 model-based effective reproduction numbers before and after the implementation of NPIs from model  
parameters (Eq. 7). We then estimate the impact of NPIs as the relative reduction in the reproduction  
146 number<sup>9</sup>. We compare the impacts estimated using the fixed  $\tau$  model and the free  $\tau$  model. That is,  
we compare the impact estimate assuming that NPIs started at the official date  $\tau^*$ , versus the estimate  
148 when inferring the effective start of NPIs from the data. Figure 4 demonstrates that estimates from the  
fixed  $\tau$  model (y-axis) are consistently lower than estimates from the free  $\tau$  model (x-axis), except in  
150 Sweden and Denmark, in which estimation uncertainty is high. These results suggest that the impact  
of past NPIs is likely to be under-estimated by health officials and researchers if they assume NPIs  
152 effectively start at their official dates. The estimated impact can then be interpreted as ineffectiveness  
of the NPIs, leading to more aggressive NPIs being applied.

154 Health officials might also assess the impact of NPIs by comparing model expectations to actual  
number of daily cases at a specific number of days after the intervention began. However, a significant  
156 difference between the official and effective start of interventions can invalidate such assessments.  
This is illustrated in Figure 5 using data and parameters from Italy: a lockdown was officially ordered  
158 on Mar 10 ( $\tau^*$ ), but its late effect on the infection dynamics starts on Mar 17 ( $\hat{\tau}$ ). If health officials  
assume that the dynamics change exactly at  $\tau^*$ , they would expect the number of cases to be within the  
160 red lines (posterior predictions assuming  $\tau = \tau^*$ ). This would lead to a significant under-estimation,  
which might be interpreted as ineffectiveness of NPIs, and therefore more strict guidelines. However,

162 the number of cases would actually follow the blue lines (posterior predictions using  $\tau = \hat{\tau}$ ), which are a good fit to the real data (stars).



**Figure 4: Impact of NPIs is under-estimated when assuming they start at the official date.** Shown are estimates of the relative reductions in  $R$  (the effective reproduction number), which measures the impact of NPIs on disease transmission, in 11 European countries. The y-axis shows estimates when assuming the start of NPIs is fixed at the official date (fixed  $\tau$ ); the x-axis shows estimates when inferring the effective start of NPIs from the data (free  $\tau$ ). The dashed line shows a one-to-one correspondence. Markers and bars denote the posterior median and 50% credible intervals (HDI). The relative reductions in  $R$  are consistently lower for the fixed  $\tau$  model (below the dashed line), except in Sweden and Denmark in which uncertainty is high.



**Figure 5: Late effective start of NPIs leads to under-estimation of daily confirmed cases.** Real number of daily cases in Italy in black (markers: data; line: smoothed using a Savitzky-Golay filter with window length 3). Model posterior predictions are shown as coloured lines (1,000 draws from the posterior distribution). Shaded box illustrates a serial interval of seven days<sup>1</sup> after the start of NPIs. **(A)** Using the official date  $\tau^*$  for the effective start of the NPI, Mar 11, the model under-estimates the number of cases seven days after the start of the NPI. **(B)** Using the estimated date  $\hat{\tau}$  for the effective start of the NPI, Mar 17, the model precisely estimates the number of cases seven days after the start of the NPI. Here, model parameters are best estimates for Italy (Table 2).

## 164 Discussion

We have inferred the effective start date of NPIs in several geographical regions using an SEIR model  
166 under an MCMC parameter estimation framework. We find examples of both late and early effective  
start of NPIs relative to the official date (Figure 1).

168 In most investigated regions we find late effective start of NPIs. For example, in Italy and in Wuhan,  
China, the effective start of the lockdowns seems to have occurred five or more days after the official  
170 date (Figure 2). This difference might be explained, in some cases, by low compliance or non-  
adherence to guidelines: In Italy, for example, the government plan to implement a lockdown in  
172 the Northern provinces leaked to the public, resulting in people leaving these provinces before the  
lockdown started<sup>11</sup>. Late effect of NPIs may also be due to the time required by both the government  
174 and the citizens to prepare for a lockdown, and for new guidelines to be adopted by the population.

In contrast, in some regions we inferred reduced transmission rates even before official lockdowns  
176 were implemented, although this is only conclusive in Germany (Figure 3). An early effective date  
might be due to early adoption of social distancing and similar behavioural adaptations in parts of  
178 the population, possibly due to earlier NPIs or NPIs being applied in other regions. Adoption of

these behaviours may occur via media and social networks, rather than official guidelines, and may be  
180 influenced by increased risk perception due to domestic or international COVID-19 reports<sup>2</sup>. Indeed,  
the evidence supports a change in infection dynamics (i.e. a model with fixed or free  $\tau$ ) even for  
182 Sweden (Table S1, Table S2, Figure S3), where a lockdown was not implemented\*.

Interestingly, the effective start of NPIs in France and Spain is estimated to have started on Mar 24  
184 and 25, respectively, although the official NPI dates differ significantly: the first NPI in France is only  
one day before the last NPI in Spain. The number of daily cases was similar in both countries until  
186 Mar 8, but diverged by Mar 13, reaching much higher numbers in Spain (Figure S5). This may suggest  
correlations between effective starts of NPIs due to global or international events.

188 As expected, we have found that the evidence supports a model in which the transmission rate changes  
at a specific time point over a model with a constant transmission rate (Tables S1 and S2). It may be  
190 interesting to check if the evidence supports a model with *two* or more change-points, rather than one.  
Multiple change-points could reflect escalating NPIs (e.g. school closures followed by lockdowns),  
192 or an intervention followed by a relaxation. However, interpretation of such models will be harder,  
as multiple change-points are also likely to result in parameter unidentifiability, for example due to  
194 simultaneous implementations of NPIs<sup>9</sup>.

As different countries experiment with various intervention strategies, we expect similar shifts to  
196 occur: in some cases the population will be late to comply with new guidelines, whereas in other cases  
the population will adopt either restrictions or relaxations even before they are formally announced.  
198 Attempts to assess the impact of NPIs<sup>4,9</sup> generally assume they start at their official date, and that a  
significant change in the dynamics can be observed roughly seven days after the start of NPIs (due to  
200 the characteristic serial interval of COVID-19<sup>1</sup>). However, late and early effective start of NPIs, such  
as we have inferred, can bias these assessments and lead to wrong conclusions about the impact of  
202 NPIs (Figures 4 and 5).

Our results highlight the complex interaction between personal, regional, and global determinants of  
204 behavioral response to an epidemic. Therefore, we emphasize the need to further study heterogeneity  
in compliance and behavior over both time and space. This can be accomplished both by surveying  
206 differences in compliance within and between populations<sup>3</sup>, and by incorporating specific behavioral  
models into epidemiological models<sup>2,7,21</sup>.

## 208 Models and Methods

**Data.** We use daily confirmed case data  $\mathbf{X} = (X_1, \dots, X_T)$  from 12 regions during Jan–Apr 2020. These  
210 incidence data summarise the number of individuals  $X_t$  tested positive for SARS-CoV-2 (using RT-qPCR tests)  
on each day  $t$ . Data for Wuhan, China, from Jan 10 to Feb 8, retrieved from Pei and Shaman<sup>18</sup>. Data for  
212 11 European countries, from Feb 20 to Apr 24, retrieved from Flaxman et al.<sup>9</sup>. Where there were multiple  
sequences of days with zero confirmed cases (e.g. France), we cropped the data to begin with the last sequence  
214 so that our analysis focuses on the first sustained outbreak rather than isolated imported cases. For official NPI  
dates see Table 1.

216 **SEIR model.** We model SARS-CoV-2 infection dynamics by following the number of susceptible  $S$ , exposed  
 $E$ , reported infected  $I_r$ , unreported infected  $I_u$ , and recovered  $R$  individuals in a population of size  $N$ . This  
218 model distinguishes between reported and unreported infected individuals: the reported infected are those that  
have enough symptoms to eventually be tested and thus appear in daily case reports, to which we fit the model.  
220 This model is inspired by Li et al.<sup>17</sup> and Pei and Shaman<sup>18</sup>, who used a similar model with multiple regions

---

\*Sweden banned public events on Mar 12, encouraged social distancing on Mar 16, and closed schools on Mar 18<sup>9</sup>.

and constant transmission and reporting rates to study COVID-19 dynamics in China and in the continental US.

Susceptible ( $S$ ) individuals become exposed due to contact with reported or unreported infected individuals ( $I_r$  or  $I_u$ ) at a rate  $\beta_t$  or  $\mu\beta_t$ , respectively. The parameter  $0 < \mu < 1$  represents the decreased transmission rate from unreported infected individuals, who are often subclinical or even asymptomatic<sup>8,20</sup>. The transmission rate  $\beta_t \geq 0$  may change over time  $t$  due to behavioural changes of both susceptible and infected individuals. Exposed individuals, after an average latent period of  $Z$  days, become reported infected with probability  $\alpha_t$  or unreported infected with probability  $(1 - \alpha_t)$ . The reporting rate  $0 < \alpha_t < 1$  may also change over time due to changes in human behaviour. Infected individuals remain infectious for an average period of  $D$  days, after which they either recover, or become ill enough to be quarantined. In either case, they no longer infect other individuals, and therefore effectively become recovered ( $R$ ). The model is described by the following set of equations,

$$\begin{aligned} \frac{dS}{dt} &= -\beta_t S \left( \frac{I_r}{N} + \mu \frac{I_u}{N} \right) \\ \frac{dE}{dt} &= \beta_t S \left( \frac{I_r}{N} + \mu \frac{I_u}{N} \right) - \frac{E}{Z} \\ \frac{dI_r}{dt} &= \alpha_t \frac{E}{Z} - \frac{I_r}{D} \\ \frac{dI_u}{dt} &= (1 - \alpha_t) \frac{E}{Z} - \frac{I_u}{D} \\ \frac{dR}{dt} &= \frac{I_r}{D} + \frac{I_u}{D}, \end{aligned} \quad (1)$$

where  $N$  is the population size. The initial numbers of exposed  $E(0)$  and unreported infected  $I_u(0)$  are free model parameters (i.e. inferred from the data), whereas the initial number of reported infected and recovered is assumed to be zero,  $I_r(0) = R(0) = 0$ , and the number of susceptible is  $S(0) = N - E(0) - I_u(0)$ .

**Likelihood function.** For a given vector  $\theta$  of model parameters the *expected* cumulative number of reported infected individuals ( $I_r$ ) until day  $t$ , following Eq. 1, is

$$Y_t(\theta) = \int_0^t \alpha_s \frac{E(s)}{Z} ds, \quad Y_0 = 0. \quad (2)$$

We assume that reported infected individuals are confirmed and therefore observed in the daily case report of day  $t$  with probability  $p_t$  (note that an individual can only be observed once, and that  $p_t$  may change over time, but  $t$  is a specific date rather than the time elapsed since the individual was infected). We denote by  $X_t$  the *observed* number of confirmed cases in day  $t$ , and by  $\tilde{X}_t$  the cumulative number of confirmed cases until end of day  $t$ ,

$$\tilde{X}_t = \sum_{i=1}^t X_i. \quad (3)$$

Therefore, at day  $t$  the number of reported infected yet-to-be confirmed individuals is  $(Y_t(\theta) - \tilde{X}_{t-1})$ . We assume that  $X_t$  conditioned on  $\tilde{X}_{t-1}$  is Poisson distributed, such that

$$\begin{aligned} (X_1 | \theta) &\sim Poi(Y_1(\theta) \cdot p_1), \\ (X_t | \tilde{X}_{t-1}, \theta) &\sim Poi((Y_t(\theta) - \tilde{X}_{t-1}) \cdot p_t), \quad t = 2, \dots, T. \end{aligned} \quad (4)$$

Hence, the *likelihood function*  $\mathcal{L}(\theta | \mathbf{X})$  for a parameter vector  $\theta$  given the confirmed case data  $\mathbf{X} = (X_1, \dots, X_T)$  is defined by the probability to observe  $\mathbf{X}$  given  $\theta$ ,

$$\mathcal{L}(\theta | \mathbf{X}) = P(\mathbf{X} | \theta) = P(X_1 | \theta) \cdot P(X_2 | \tilde{X}_1, \theta) \cdots P(X_T | \tilde{X}_{T-1}, \theta). \quad (5)$$

252 **NPI model.** To model non-pharmaceutical interventions (NPIs), we set the start of the NPIs to day  $\tau$  and  
253 define

$$254 \quad \beta_t = \begin{cases} \beta, & t < \tau \\ \beta\lambda, & t \geq \tau \end{cases}, \quad \alpha_t = \begin{cases} \alpha_1, & t < \tau \\ \alpha_2, & t \geq \tau \end{cases}, \quad p_t = \begin{cases} 1/9, & t < \tau \\ 1/6, & t \geq \tau \end{cases}, \quad (6)$$

255 where  $0 < \lambda < 1$ . The values for  $p_t$  follow Li et al.<sup>17</sup>, who estimated the average time between infection and  
256 reporting in Wuhan, China, at 9 days before the start of NPIs and 6 days after start of NPIs.

Following Li et al.<sup>17</sup>, the effective reproduction numbers before and after the start of NPIs are

$$258 \quad R_1 = \alpha_1\beta D + (1 - \alpha_1)\mu\beta D, \\ R_2 = \alpha_2\lambda\beta D + (1 - \alpha_2)\mu\lambda\beta D. \quad (7)$$

The relative reduction in the effective reproduction number due to NPIs is  $\frac{R_1 - R_2}{R_1}$ .

260 **Parameter estimation.** To estimate the model parameters from the daily case data  $\mathbf{X}$ , we apply a Bayesian  
261 inference approach. Model fitting was calibrated for case data up to Mar 28, and then applied to data up to  
262 Apr 11 (for Wuhan, China, model fitting was applied for data up to Feb 8.) We start our model  $\Delta t$  days<sup>11</sup> before  
263 the outbreak (defined as consecutive days with increasing confirmed cases) in each country. The model in Eqs. 1  
264 and 6 is parameterised by the vector  $\theta$ , where

$$\theta = (Z, D, \mu, \beta, \lambda, \alpha_1, \alpha_2, E(0), I_u(0), \Delta t, \tau). \quad (8)$$

266 The likelihood function is defined in Eq. 5. We defined the following prior distributions on the model parameters  
 $P(\theta)$ :

$$\begin{aligned} Z &\sim Uniform(2, 5) \\ D &\sim Uniform(2, 5) \\ \mu &\sim Uniform(0.2, 1) \\ \beta &\sim Uniform(0.8, 1.5) \\ \lambda &\sim Uniform(0, 1) \\ 268 \quad \alpha_1, \alpha_2 &\sim Uniform(0.02, 1) \\ E(0) &\sim Uniform(0, 3000) \\ I_u(0) &\sim Uniform(0, 3000) \\ \Delta t &\sim Uniform(1, 5) \\ \tau &\sim TruncatedNormal\left(\frac{\tau^* + \tau^0}{2}, \frac{\tau^* - \tau^0}{2}, 5, T - 2\right), \end{aligned} \quad (9)$$

where the prior for  $\tau$  is a truncated normal distribution shaped so that the date of the first and last NPI,  $\tau^0$   
270 and  $\tau^*$  (Table 1), are at minus and plus one standard deviation, and taking values only between 5 and  $T - 2$ ,  
271 where  $T$  is the number of days in the data  $\mathbf{X}$ . We also tested an uninformative uniform prior,  $Uniform(1, T - 2)$ .  
272 WAIC (Eq. 10) of a model with this uniform prior was either higher, or lower by less than 2, compared to  
273 WAIC of a model with the truncated normal prior. The uninformative prior resulted in non-negligible posterior  
274 probability for unreasonable  $\tau$  values, such as Mar 1 in the United Kingdom. We therefore decided to use the  
275 more informative truncated normal prior for  $\tau$ . Other priors follow Li et al.<sup>17</sup>, with the following exceptions.  
276  $\lambda$  is used to ensure transmission rates are lower after the start of the NPIs ( $\lambda < 1$ ). We checked values of  $\Delta t$   
277 larger than five days and found they generally produce lower likelihood and unreasonable parameter estimates,  
278 and therefore chose  $Uniform(1, 5)$  as the prior for  $\Delta t$ . We also tried to estimate the value of  $p_t$  before and after  
279  $\tau$ , instead of keeping it fixed at 1/9 and 1/6. The model with fixed values was supported by the evidence (lower  
280 WAIC, see below) in 9 of 12 countries. Moreover, the estimates for Wuhan, China were 1/9 and 1/6, as in Li  
et al.<sup>17</sup>.

282 The posterior distribution of the model parameters  $P(\theta | \mathbf{X})$  is estimated using the affine-invariant ensemble  
283 sampler for Markov chain Monte Carlo (MCMC)<sup>13</sup>, implemented in the emcee Python package<sup>10</sup>. We use

284 the default configuration using the stretch move with stretch scale parameter  $a = 2$ . For the main analysis we  
 285 use 50 chains (or walkers) per region, with 7M samples per chain (no thinning was applied). The *integrated*  
 286 *autocorrelation time* (IAT)<sup>10,13</sup> was averaged across parameters and chains for each region. The average IAT  
 287 was between 32K for Wuhan, China, and 187K for Germany (Figure S6). We examined the trace plots for  $\tau$   
 288 in all regions (Figure S7). All chains seem to converge to the stationary distribution, in most cases before 2M  
 289 samples. Thus, we discarded the first 2M samples as burn-in samples. The only exception is Spain, in which  
 290 a single chain converged at around 6M samples. We considered this chain as part of the burn-in and removed  
 291 it from the analysis. Therefore, 50 chains with 5M samples and IAT between 32K and 187K give an effective  
 292 sample size between 1,336 and 7,812. We ran additional chains with 2M samples and a different initialization  
 293 (i.e. seed). The estimated posterior distributions of  $\tau$  were very similar to our main analysis, further increasing  
 294 our confidence in the convergence of our inference.

**Model comparison.** We perform model selection using two methods. First, we compute WAIC (widely  
 295 applicable information criterion)<sup>12</sup>,

$$WAIC(\theta | \mathbf{X}) = -2 \log \mathbb{E}[\mathcal{L}(\theta | \mathbf{X})] + 2\mathbb{V}[\log \mathcal{L}(\theta | \mathbf{X})] \quad (10)$$

298 where  $\mathbb{E}[\cdot]$  and  $\mathbb{V}[\cdot]$  are the expectation and variance operators taken over the posterior distribution  $P(\theta | \mathbf{X})$ .  
 299 We compare models by reporting their relative WAIC; lower is better (Table S1).

300 We also plot posterior predictions: we sample 1,000 parameter vectors from the posterior distribution  $P(\theta | \mathbf{X})$   
 301 fitted to data up to Apr 11, use these parameter vectors to simulate the SEIR model (Eq. 1) up to Apr 24, and plot  
 302 the predicted dynamics (Figure S3). Both the accuracy (i.e. overlap of data and prediction) and the precision  
 303 (i.e. the tightness of the predictions) are good ways to visually compare models. We also compute the expected  
 304 posterior RMSE (root mean squared error) of these predictions (Table S2).

## Declarations

306 **Ethics approval and consent to participate.** Not applicable.

**Consent for publication.** Not applicable.

308 **Availability of data and materials.** We use Python 3 with NumPy, Matplotlib, SciPy, Pandas, Seaborn, and  
 309 emcee. Source code will be publicly available under a permissive open-source license at [github.com/yoavram-lab/EffectiveNPI](https://github.com/yoavram-lab/EffectiveNPI). We used freely available data, see source code repository and above *Data* section for details.

312 **Competing interests.** The authors declare that they have no competing interests.

314 **Funding.** This work was supported in part by the Israel Science Foundation (3811/19 and 552/19, YR). The funding body had no role in designing, performing, or analyzing the study.

316 **Authors' contributions.** UO and YR designed the research. IK and YR performed the research and wrote the manuscript. All authors read and approved the final manuscript.

**Acknowledgements.** We thank Yinon M. Bar-On, Lilach Hadany, Oren Kolodny, and Zohar Yakhini.

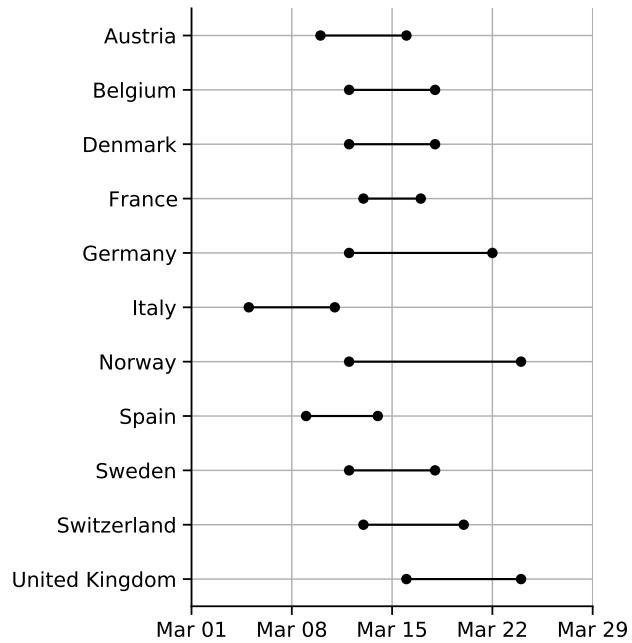
Region	$\tau^*$	$\hat{\tau}$	$HDI_{75\%}$	$HDI_{95\%}$	Z	D	$\mu$	$\beta$	$\alpha_1$	$\lambda$	$\alpha_2$	$E(0)$	$I_u(0)$	$\Delta t$		
Austria	Mar 16	Mar 21	-2.8240	2.5130	-4.2967	3.5110	3.9362	0.63066	0.1468	0.1592	0.1370	0.2667	128.6634	111.2741	2.2407	
Belgium	Mar 18	Mar 27	-1.5528	2.3908	-3.4283	2.3908	4.0266	3.6358	0.5031	0.0780	0.2536	0.4572	0.1927	327.7634	417.8771	2.1455
Denmark	Mar 18	Mar 11	-5.3104	2.5859	-7.4123	15.7701	4.0140	3.4301	0.4149	0.0594	0.1546	0.6977	0.2041	268.2553	370.4863	2.2678
France	Mar 17	Mar 24	-0.4310	0.5557	-1.4388	2.6488	4.2107	3.0919	0.4713	1.0555	0.3823	0.3789	0.4142	412.7334	1324.2711	1.6487
Germany	Mar 22	Mar 19	-0.5544	0.9207	-0.9874	0.9205	3.3868	3.6944	0.6963	1.1670	0.1464	0.2735	0.3464	555.4142	512.6100	2.1134
Italy	Mar 11	Mar 17	-0.9537	2.4978	-2.0064	3.0463	4.1810	2.6012	0.5307	0.9845	0.5554	0.3819	0.4918	1046.1239	1934.9495	1.6776
Norway	Mar 24	Mar 22	-3.9815	4.5891	-12.0256	5.5891	4.0587	3.1077	0.3949	1.0343	0.1564	0.3105	0.2447	471.9163	828.4834	2.0465
Spain	Mar 14	Mar 25	-0.7436	0.6951	-1.4311	1.6951	4.0898	3.2785	0.5791	1.1420	0.3861	0.2521	0.3091	263.3634	866.9249	1.6239
Sweden	Mar 18	Mar 09	-4.9651	4.0108	-6.9652	11.2502	4.0167	3.3536	0.3691	1.0414	0.1304	0.7359	0.3024	398.0938	541.5147	2.4724
Switzerland	Mar 20	Mar 19	-1.5556	1.2313	-1.7802	3.1982	3.9322	3.5641	0.6174	1.1477	0.1617	0.2452	0.2787	277.1029	312.9546	2.0510
United Kingdom	Mar 24	Mar 26	-1.4515	1.2782	-2.1466	2.2780	3.9944	3.4811	0.6180	1.1208	0.1959	0.4468	0.2597	288.9485	330.3028	2.0867
Wuhan, China	Jan 23	Feb 02	-1.4380	2.0284	-2.9716	2.2892	3.7473	3.6828	0.6026	1.1391	0.2805	0.1825	0.3511	610.6800	544.0883	2.3823

**Table 2: Parameter estimates for different regions.** See Eq. 1 for model parameters. All estimates are posterior medians. 75% and 95% credible intervals (HDI) are given for  $\tau$  in days relative to  $\hat{\tau}$ .  $\tau^*$  is the official last NPI date (Table 1).

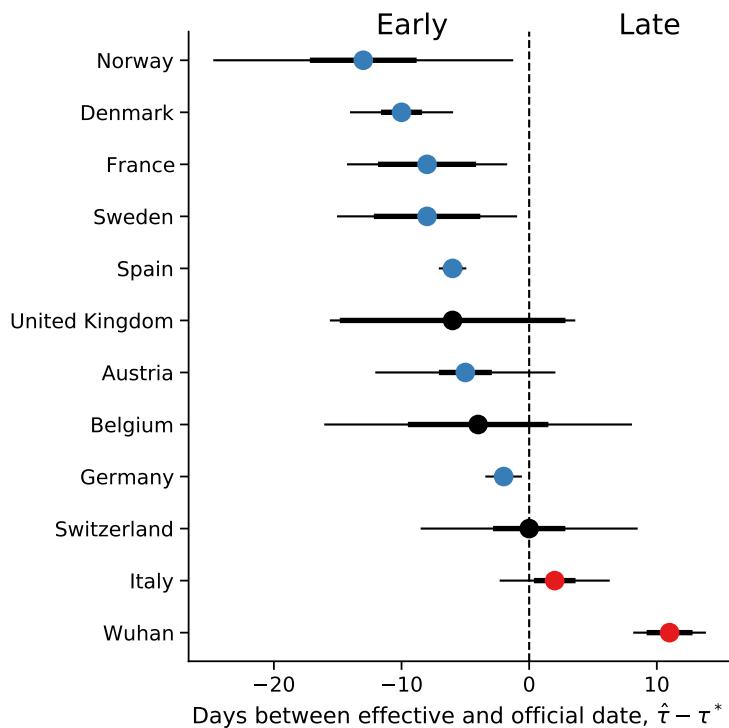
- [1] Ali, S. T., Wang, L., Lau, E. H. Y., Xu, X.-K., Du, Z., Wu, Y., Leung, G. M. and Cowling, B. J. 2020, ‘Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions’, *Science* (80-. ). **9004**(July), eabc9004.
- [2] Arthur, R. F., Jones, J. H., Bonds, M. H. and Feldman, M. W. 2020, ‘Complex dynamics induced by delayed adaptive behavior during outbreaks’, *bioRxiv* pp. 1–23.
- [3] Atchison, C. J., Bowman, L., Vrinten, C., Redd, R., Pristera, P., Eaton, J. W. and Ward, H. 2020, ‘Perceptions and behavioural responses of the general public during the COVID-19 pandemic: A cross-sectional survey of UK Adults’, *medRxiv* p. 2020.04.01.20050039.
- [4] Banholzer, N., Weenen, E. V., Kratzwald, B. and Seeliger, A. 2020, ‘The estimated impact of non-pharmaceutical interventions on documented cases of COVID-19 : A cross-country analysis’, *medRxiv*.
- [5] Bar-On, Y. M., Flamholz, A., Phillips, R. and Milo, R. 2020, ‘SARS-CoV-2 (COVID-19) by the numbers’, *Elife* **9**.
- [6] Dunn, A. G., Leask, J., Zhou, X., Mandl, K. D. and Coiera, E. 2015, ‘Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: An observational study’, *J. Med. Internet Res.* **17**(6), e144.
- [7] Fenichel, E. P., Castillo-Chavez, C., Ceddia, M. G., Chowell, G., Gonzalez Parrae, P. A., Hickling, G. J., Holloway, G., Horan, R., Morin, B., Perrings, C., Springborn, M., Velazquez, L. and Villalobos, C. 2011, ‘Adaptive human behavior in epidemiological models’, *Proc. Natl. Acad. Sci. U. S. A.* **108**(15), 6306–6311.
- [8] Ferretti, L., Wymant, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D. and Fraser, C. 2020, ‘Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing’, *Science* (80-. ). **368**(6491), eabb6936.
- [9] Flaxman, S., Mishra, S., Gandy, A., Unwin, H. J. T., Mellan, T. A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J. W., Monod, M., Ghani, A. C., Donnelly, C. A., Riley, S. M., Vollmer, M. A. C., Ferguson, N. M., Okell, L. C. and Bhatt, S. 2020, ‘Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe’, *Nature* (March), 1–35.
- [10] Foreman-Mackey, D., Hogg, D. W., Lang, D. and Goodman, J. 2013, ‘emcee : The MCMC Hammer’, *Publ. Astron. Soc. Pacific* **125**(925), 306–312.
- [11] Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R. and Rinaldo, A. 2020, ‘Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures’, *Proc. Natl. Acad. Sci. U. S. A.* p. 202004978.
- [12] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. 2013, *Bayesian Data Analysis, Third Edition*, Chapman & Hall/CRC Texts in Statistical Science, Taylor & Francis.
- [13] Goodman, J. and Weare, J. 2010, ‘Ensemble Samplers With Affine Invariance’, *Commun. Appl. Math. Comput. Sci.* **5**(1), 65–80.
- [14] Kass, R. E. and Raftery, A. E. 1995, ‘Bayes Factors’, *J. Am. Stat. Assoc.* **90**(430), 773.
- [15] Kaufman, M. R., Cornish, F., Zimmerman, R. S. and Johnson, B. T. 2014, ‘Health behavior change models for HIV prevention and AIDS care: Practical recommendations for a multi-level approach’, *J. Acquir. Immune Defic. Syndr.* **66**(SUPPL.3), 250–258.

- [16] Kruschke, J. K. 2014, *Doing Bayesian Data Analysis A Tutorial with R, JAGS, and Stan*, Academic Press.
- [17] Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W. and Shaman, J. 2020, ‘Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)’, *Science* (80-. ). p. eabb3221.
- [18] Pei, S. and Shaman, J. 2020, ‘Initial Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US’, *medRxiv* p. 2020.03.21.20040303.
- [19] Smith, L. E., Mottershaw, A. L., Egan, M., Waller, J., Marteau, T. M. and Rubin, G. J. 2020, ‘The impact of believing you have had COVID-19 on behaviour : Cross-sectional survey’, *medRxiv* pp. 1–20.
- [20] Thompson, R. N., Lovell-Read, F. A. and Obolski, U. 2020, ‘Time from Symptom Onset to Hospitalisation of Coronavirus Disease 2019 (COVID-19) Cases: Implications for the Proportion of Transmissions from Infectors with Few Symptoms’, *J. Clin. Med.* **9**(5), 1297.
- [21] Walters, C. E. and Kendal, J. R. 2013, ‘An SIS model for cultural trait transmission with conformity bias’, *Theor. Popul. Biol.* **90**, 56–63.
- [22] Wiyeh, A. B., Cooper, S., Nnaji, C. A. and Wiysonge, C. S. 2018, ‘Vaccine hesitancy ’outbreaks’: using epidemiological modeling of the spread of ideas to understand the effects of vaccine related events on vaccine hesitancy’, *Expert Rev. Vaccines* **17**(12), 1063–1070.
- [23] Zipfel, C. M. and Bansal, S. 2020, ‘Assessing the interactions between COVID-19 and influenza in the United States’, *medRxiv* (February), 1–13.

# Supplementary Material



**Figure S1: Official start of non-pharmaceutical interventions.** See Table 1 for more details. Wuhan, China is not shown.



**Figure S2: Official vs. effective start of non-pharmaceutical interventions estimated up to Mar 28.** The difference between  $\tau$  the effective and  $\tau^*$  the official start of NPIs estimated from case data up to Mar 28, 2020, shown for different regions. Here,  $\hat{\tau}$  is the marginal posterior median.  $\tau^*$  is the last NPI date (a lockdown everywhere by Sweden, see Table 1). Thin and bold lines show 95% and 75% credible intervals (HDI<sup>16</sup>), respectively. Inference performed similarly to main inference, but with data up to Mar 28 and only 1M samples per chain with 600K burn-in.

WAIC

Country	No	Fixed	Free
Austria	219.49	95.06	<b>35.96</b>
Belgium	148.37	98.41	<b>49.16</b>
Denmark	44.36	<b>40.30</b>	43.11
France	581.59	255.14	<b>172.08</b>
Germany	1029.36	327.50	<b>174.90</b>
Italy	898452.34	5484.56	<b>80.18</b>
Norway	70.03	42.04	<b>39.79</b>
Spain	1476.46	647.34	<b>128.58</b>
Sweden	32.53	<b>30.06</b>	31.10
Switzerland	265.80	83.95	<b>63.89</b>
United Kingdom	258.18	117.54	<b>68.17</b>
Wuhan China	107.31	94.00	<b>73.75</b>

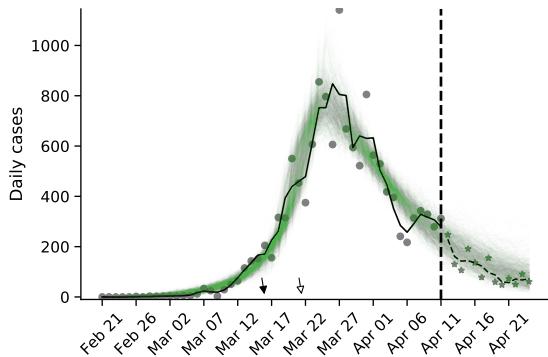
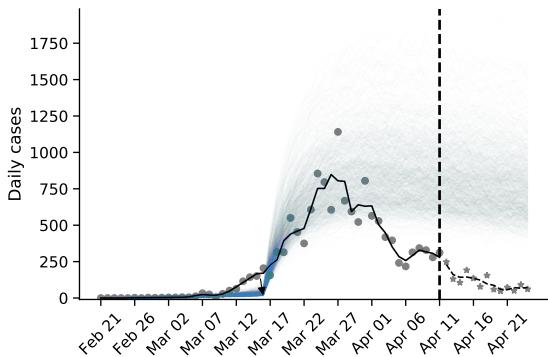
**Table S1: WAIC values for the different models.** WAIC (widely applicable information criterion; Eq. 10)<sup>12</sup> values for models with: no  $\tau$  at all, *No*;  $\tau$  fixed at the official last NPI date  $\tau^*$ , *Fixed*; and free parameter  $\tau$ , *Free*. WAIC values are scaled as a deviance measure: lower values imply higher predictive accuracy and a difference of 2 is a popular threshold for model comparison<sup>14</sup>. Bold values emphasize cases in which the *Free* model has the lowest WAIC.

RMSE

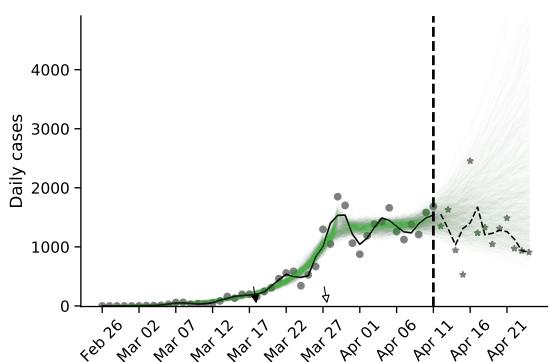
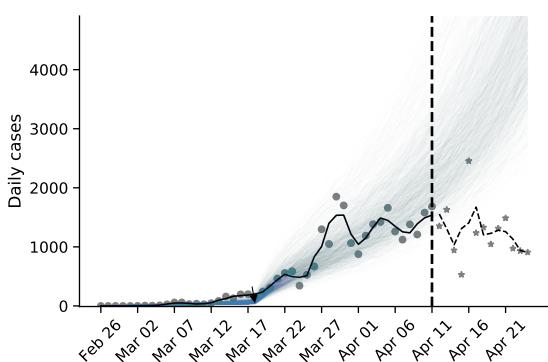
Country	No	Fixed	Free
Austria	926.7	740.7	<b>58.9</b>
Belgium	3286.0	2541.0	<b>902.9</b>
Denmark	434.8	1058.0	<b>400.0</b>
France	10530.0	8206.0	<b>1816.0</b>
Germany	14240.0	17760.0	<b>1952.0</b>
Italy	14160.0	9536.0	<b>1106.0</b>
Norway	292.2	578.2	<b>63.3</b>
Spain	17840.0	14410.0	<b>1704.0</b>
Sweden	775.9	1205.0	<b>596.4</b>
Switzerland	1839.0	1721.0	<b>214.0</b>
United Kingdom	14700.0	14780.0	<b>2735.0</b>

**Table S2: Posterior RMSE of out-of-sample predictions with the different models.** Expected posterior predictive RMSE (root mean squared error) for models with: no  $\tau$  at all, *No*;  $\tau$  fixed at the official last NPI date  $\tau^*$ , *Fixed*; and free parameter  $\tau$ , *Free*. In all cases, the model with free parameter  $\tau$  has the lowest RMSE. Models were fitted to case data up to Apr 11, 2020, and then used to generate 1,000 predictions up to Apr 24 by sampling model parameters from the posterior distribution. These predictions were then compared to the real data using RMSE, and the mean RMSE value is shown in the table for each country and model. Bold values emphasize cases in which the *Free* model has the lowest RMSE.

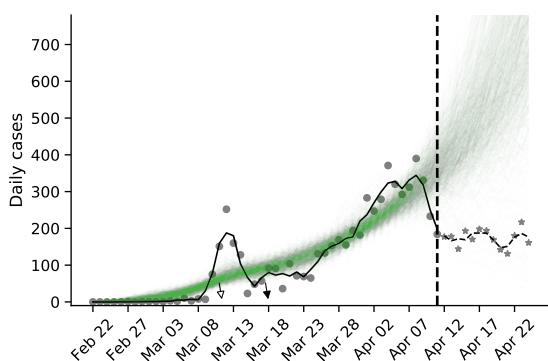
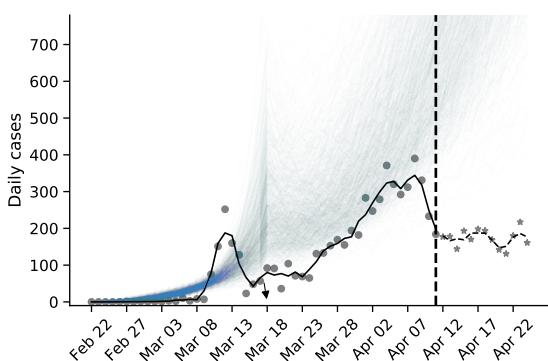
### Austria



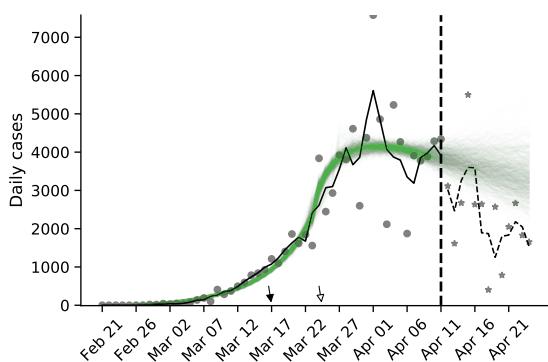
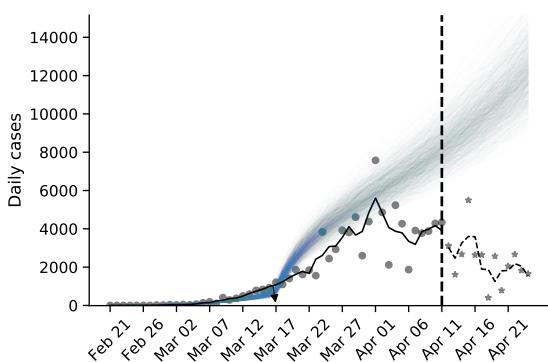
### Belgium



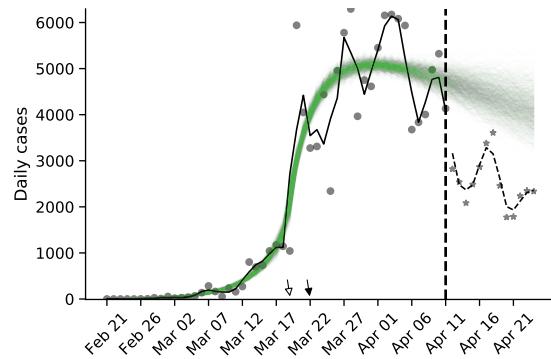
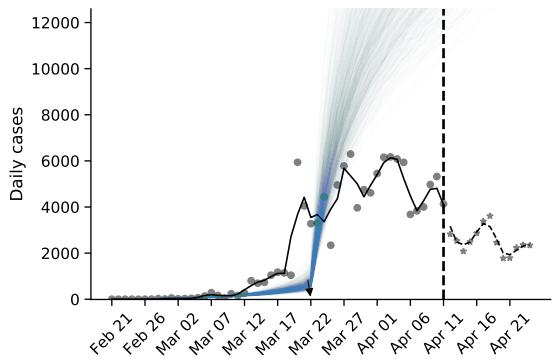
### Denmark



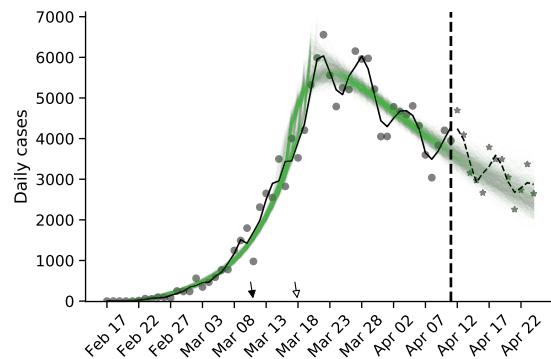
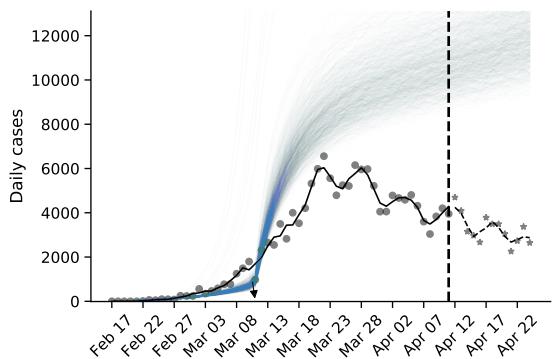
### France



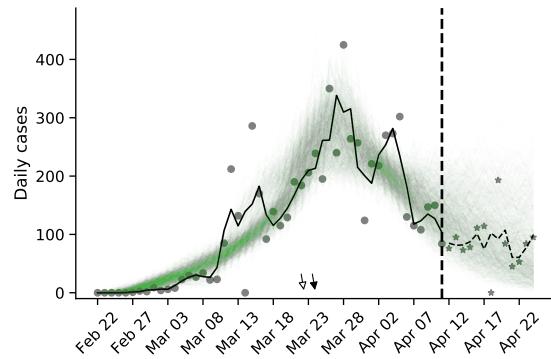
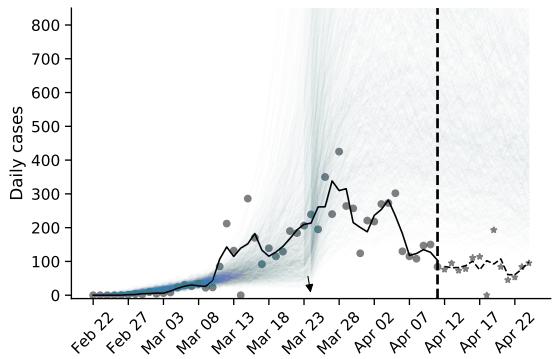
## Germany



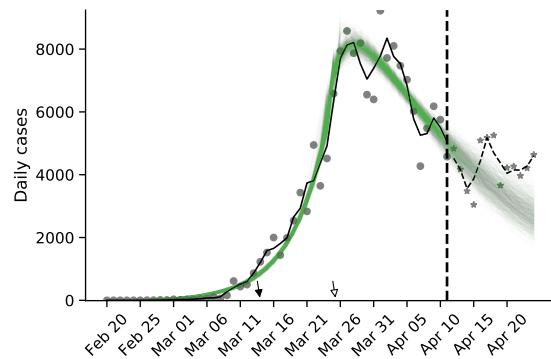
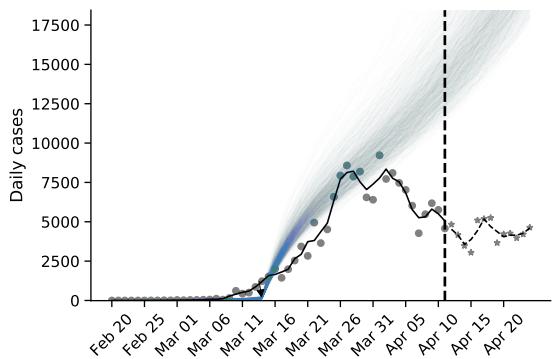
## Italy

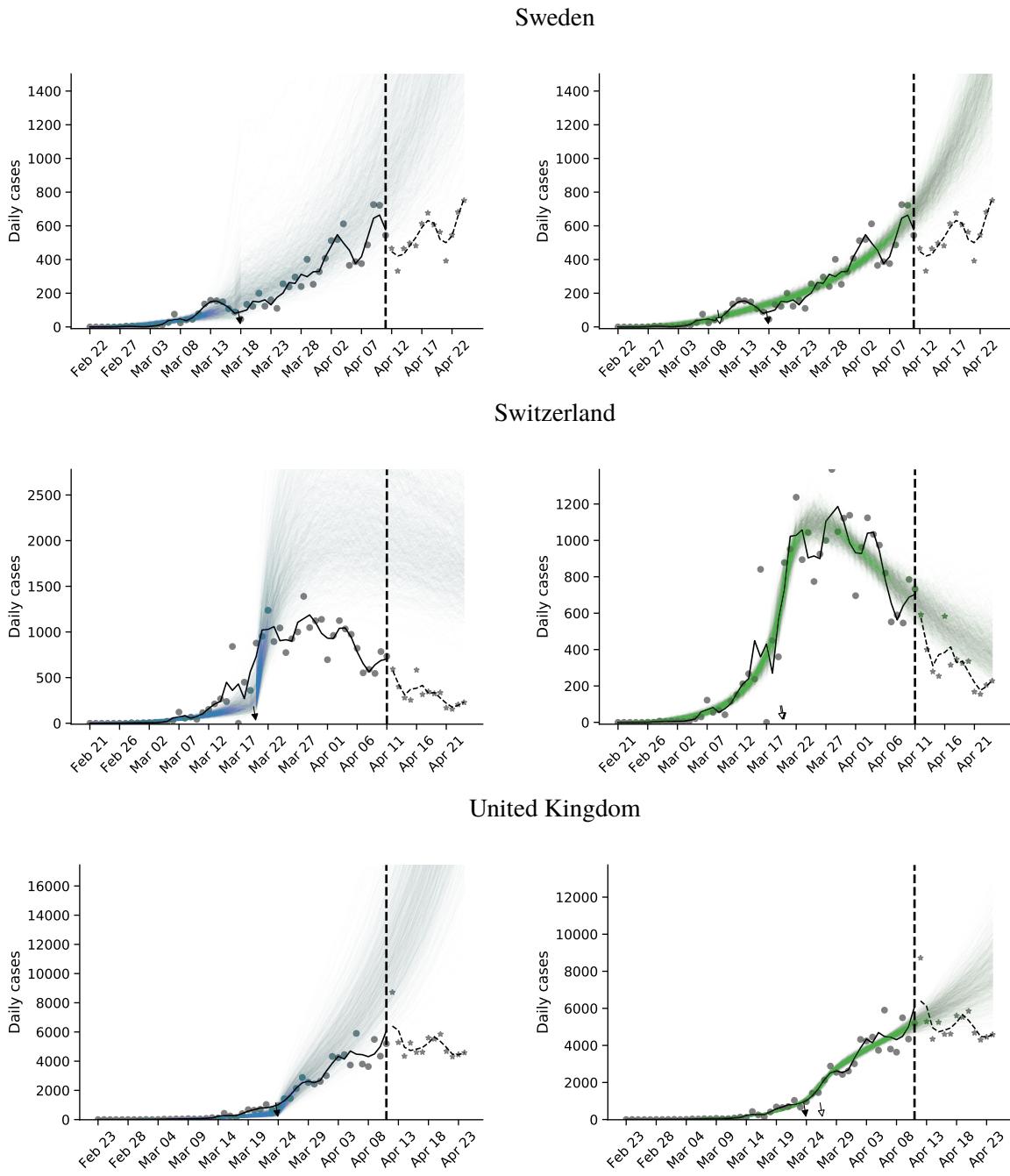


## Norway

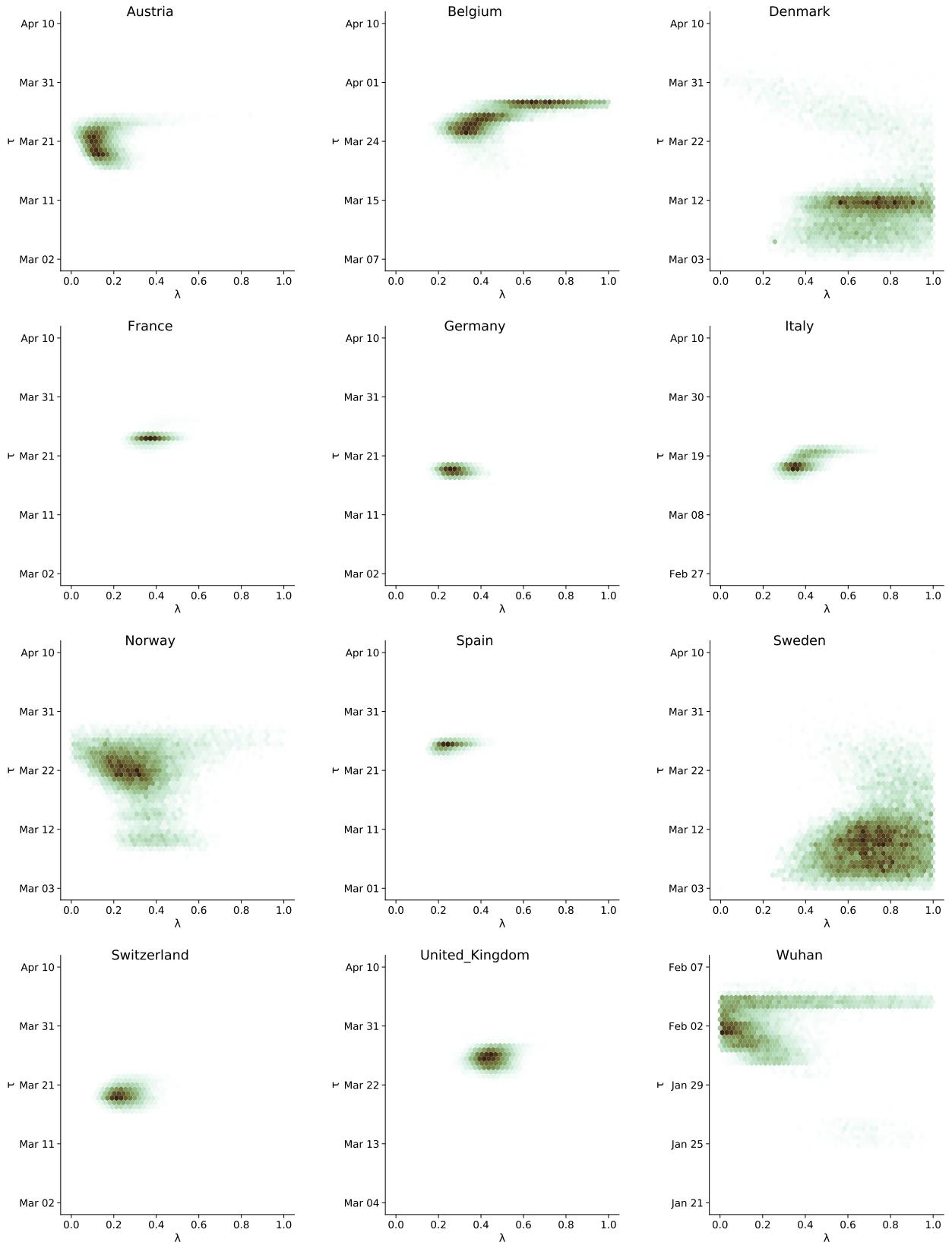


## Spain

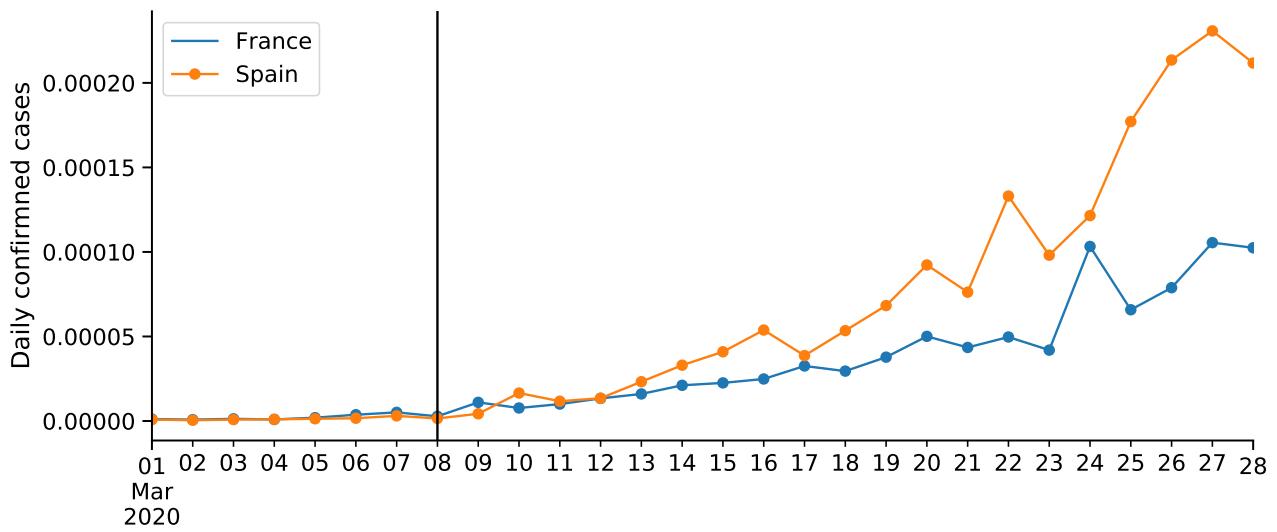




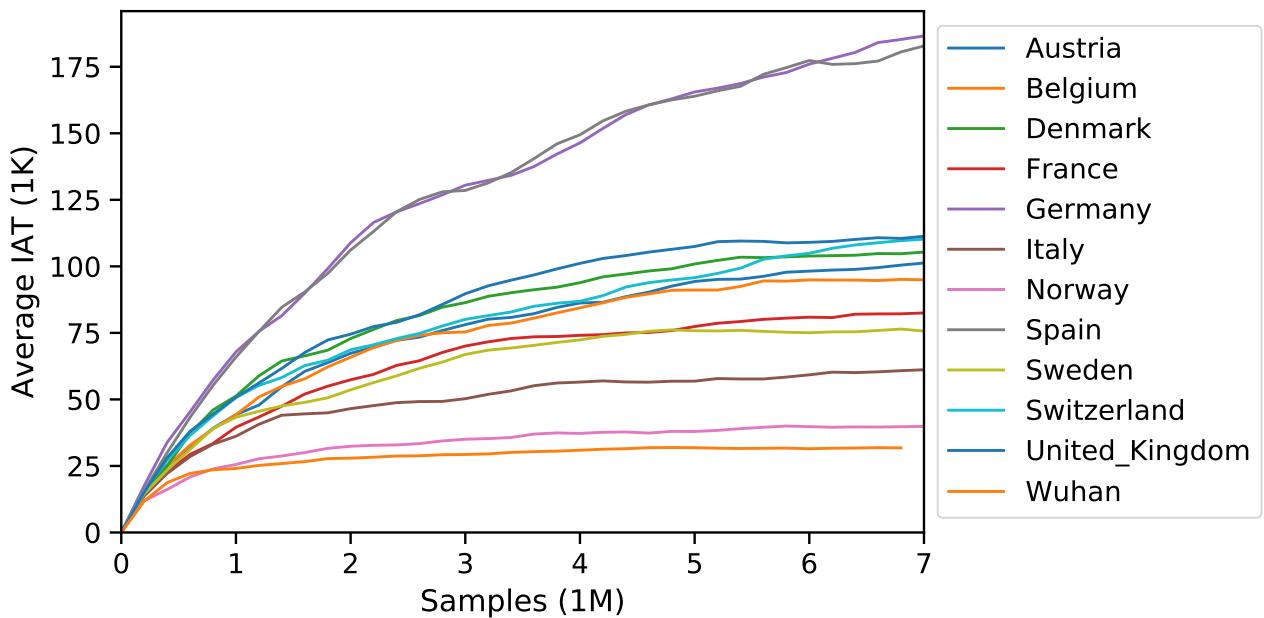
**Figure S3: Posterior prediction plots for 11 European countries.** The vertical dashed line represents Apr 11, 2020. Circles and stars represent daily case data up to and after Apr 11, respectively. Black and white arrows denote the official  $\tau^*$  and effective  $\hat{\tau}$  start of NPIs, respectively. Black lines represent a smoothing of the data points using a Savitzky-Golay filter with window length 3. Coloured lines represent posterior predictions from a model with fixed  $\tau$  (blue) and free  $\tau$  (green). Models were fitted with data up to Apr 11. The predictions are generated by drawing 1,000 parameters sets from the posterior distribution, and then generating a daily case count using the SEIR model (Eq. 1) up to Apr 24. Note the differences in the y-axis scale. Posterior predictions with the free  $\tau$  model predict the out-of-sample data well for all countries except Denmark and Sweden, but poorly with the fixed  $\tau$  model. The predictions of the model without  $\tau$  (not shown) are even worse.



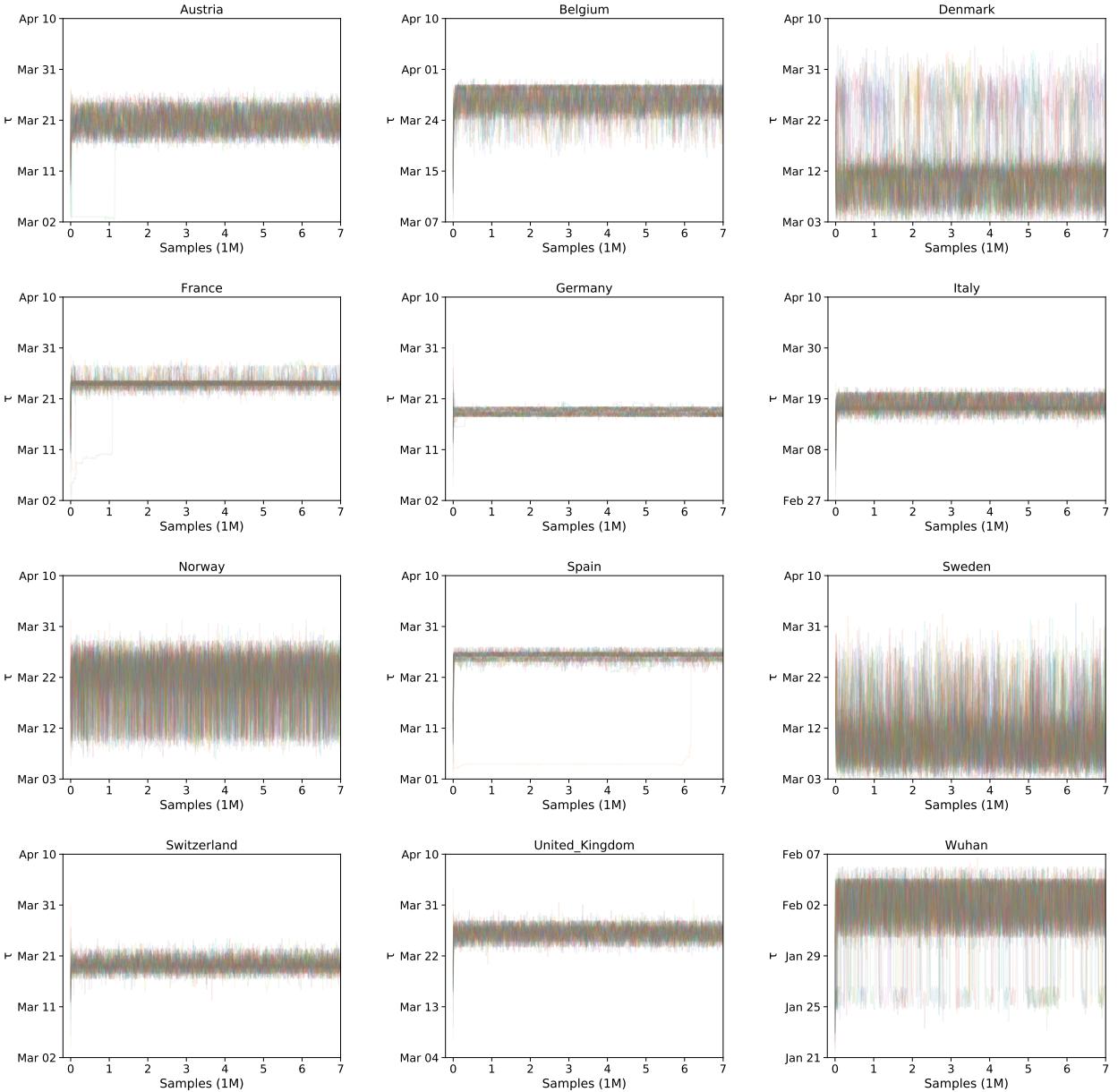
**Figure S4: Joint posterior density plots for  $\tau$  and  $\lambda$ .** The high values of  $\lambda$  estimated in Denmark and Sweden reduce the effect of the NPIs thereby making the inference of  $\tau$  more difficult, resulting in wide posterior distributions. A correlation between the parameters is also evident in Norway. In Belgium and in Wuhan a later  $\tau$  gives a wide estimate for  $\lambda$ . In comparison, Austria, France, Germany, Italy, Spain, Switzerland, and UK have a narrow joint distribution.



**Figure S5: COVID-19 daily confirmed cases in France and Spain.** Number of cases proportional to population size (as of 2018). Vertical line shows Mar 8, the effective start of NPIs  $\hat{\tau}$  in both countries. Data from Flaxman et al.<sup>9</sup>.



**Figure S6: Integrated autocorrelation time (IAT)**<sup>10,13</sup>, averaged across model parameters and MCMC chains, as a function of the number of samples in the chains. IAT is less than 187K in all cases, while chain length is 5M after burn-in period of 2M. With 50 chains per region, this gives at least 1,335 uncorrelated samples for estimating the posterior distribution.



**Figure S7: Trace plots for  $\tau$ .** The value of  $\tau$  at for sequential samples of 50 chains per region. To facilitate visualization, chains were thinned 1:10,000 for the trace plots, but not for posterior estimation. Line transparency was set at  $\alpha = 0.1$  and chains cycle through different colors. In France and Austria some chains converged relatively late but before the burn-in period (2M iterations). In Spain, a single chain converged very late (after roughly 6M iterations) and was removed from further analysis. The y-axis range is the support of the prior on  $\tau$ .