

## A MODELLING FRAMEWORK FOR THE ANALYSIS OF THE TRANSMISSION OF SARS-COV2

BY ANASTASIA CHATZILENA<sup>1</sup>, NIKOLAOS DEMIRIS<sup>2</sup> AND KONSTANTINOS KALOGEROPOULOS<sup>3</sup>

<sup>1</sup>*Department of Economics, Athens University of Economics and Business, [xatzilenan@auer.gr](mailto:xatzilenan@auer.gr)*

<sup>2</sup>*Department of Statistics, Athens University of Economics and Business, [nikos@auer.gr](mailto:nikos@auer.gr)*

<sup>3</sup>*Department of Statistics, London School of Economics and Political Science, [k.kalogeropoulos@lse.ac.uk](mailto:k.kalogeropoulos@lse.ac.uk)*

Action plans against the current SARS-CoV-2 pandemic have been implemented around the globe in order to reduce transmission. The reproduction number has been found to respond to public health interventions changing throughout the pandemic waves. However, the actual global burden of SARS-CoV-2 remains unknown due to severe under-ascertainment of cases. The use of reported deaths has been pointed out as a more reliable source of information, likely less prone to under-reporting. Given that daily deaths occur from past infections weighted by their probability of death, one may infer the true number of infections accounting for their age distribution, using the data on reported deaths. We adopt this framework and assume that the transmission dynamics generating the total number of underlying infections can be described by a continuous time transmission model expressed through a system of non-linear ordinary differential equations, where the transmission rate and consequently the reproduction number are stochastic. We model the transmission rate as a diffusion process, with distinct volatility values for each pandemic wave, allowing to reveal both the effect of control strategies and the changes in individuals behavior. We study the case of 6 European countries and estimate the time-varying reproduction number ( $R_t$ ) as well as the true cumulative number of infected individuals using Stan's No-U-Turn sampler variant of Hamiltonian Monte Carlo using. As we estimate the true number of infections through deaths, we offer a more accurate estimate of  $R_t$ . We also provide an estimate of the daily reporting ratio and discuss the effects of changes in mobility and testing to the inferred quantities.

**1. Introduction.** The current SARS-CoV-2 pandemic which originated in December 2019 in the city of Wuhan, China and spread rapidly across the globe, has had devastating economic and social consequences, in addition to the severe loss of human life. Early on the pandemic, there have been consistent efforts from national health authorities to publicly report the daily counts of laboratory-confirmed cases and deaths in real-time, however, it became more than apparent that surveillance was going to be a challenging part in our quantitative understanding due to the low diagnostic capacity of many asymptomatic and mild infections. In response to the rising numbers of reported cases and deaths, which characterized the pandemic waves in early and late 2020, many European countries, have implemented several control strategies, ranging from social distancing recommendations to large-scale lockdowns. Aiming to reduce a key epidemiological parameter, the time-varying reproduction number, these control strategies have been changing over time and different countries have adopted different action plans, expressing different policy decisions, social mechanisms, health systems capacity and transmission dynamics.

---

*Keywords and phrases:* SARS-COV2, compartmental models, COVID-19, time-varying reproduction number, Hamiltonian Monte Carlo, reporting ratio, Bayesian inference, Stan.

The evaluation of the effectiveness of the implemented preventive measures in each country, as reflected in the reduction of transmission, is not straightforward. Estimation of the time-varying transmission rate and consequently the time-varying reproduction number relies on surveillance data, which are usually biased and incomplete. It is very difficult to evaluate the actual burden of the disease when we deal with such a highly transmissible disease as COVID-19, with many asymptomatic and mild symptomatic infections which are not detected by health systems (Jombart et al., 2020; Li et al., 2020a; Verity et al., 2020). Some large-scale seroprevalence studies (Ward et al., 2021; Pollán et al., 2020) aimed to estimate the actual number of infections and found severe under-ascertainment. Depending on the testing capacities imposed by healthcare resource constraints and the adopted testing and tracing policies, the level of under-ascertainment has been changing over time and across countries.

The number of reported deaths is a more reliable indication of which countries around the globe have faced the most severe effects of the SARS-CoV-2 pandemic and even though, the reporting of deaths may vary over time and across countries, data on reported deaths are likely less prone to under-reporting. Therefore, given that daily deaths occur from past infections weighted by their probability of death, we can infer the total number of infections using the data on reported deaths (Jombart et al., 2020; Flaxman et al., 2020).

This work uses a model-based approach to estimate the transmissibility of SARS-CoV-2 and the effect of the adopted control measures across 6 European countries. Our proposed model is then fit to data from country pairs with similar population demographics and health and social welfare infrastructures, to gain insights on the COVID-19 pandemic.

The paper is organised as follows: In Section 2 we present the developed and adopted methods. More specifically, we provide an overview of the available data and present analytically our model. We consider an extension of the deterministic SEIR (susceptible-exposed-infected-recovered) compartmental model using a stochastic time-varying transmission rate to capture the effect of both the control measures and the behavioural changes (Dureau, Kalogeropoulos and Baguelin, 2013). Section 3 contains the results of our empirical analysis. We implement our suggested model in a Bayesian framework, using Hamiltonian Monte Carlo employing the Stan software, by fitting our model to daily reported deaths for Greece, Portugal, United Kingdom, Germany, Sweden and Norway. We then examine how we can combine our estimates of the total number of daily cases with data on daily laboratory-confirmed cases and estimate the daily reporting ratio. Finally, through a multivariate regression analysis, we disentangle the effects of preventive measures and testing policies on the estimated total cases, the time-varying transmission rate and the reporting rate, using only publicly available data. Section 4 concludes and provides some relevant discussion. All the data, R and Stan code files are available at <https://github.com/anastasiachtz/seir-gbm.git>.

## 2. Methods.

**2.1. The data to date.** Publicly available datasets containing surveillance data on new confirmed cases and deaths per day and per country or region, are maintained in the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) (Dong, Du and Gardner, 2020) based on various sources. To estimate our model parameters describing the mechanisms of disease spread we use only data on the reported number of deaths as a more reliable source of information compared to laboratory-confirmed cases.

The number of laboratory-confirmed daily cases constitutes a biased source of information, primarily due to the high proportion (Lavezzo et al., 2020; Jombart et al., 2020; Li et al., 2020a; Verity et al., 2020) of mild or asymptomatic infections which are not typically reported since people may not seek medical care or testing. This phenomenon intensifies during the

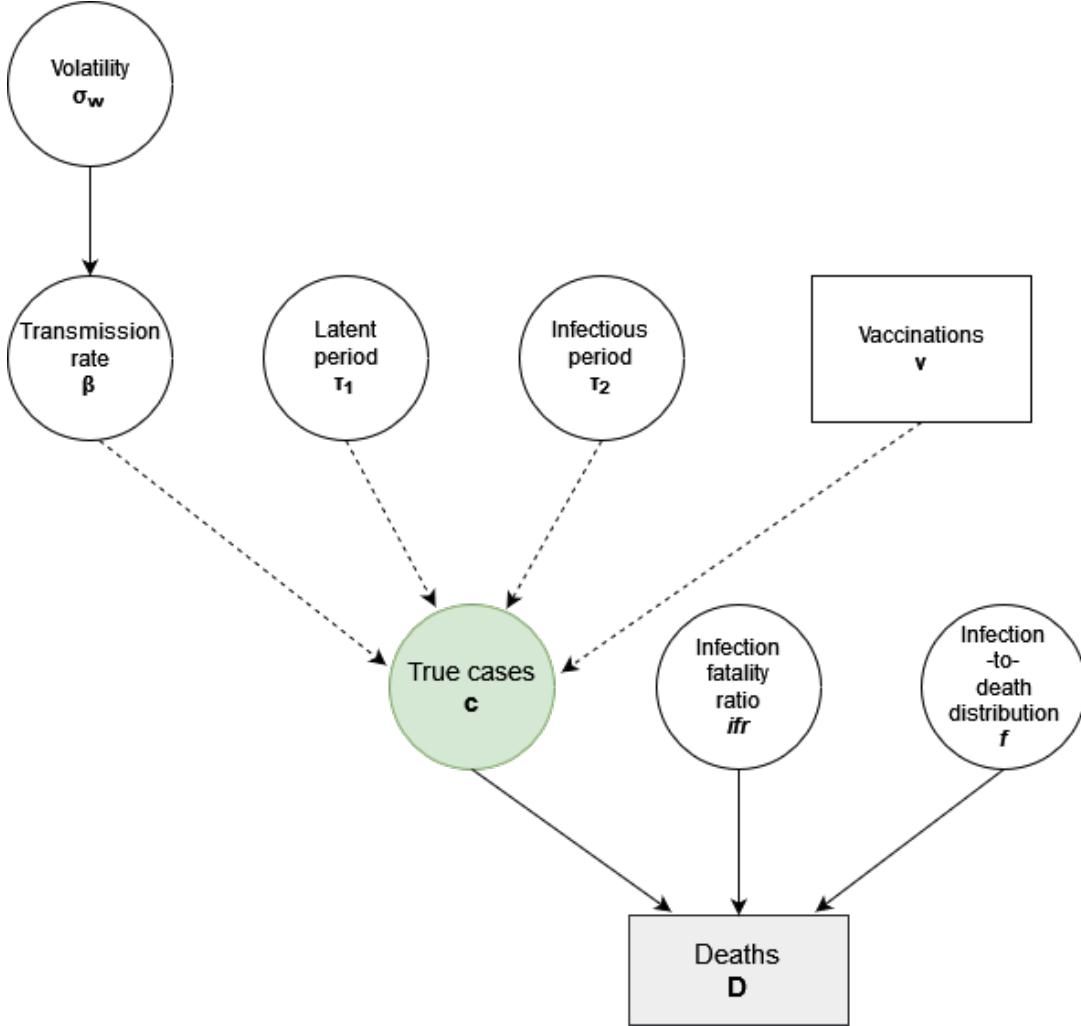
pandemic waves, when the health care systems are overwhelmed, people with mild COVID-19 symptoms are advised to avoid health care unless it is necessary. Taking also into account the testing capacity constraints of each country, which has been changing over time, casts further doubts regarding the quality of the daily cases, as a source of data, towards estimating the actual number of infected people during the pandemic. The latter has been confirmed by some large-scale seroprevalence studies (Ward et al., 2021; Pollán et al., 2020). In this regard, we consider the total number of infections to be unobserved (latent).

Reported deaths are offering a more reliable quantitative understanding of the pandemic despite their own limitations. Flaxman et al. (2020) suggested calculating backwards from observed deaths to the number of infections. The number of deaths attributed to COVID-19 has been considered less prone to under-reporting since deaths mainly arise from severe cases who are more likely to have been tested while seeking health care or after death. Therefore, we assume that the number of unreported deaths in the countries under study is negligible. However, early in the pandemic, in the absence of European and international standards, some deaths due to COVID-19 may have not been recorded. We begin our analysis 2 weeks prior each country reports 10 cumulative deaths following Flaxman et al. (2020), with knowledge of the incomplete nature of these early data since our estimates are not affected by them. We are aware that the timing of the reporting procedures may differ between countries and reporting delays may exist, but we consider that they are relatively minor for the countries under study, however, we incorporate uncertainty within our observational model. In our context, deaths offer a window to the past, revealing the infections which led to them, but they can not be used in real-time analysis to inform current infections unless additional assumptions are imposed.

**2.2. Modelling framework.** Based only on publicly available data sources on the reported number of deaths, we develop a modelling framework where we link the daily new infections to these reported deaths, and infections are generated by a stochastic transmission SEIR compartmental model, adding another level of hierarchy to the model. Also, given the significant vaccine rollout, we extend our model to reflect the impact of vaccinations. Using a directed acyclic graph (Figure 1) we represent the final model’s structure and offer a transparent link between our transmission and observational model. The complexity of such a model is addressed within a Bayesian framework, and statistical inference is carried out using Stan’s implementation of Hamiltonian Monte Carlo (Stan Development Team, 2018). The analysis period ranges from March 2020 to June 2021. Death count data for Greece, Portugal, Germany, United Kingdom and Norway are obtained from CSSE at JHU (Dong, Du and Gardner, 2020), while data for Sweden were directly obtained from Folkhälsomyndigheten, the Public Health Agency of Sweden. Swedish Public Health Agency adjusts the daily number of deaths ex-post, correcting for the reporting delay, resulting in significantly different counts compared to their initial reports. We examine all possible discrepancies between data reported by national health authorities and data maintained in the COVID-19 Data Repository by CSSE at JHU and use the most integrated dataset for each country. Detailed references for each data source are presented in the supplementary material.

**Observation Model.** The base of the model is to link the daily new infections to the data on reported deaths. The infection fatality ratio serves as a bridge between deaths and true infections, in the sense that deaths on any day occur from previously acquired infections according to their probability of death given infection. Following Flaxman et al. (2020), we assume that the expected number of deaths at time  $t$ ,  $d_t$ , is a function of the unobserved true past infections  $c_{t-s}$ , weighted by the distribution of time from infection to death,  $f$  (Verity et al., 2020), and multiplied by their probability of death i.e the infection fatality ratio,

Fig 1: Directed acyclic graph of the model. Square nodes represent observable quantities and circles are latent quantities. Solid arrows represent stochastic dependencies and dashed arrows represent deterministic dependencies.



*ifr* (Levin et al., 2020). Therefore, the expected number of deaths at time  $t$  can be expressed as,

$$(1) \quad d_t = ifr_j * \sum_{\tau=1}^{t-1} c_\tau f_{t-\tau}$$

where  $c_\tau$  are the unobserved true past infections which will be linked to the solution of a system of ODEs (6). The infection-to-death distribution is the sum of two independent Gamma distributions, the estimated infection-to-onset distribution and the estimated onset-to-death distribution (Verity et al., 2020),  $f \sim \Gamma(6.29, 0.26)$  and it is discretized by  $f_1 = \int_0^{1.5} f(\tau)d\tau$  and  $f_s = \int_{s-0.5}^{s+0.5} f(\tau)d\tau$  for  $s = 2, 3, \dots$ .

The overall *ifr* depends on the age distribution of the infections, which changes over time. Until the end of 2020 in our period under study, we identify the points in time where the age distribution of reported infections changes, especially for the 50 – 69 and over 70 age groups which have both the lower under-reporting rates and the higher age-specific *ifr*. We evaluate different overall *ifr* across these different periods using estimates of the age-specific

*ifr* reported by the Centers for Disease Control and Prevention (CDC) based on [Levin et al. \(2020\)](#), which we then adjust taking into account the improvement of health infrastructures after the first few months of the pandemic and the emergence of more lethal and transmissible variants at the end of 2020([Davies et al., 2021; Volz et al., 2021](#)). Our central calculations of the overall *ifr* for each time period  $j$  in the year 2020, of a relatively constant proportion of reported cases per age group, are based on

$$(2) \quad \overline{ifr_j} = \sum_{g=1}^G ifr_g \frac{c_g^{rep}}{C^{rep}}$$

where  $ifr_g$  is the age-specific *ifr* for age group  $g$ ,  $g = 1, \dots, G$ ,  $c_g^{rep}$  is the cumulative number of reported cases of age-group  $g$  and  $C^{rep}$  is the cumulative number of all reported cases. Given our calculations on mean  $ifr_j$ s for every country, we assign Beta priors for each  $ifr_j$  with low variance.

At the end of 2020 and the beginning of 2021, COVID-19 vaccines became available in several countries, including the countries under study. Vulnerable groups at the highest risk of severe disease were prioritized by the designed vaccination programs, therefore older age groups were immunised first. Protection of older age groups and an increase in the number of infections in younger age groups in the presence of more transmissible variants, led to a significant decrease in *ifr*. Based on the rate of decrease of the estimated overall *ifr* in the UK by [Birrell, Blake and van Leeuwen \(2020\)](#), we re-adjust the overall *ifr* in all countries, taking into account the timing of their immunisation programs. Given the existence of asymptomatic infections and the inability of mass testing, the confirmed number of infections will always be smaller than the actual number of infections. We use the latter as a rule of thumb and update the overall *ifr* if the estimated number of infections is larger than the reported number of infections.

The reported deaths  $D_t$  at time  $t$  are assigned a Negative Binomial distribution with mean  $d_t$  and variance  $d_t + \frac{d_t^2}{\phi}$ , i.e.

$$(3) \quad D_t \sim \text{Negative Binomial} \left( d_t, d_t + \frac{d_t^2}{\phi} \right)$$

where  $d_t$  is the expected number of deaths at time  $t$  and  $1/\phi$  controls the overdispersion,  $1/\phi \sim C^+(0, 5)$ .

*Transmission Model.* Several epidemiological models have been proposed ([Birrell et al., 2021; Flaxman et al., 2020; Lemaitre et al., 2020; Dehning et al., 2020; Wood, 2021](#)) attempting to describe the transmission dynamics of SARS-CoV-2 and the effects of preventative measures on these dynamics. Non-pharmaceutical interventions such as social distancing recommendations, limitations on the size of indoor and outdoor gatherings, promotion of teleworking, self-isolation of symptomatic individuals, school closures and ultimately stay-at-home measures, primarily aim to limit the contact rate between individuals while also affecting the relative infectiousness of infected individuals. These control strategies have been changing over time, different action plans have been adopted according to the epidemiological situation of each country and local communities have responded differently to these measures between pandemic waves.

To capture the consequent time-varying nature of the transmission rate we consider a stochastic expansion of the well-known deterministic SEIR compartmental model ([Anderson and May, 1992](#)) assuming a homogeneously mixing population in which all individuals are equally susceptible and equally infectious if they become infected. Instead of a constant transmission rate between susceptible and infected individuals, we assume that it follows a

stochastic process (Dureau, Kalogeropoulos and Baguelin, 2013). Also, given that we extend the period under study to include a significant vaccine rollout to the reviewed countries, we modify our model to include data on vaccinations. The main idea is that vaccinated individuals, once they obtain immunity, they are no longer part of the transmission process. Since there are different types of COVID-19 vaccines available in the countries under study, we assume an average time to obtain immunity irrespective of the type of the distributed vaccines.

Specifically, we consider that the transmission dynamics resulting to the generated true infections in each country, are expressed as the solution of the following system of non-linear ordinary differential equations (ODEs):

$$(4) \quad \begin{aligned} \frac{dS_t}{dt} &= -\beta_t S_t \frac{(I_{1t} + I_{2t})}{N} - \rho \nu_{t-U} \\ \frac{dE_{1t}}{dt} &= \beta_t S_t \frac{(I_{1t} + I_{2t})}{N} - \gamma_1 E_{1t} \\ \frac{dE_{2t}}{dt} &= \gamma_1 E_{1t} - \gamma_1 E_{2t} \\ \frac{dI_{1t}}{dt} &= \gamma_1 E_{2t} - \gamma_2 I_{1t} \\ \frac{dI_{2t}}{dt} &= \gamma_2 I_{1t} - \gamma_2 I_{2t} \\ \frac{dR_t}{dt} &= \gamma_2 I_{2t} + \rho \nu_{t-U} \end{aligned}$$

where  $S_t$  represents the number of susceptible,  $E_t$  the number of exposed, but not yet infectious,  $I_t$  the number of infected and  $R_t$  the number of recovered individuals at time  $t$ . The total population size of each country is denoted by  $N$  (with  $N = S_t + E_t + I_t + R_t$ ). In order to allow the latent and infectious periods to be gamma distributed, we assume that each of the E and I compartments are defined by two classes. Hence,  $\gamma_1$  denotes the rate at which the exposed individuals become infective so that  $\frac{2}{\gamma_1}$  is the mean latent period and  $\gamma_2$  denotes the recovery rate so that  $\frac{2}{\gamma_2}$  is the mean infectious period.

The vaccine efficacy is denoted by  $\rho$ , and  $\nu_{t-U}$  is the reported number of individuals who received the first dose of a vaccine  $U$  days prior to time  $t$ . We consider a simple vaccination scenario, where vaccinated individuals are accounted as removed 45 days after the first dose of any of the available vaccines, which is the average time to obtain immunity given that during this time interval they have also obtained the second dose if it is necessary. We assume that vaccinated individuals do not have reduced transmissibility, thus they remain fully susceptible and they are considered removed only 45 days after the date of their vaccination. While the vaccines substantially weaken the link between cases and deaths, they are not perfect. To account for not perfect vaccine efficacy, we consider that vaccinated individuals move to the removed population proportionally to the vaccines efficacy, which we set equal to 90% as an average efficacy of the different types of the distributed vaccines.

The transmission rate at time  $t$  is denoted by  $\beta_t$ , for which we assume the following stochastic differential equation (SDE)

$$(5) \quad \begin{aligned} d\eta_t &= \mu(\eta_t, \theta_\eta) + \sigma(\eta_t, \theta_\eta) dB_t \\ \eta_t &= g(\beta_t). \end{aligned}$$

The model in (5) may be viewed as the prior for the transmission rate trajectory  $\beta_t$ . The function  $g(\cdot)$  transforms to the real line and is typically set to the logarithm  $\log(\cdot)$ . The drift

function  $\mu(\cdot)$  determines the mean change in  $\beta_t$  and is being set to 0 as we a-priori assume that upward and downward movements are equally likely. The function  $\sigma(\eta_t, \theta_\eta)$  reflects the volatility and  $B_t$  denotes standard Brownian motion. Our starting point is a constant volatility assumption,  $\sigma(\cdot) \equiv \sigma$ , but this is relaxed by introducing specific change-points across different waves to capture potential different responses resulting from adaptive human behavior. Given these specifications, we get the geometric Brownian motion as the prior for the  $\beta_t$  trajectory.

In order to link with the available observations, the model-implied daily new infections, denoted by  $c_t$ , are needed

$$(6) \quad c_t = \int_{t-1}^t \gamma_1 E_{2s} ds.$$

The above integral requires solving the ODE in (4) together with the SDE in (5), in fact it can be viewed as a hypo-elliptic diffusion. As such, it cannot be solved analytically. This implies that is the exact likelihood function for the observed data is not tractable. For this reason, we adopt the data augmentation framework of [Dureau, Kalogeropoulos and Baguelin \(2013\)](#), in the spirit of [Roberts and Stramer \(2001\)](#), that employs time-discretization via the Euler approximation. The fineness of the discretization can be chosen by the user to control the approximation error. For illustration purposes, let us consider  $\beta_t$  to be constant between each each pair of days, i.e. for each  $[t-1, t]$ , noting that smaller intervals can also be used. The model in (5) then implies that  $\eta_t | \eta_{t-1} \sim \mathcal{N}(\eta_{t-1}, \sigma^2)$  and the solution of the ODEs in (4) can be approximated using the trapezoidal rule. More sophisticated Runge-Kutta methods can also be used but the computational cost was very high compared to the trapezoidal rule.

*Prior specification.* To complete the model specification, we consider Gamma prior distributions for the rate of loss of latency and the recovery rate, with small variances, reflecting 2 days average latent period and 4-5 days average infectious period ([Li et al., 2020b](#); [Liu et al., 2020](#)). Due to the fact that the testing capacity of the countries under study has been scaled up especially during 2021, we assume that new cases are isolated more consistently which can be translated in shorter infectious period. Therefore, even though we consider an average infectious period of 5 days during the first year of the pandemic, we adopt a shorter average infectious period of 4 days for the last several months. Finally, a half-Cauchy prior is assigned for the volatility of the Brownian motion,  $\sigma_w \sim C^+(0, 5)$  for each pandemic wave  $w$ .

*Computation.* Inference for ODE-based models represents a non-trivial statistical problem. As the chosen MCMC algorithm explores the parameter space effectively we are solving an increasing number of ODE systems and the ODEs' behaviour varies for different parameter values ([Grinsztajn et al., 2020](#)). Therefore, sophisticated systems of ODEs can be computationally intensive. While fitting our stochastic transmission SEIR model, including vaccinations for 6 European countries, for an extended time period the aforementioned issue became apparent. As the number of observations increased, so did the number of parameters given the time-varying nature of the transmission rate, resulting in relatively small effective sample sizes indicating slow mixing. Inspecting the trace plots and the fit to the observation process per chain, indicated that some chains were stuck at a local mode, thus affected by their initial values.

In contrast, the corresponding SIR model was less sensitive to initial values and computationally less intensive. Therefore we first fit the SIR model and then use the posterior estimates to uniformly draw initial values for the SEIR model. Details on the SIR model's specification and indicative results for Greece can be found in the supplementary material. For the analyses we use 3 chains, each with 1000 iterations of which the first 500 are warm-up to automatically tune the sampler, leading to a total of 1500 posterior samples. We examine the convergence of

the parameters by inspecting the trace plots of all chains and by checking the  $\hat{R}$  convergence statistic and effective sample sizes reported by Stan. In all cases it appears that the chains converged reasonably well with relatively low effective sample sizes only for the volatility parameters of the Brownian motion. Volatilities are top-level parameters in our hierarchical model and we have very little information for them. Other non-informative priors have been also tested for  $\sigma_w$  giving similar results. All diagnostics and analytical results are provided in the supplementary material.

### 3. Results.

**3.1. Estimates of key epidemiological quantities.** In order to access the time course of the pandemic in the 6 European countries under study, estimates on the time-varying reproduction number, the number of daily infections and the cumulative infections per country are reported.

For the stochastic transmission SEIR model, the time-varying reproduction number,  $R_t$ , defined as the average number of secondary cases generated by a typical infectious individual on each day, is denoted by  $2\beta_t/\gamma_2$ . Figures 2, 3, 4 summarize our results on  $R_t$ , for each country pair. Estimates of  $R_t$  at the start of the epidemic must be viewed with caution, with respect to magnitude, since early data on deaths can not accurately reflect the transmission dynamics of local, not sufficiently widespread, infection. For  $R_t$ , if not stated otherwise, we report posterior medians and credible intervals based on the 2.5% and 97.5% quantile of the posterior samples. Analytical results for all the inferred parameters characterizing transmission, can be found at the supplementary material, including the parameters appearing in the observational process. We examine countries in pairs based on their similarities in terms of demographics and health and social welfare infrastructures. Even though our study includes only European countries, the adopted preventive measures may differ significantly between countries and so may the response of the local populations to those measures, as captured by the variation in the transmission rate.

In early March 2020, after the detection of the first SARS-CoV-2 infections, both Greece and Portugal (Fig. 2) introduced sequentially several control measures aiming to prevent large-scale outbreaks. We estimate that early measures taken by both countries, such as cancellation of large public events and closure of educational facilities, managed to drop  $R_t$  significantly, before their nationwide lockdowns. In Greece, lockdown sustained a low spread of SARS-CoV-2 until the beginning of the summer, as reflected in the estimated  $R_t$ , which remained lower than 1, as opposed to Portugal for which we estimate temporary fluctuations of  $R_t$  well above 1 during the same period. Between August and September, some months after the restrictive measures were eased, there was a gradual re-emergence of SARS-CoV-2 transmission in both countries. We estimate that the gradual reintroduction of social distancing measures led to a fall in  $R_t$ , before the imposition of lockdowns, which sustained lower  $R_t$  in the short run. However, in early 2021, in the wake of Christmas and New Year's relaxed measures, as well as the establishment of more transmissible variants, a steady increase in  $R_t$  is estimated in Greece. As a consequence, control measures were even strengthened which resulted in lower transmission levels until June, as reflected in the estimated  $R_t$ . In Portugal on the other hand, even though a state of emergency was declared in early November, the transmission wasn't tamed easily, and in conjunction with relaxed measures during Christmas, we estimate that  $R_t$  reached its highest level since the first wave, at the end of December 2020. Hospitals were pushed to the limit of their capacity during Portugal's third wave which was driven by the highly transmissible alpha variant (Davies et al., 2021; Volz et al., 2021). Eventually, stricter lockdown rules were imposed, managing to slow down  $R_t$ .

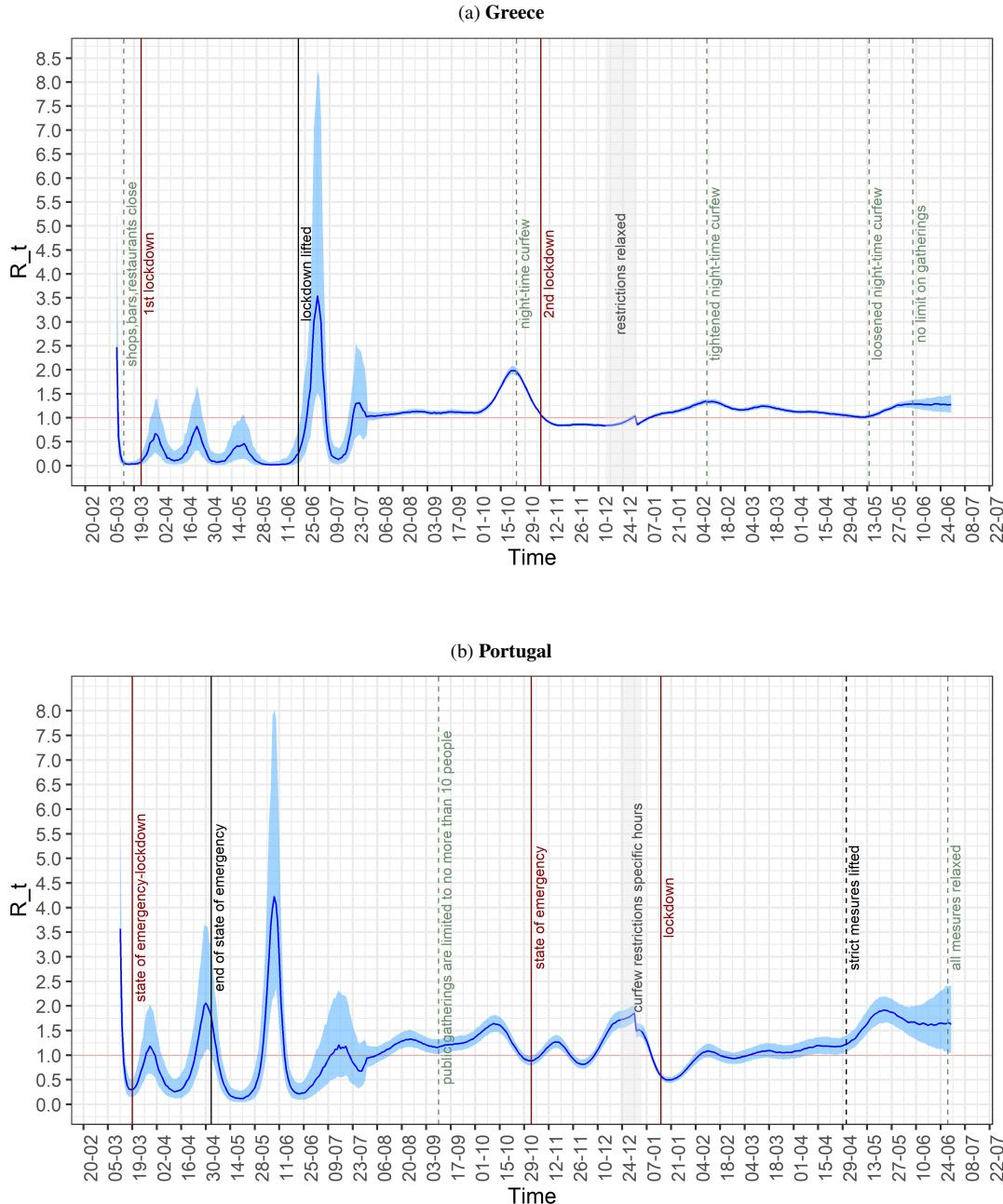


Fig 2: Greece-Portugal - Time-varying reproduction number. Medians(lines) and 50% CI(shaded areas).

Figure 3 illustrates the progression of estimated  $R_t$  in United Kingdom and Germany. In Germany large events were cancelled and schools, as well as non-essential shops, were closed by the middle of March, therefore as we estimate,  $R_t$  dropped significantly even before the partial lockdown, which however sustained  $R_t$  levels well below 1. Elevated testing and tracing policy early in the outbreak, allowed Germany to start lifting restrictions in early May while maintaining low transmission. Regarding the United Kingdom, we estimate that by the middle of March the slow introduction of social distancing measures, before the nationwide lockdown, managed to reduce  $R_t$  below 1. A gradual re-emergence of SARS-CoV-2 transmission after

summer gave rise to  $R_t$ , which by mid-September is estimated to be constantly above 1 in both countries. Control measures and ultimately lockdown tamed transmission in the short-run in both countries, however, we estimate that the significant drop in  $R_t$  was prominent before the implemented lockdowns. Increased transmissibility of the alpha variant as well as relaxed restrictions during Christmas, resulted in a rapid rise in infections both in the United Kingdom and Germany. Both countries implemented stricter lockdowns while the estimated  $R_t$  was decreasing. Despite the stricter measures and a significant vaccine rollout in both countries, we estimate that  $R_t$  remained slightly over 1 up to the end of June.

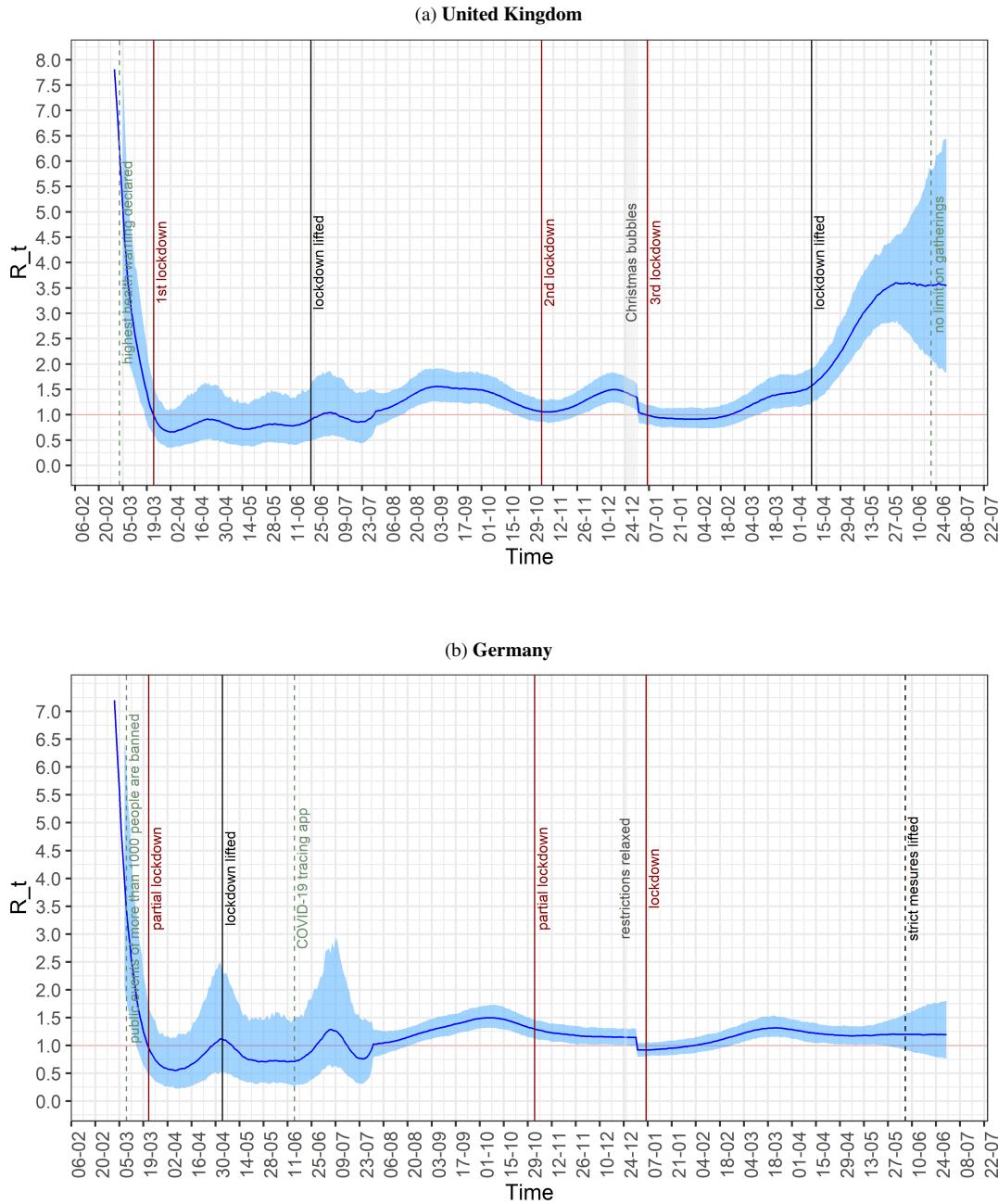


Fig 3: United Kingdom-Germany - Time-varying reproduction number. Medians(lines) and 95% CI(shaded areas).

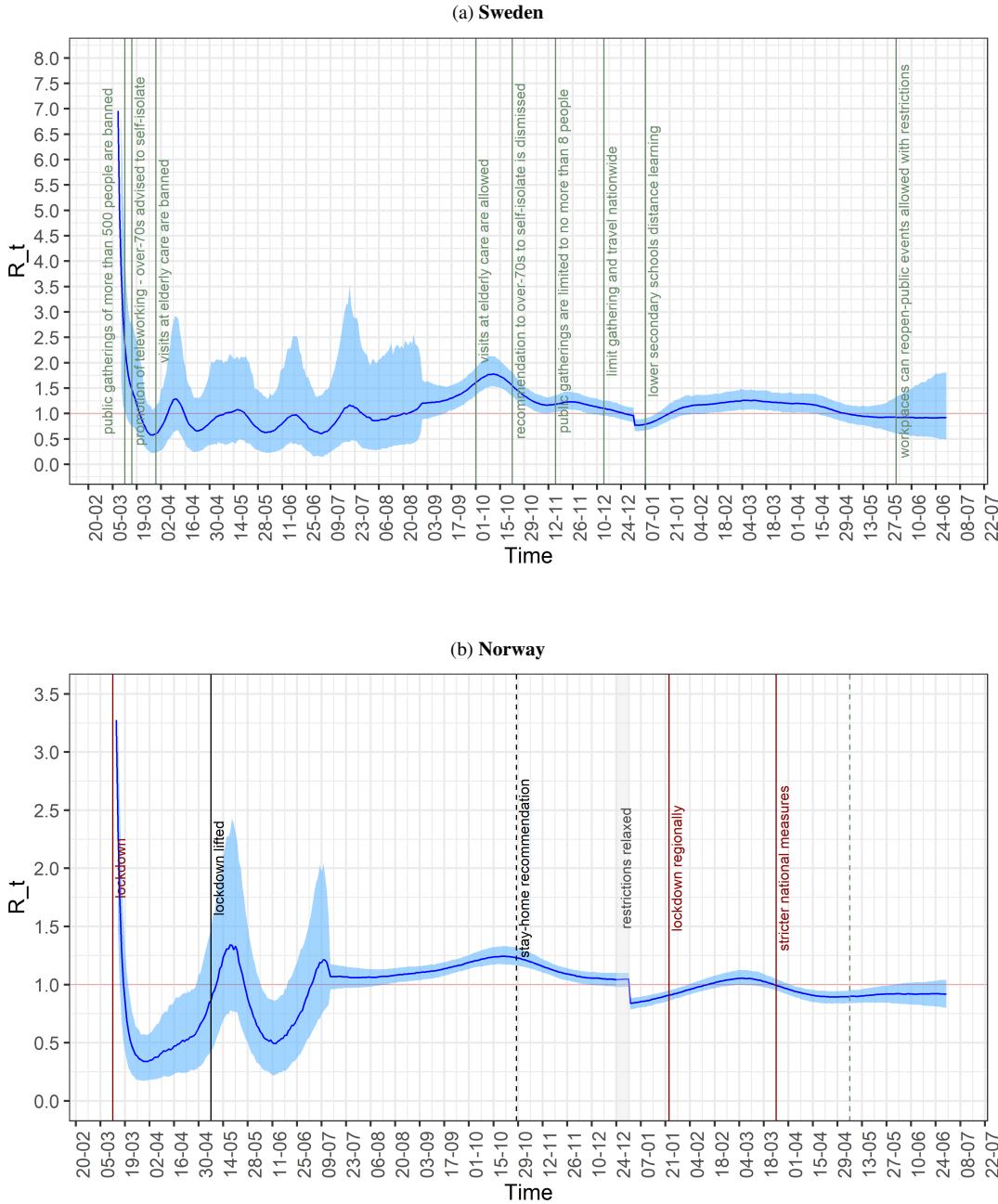


Fig 4: Sweden-Norway - Time-varying reproduction number. Medians(lines) and 95% CI(shaded areas)

While Norway's public health response to COVID-19 was similar to that of many European countries including those under study, Sweden adopted less restrictive measures with no general lockdown. Figure 4 presents our results on the estimated  $R_t$  for Sweden and Norway during the period under study. The day the global pandemic was declared, Norway acted quickly, imposing a lockdown with school closures and rigorous testing. As we estimate, the early measures resulted in a substantial reduction in  $R_t$  well below 1. Contrary to the response of Norway, Sweden sequentially introduced non-pharmaceutical interventions, cancelling large public events and recommending social distancing measures especially for more vulnerable

groups, however, all businesses, as well as schools, continued to operate. We estimate that  $R_t$  managed to drop below 1, although at a smaller pace compared to Norway, but Sweden faced excess transmission in elderly care homes leading to many deaths. A resurgence of infections during the second pandemic wave, led Sweden to introduce stricter control strategies, similar to those of other countries. We estimate that  $R_t$  started increasing in September in both countries, however, the estimated  $R_t$  in Sweden peaked at a much higher level in October, compared to Norway. Re-introduction of control measures dropped the estimated  $R_t$  in the short-run for both countries, yet during the first months of 2021, increased transmissibility characterizing the third pandemic wave, led to an increase in  $R_t$ . Sweden faced a significantly more severe situation with an estimated  $R_t$  much higher than Norway.

The number of new daily infections is one of the main epidemiological quantities offering straightforward insight into the dynamics of the pandemic. In the supplementary material, the estimated daily new cases are presented along with the reported cases for each country, as well as the estimated aggregate infections from March 2020 to June 2021. Our findings indicate that in all countries, during the first and second pandemic wave, the estimated daily new cases are significantly higher than the laboratory-confirmed cases. In Figure 5, the estimated number of cumulative cases in the United Kingdom is presented, along with the equivalent estimate from REACT-2 ([Ward et al., 2021](#)). Seroprevalence surveys such as REACT constitute a direct approach to estimate the actual number of individuals that have been infected but have not been detected by surveillance systems. According to REACT-2, by mid-July the overall antibody prevalence in England was 6% (95% CI: 5.8 – 6.1). If we adjust the estimated overall prevalence to the population in the United Kingdom, our estimates on the total population infected coincide with the estimates from REACT-2, validating our findings.

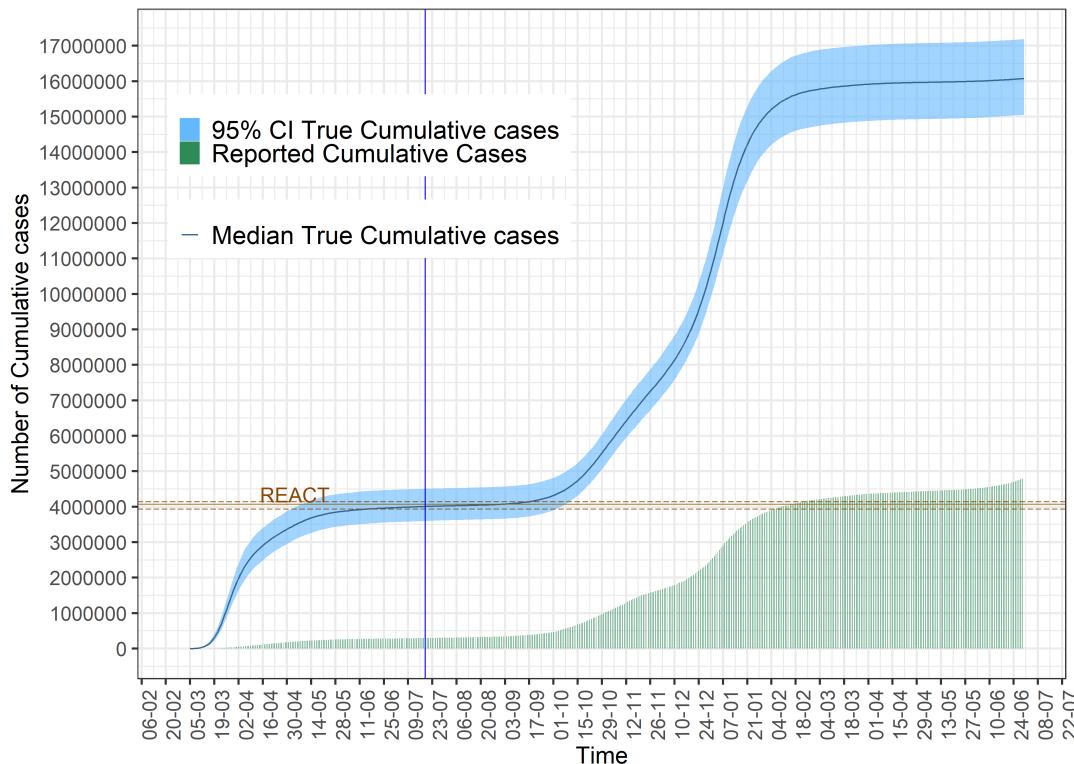


Fig 5: **United Kingdom** - Total population infected, data(bars), median(line) and 95% CI(shaded area).

**3.2. Estimates of additional epidemiological quantities.** Given our estimates on the number of true daily cases, if we combine them with data on daily laboratory-confirmed cases, we get an estimate of the number of unreported cases each day and consequently an estimate of the daily reporting ratio. Thus, using the posterior median of the estimated total number of infections at time  $t$ , denoted by  $c_t$ , we explicitly incorporate a reporting delay between infection and report,  $L$ , so the number of unreported cases can be described as  $c_t^{unrep} = c_{t-L} - c_t^{rep}$ , where  $c_t^{rep}$  is data on the number of laboratory-confirmed cases which are reported at time  $t$  (see supplementary material). We consider a time delay between infection and report ( $L$ ) equal to 6 days (Abbott et al., 2020). Similarly, we can define the daily reporting ratio as the ratio of laboratory-confirmed cases to the estimated total number of cases adjusted to their time of report i.e.

$$(7) \quad r_t = c_t^{rep} / c_{t-L}$$

Especially during the first pandemic wave, when reporting protocols had not been established, there are several days where the reported number of cases display spikes that do not represent an actual increase in cases in this particular day but inconsistencies in reporting. Given that we want to capture the general direction of the varying reporting ratio, we implement a generalized additive model smoothing to remove those resulting peaks (Wood, Pya and Säfken, 2016). We used the mgcv-package (Wood and Wood, 2015) and specify a spline based smooth with respect to time. For each country, results on the smoothed reporting ratio are presented in Figure 6.

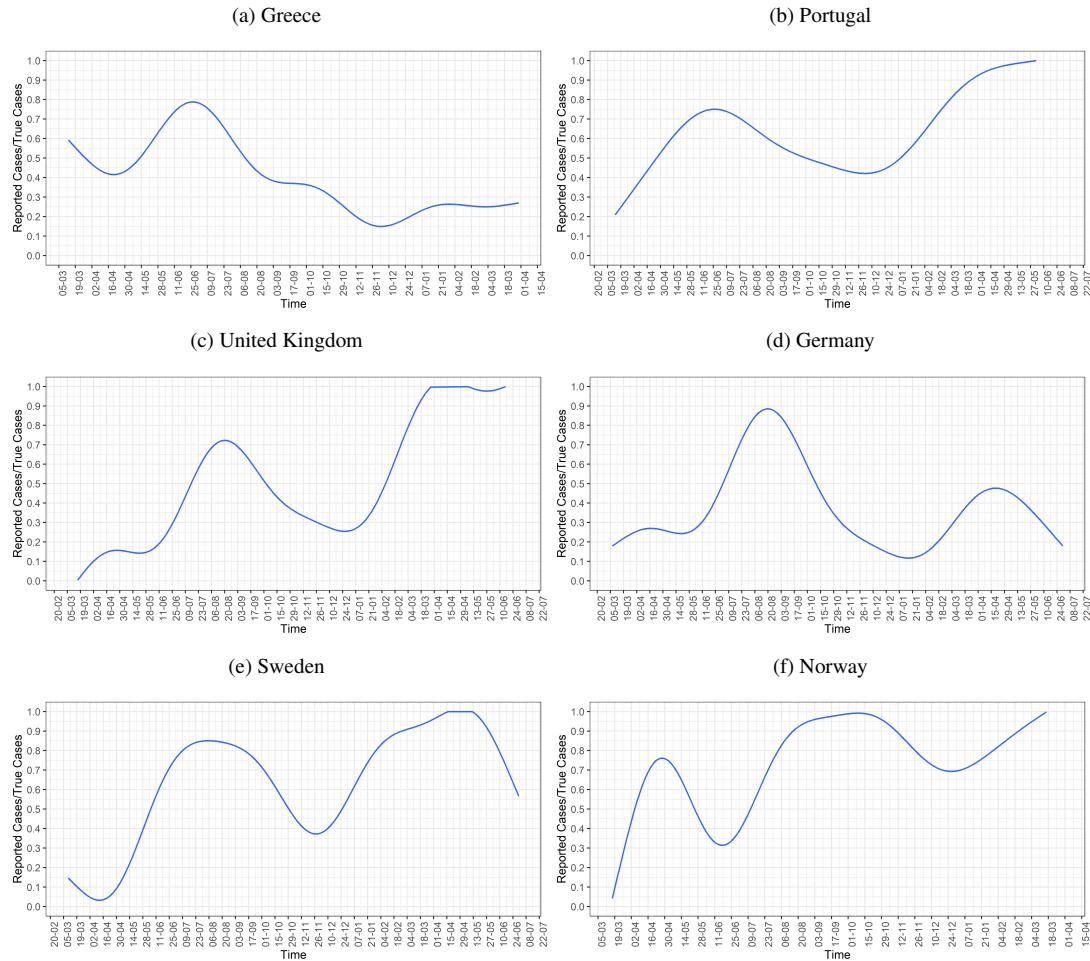


Fig 6: Smoothed daily reporting ratio.

Our findings indicate that especially during periods of high transmission there is a large proportion of under-reported infections in all countries. Even though most countries improved their testing coverage during the second and third wave, the predominance of more transmissible variants led to an increased number of infections, a large proportion of which was not detected by health systems, resulting in an estimated reporting ratio lower than 40%. Mismatches between our estimates on cases and the reported data for Sweden and Norway, calls the respective reporting ratios into question and further refinement.

**3.3. Multivariate Regression Analysis.** As noted in subsection 3.1, the estimated  $R_t$  presents various changes over time, which seem consistent with the expected changes as a result of the implemented control measures for COVID-19 in the countries under study. Essentially, the time-varying transmission rate  $\beta_t$  captures transmission dynamics as shaped by the evolution of susceptible and infected individuals which in turn is affected by several time-varying factors such as non-pharmaceutical interventions, variations of human behaviour based on sociodemographic characteristics and climatic variations, which given the nature of SARS-CoV-2 transmission encourage infectious contacts, among others. These effects are highly interdependent making it difficult to disentangle their individual contribution to  $\beta_t$  and design targeted and more efficient public health control strategies for each country. However, using publicly available data, we can investigate the link of generic measures of the effects of control strategies, to key epidemiological quantities. Therefore, in what follows, given our estimates on the time-varying transmission rate, the daily number of true cases and the daily reporting ratio, we examine their relationship to two basic measures; mobility patterns and testing policies. Given the significant impact of vaccinations on transmission dynamics, we have chosen to implement our analysis until March 2021, when the effect of non-pharmaceutical interventions was still dominant.

Social distancing measures such as stay-at-home recommendations, limitations on gatherings, closure of schools and workplaces and general restrictions on internal movement can be reflected in changes in mobility patterns (Kraemer et al., 2020; Lemaitre et al., 2020). These kinds of measures aim to limit the contact rate between individuals and therefore reduce transmission. Mobility data by Google (Google, 2021) are collected by geographical location and summarize relative changes in movement in different categories of places, such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential. We access Google mobility data for Greece, Portugal, Sweden, Norway, United Kingdom and Germany for the study period and use principal component analysis on movement trends across different categories of places, and take only the first component to form a single measure of mobility (James et al., 2013).

As to testing and tracing policies, a crude measure of the efforts to increase testing capacity in each country, is the reported number of daily tests. The objective of scaling up testing policies is to tame transmission by the early isolation of confirmed or suspected cases. However, there are many different technologies for COVID-19 testing and reporting procedures differ between countries and across time. Publicly available datasets containing data on the number of tests per day and per country, are maintained in the data portal Our World in Data (Hannah Ritchie and Roser, 2020) based on various sources and contain either both PCR and antigen tests or only PCR tests. We access these data for Greece, Portugal, Sweden, Norway and United Kingdom while for Germany we use data reported by Robert Koch Institute, Federal Ministry of Health (see supplementary material). We are aware that these data are not an accurate representation of each country's testing and tracing policy, however, we consider that they account for a significant fraction of the actual tests performed and they can still reflect possible variations in the testing policy.

Considering the definitions of the estimated time-varying transmission rate, true cases and reporting ratio, we naturally assume that they have a cross-equation error correlation.

Consequently, in order to examine their relationship to mobility patterns and testing policies, we implement a multivariate regression analysis. We use as covariates the sum of lagged mobility trends weighted by the time they are able to generate infections and the daily number of tests performed 3 to 6 days ago. So we run the following multivariate regression for each country,

$$(8) \quad \mathbf{Y}_t \sim \text{MVN} \left( \delta_1 m_t + \delta_2 \text{tests}_{t-3} + \delta_3 \text{tests}_{t-4} + \delta_4 \text{tests}_{t-5} + \delta_5 \text{tests}_{t-6}, \boldsymbol{\Sigma} \right)$$

where  $\mathbf{Y}_t = (c_t, \log(\beta_t), \text{logit}(r_t))$ . The mobility proxy  $m_t$  is described as  $\sum_{\tau=1}^{t-1} \text{mob}_\tau \pi_{t-\tau}$ , where  $\text{mob}$  is the first principal component of movement trends and  $\pi \sim \Gamma(2.6, 0.4)$  is the serial interval which is discretized by  $\pi_1 = \int_0^{1.5} \pi(\tau) d\tau$  and  $\pi_s = \int_{s-0.5}^{s+0.5} \pi(\tau) d\tau$  for  $s = 2, 3, \dots$ . The number of tests at day  $t$  is denoted by  $\text{tests}_t$  and the regression coefficients by  $\delta_i$ ,  $i = 1, \dots, 5$ .

The covariance matrix is represented by  $\boldsymbol{\Sigma}$  and we can rewrite it in terms of the correlation matrix  $\Omega$  as,  $\boldsymbol{\Sigma} = D_\sigma \Omega D_\sigma$ , where  $D_\sigma$  denotes a diagonal matrix with diagonal elements  $\sigma_i$ ,  $i = 1, 2, 3$ . Then, we specify a LKJ onion method correlation matrix distribution ([Lewandowski, Kurowicka and Joe, 2009](#)) for  $\Omega$ . The parameterization of LKJ distribution used in Stan allows to sample matrices depending on a shape parameter. The shape parameter determines whether we want to sample matrices closely to the identity matrix or more or less uniformly over positive definite matrices ([Stan Development Team, 2018](#)). Generally, the higher the value of the shape parameter, the more we consider that the correlations are close to zero, while for values less than 1, we consider higher probabilities for non-zero correlations, high in absolute value. Stan provides a more efficient, implicit parameterization of the LKJ correlation matrix density in terms of its Cholesky factor, which we use since it is also more stable. Given that our dependent variables are correlated by definition, we assume that  $\Omega \sim \text{LkjCholesky}(0.5)$  which is translated to a U-shaped prior over random correlation matrices and assign half-Normal prior on the standard deviations,  $\sigma_i \sim N^+(0, 10)$ .

For the vector of regression coefficients  $\boldsymbol{\delta}$  we use a Zellner's g-prior, which is a multivariate normal distribution with covariance matrix proportional to the inverse Fisher information matrix ([Zellner, 1986](#)), i.e.

$$(9) \quad \boldsymbol{\delta} \sim \text{MVN}(0, g\sigma_\delta^2 (X'X)^{-1})$$

where  $g$  reflects the amount of information available in the data relative to the prior, enabling an automatic scaling based on the data. A common approach which we adopt is to consider  $g = n$  which is equivalent to the prior having the same amount of information as as one observation ([Perrakis and Ntzoufras, 2018](#)).

We use the No-U-Turn sampler variant of HMC as performed in the Stan software to obtain posterior estimates of the regression coefficients for each country (see supplemental material). Our results suggest that mobility has a significant positive effect on the transmission rate for Greece, Portugal, United Kingdom and Germany, indicating that increased mobility is associated with increased transmission. The latter we would expect to hold also for the total infections, however, our results indicate a significant positive effect of mobility on total estimated cases, only in Portugal. The expected relation between the reporting ratio and mobility is not straightforward in the sense that, although increased mobility is expected to increase infections, if at the same time the number of tests performed has increased, then the effect on the reporting ratio is not clear. In Portugal, United Kingdom, Germany, Sweden and Norway our estimates reflect a positive effect of mobility on the reporting ratio while there is a negative effect in Greece. As concerns the impact of tests on transmission and infections, we would expect increased testing capacity to decrease the number of infections, as well as the transmission rate, given that confirmed infections are detected earlier and isolated.

Our findings are in line with the latter, indicating a significant negative effect of the lagged number of tests on transmission rate for Portugal, Germany, Sweden and Norway. However, an increase in the tests performed may be the result of increased transmission in the community. A positive statistically significant relation between the estimated daily cases and the number of tests performed during the previous days, is observed in most of the countries. Finally, the reporting ratio is positively associated with test numbers in the United Kingdom, Sweden and Norway. In any case, our results for Sweden and Norway should be cautiously interpreted, given the discrepancies in our estimates of the daily infections.

In general, we can not make valid conclusions concerning optimal universal strategies, based on the effect of the selected proxy measures on the inferred quantities characterising transmission.

**4. Discussion.** In this paper, we present a Bayesian approach for the estimation of temporal changes in the reproduction number of SARS-CoV-2 through data on deaths. Using a flexible stochastic extension of the SEIR model, we examine the COVID-19 pandemic in 6 European countries, inferring key epidemiological quantities such as the case reproduction number and the daily number of cases. COVID-19 outbreak control measures primarily aim to affect the transmission rate according to which the evolution of susceptible and infected individuals is determined. We assume that the generation of infections is described by an extension of the deterministic SEIR compartmental model where the transmission rate is stochastic. A proportion of these infections results to the deaths that we observe, according to a certain probability.

We estimate that during the time course of the pandemic there have been many more infections than what was detected by health care systems. Especially during the peak of each pandemic wave, the actual number of infections is significantly larger. The estimated cumulative cases can offer a measure of the actual burden of the pandemic. We show that the estimated changes in the reproduction number are consistent with the expected variation in SARS-CoV-2 transmission over time, as a result of the implemented control strategies. We estimate that all countries except Sweden, having introduced several non-pharmaceutical interventions, were able to drop  $R_t$  below 1 well short of their nationwide lockdowns. The effects of sequentially introduced interventions in a small period of time, are highly interdependent making it difficult to disentangle their individual contribution. Therefore, we can't make valid conclusions concerning optimal strategies.

The basic advantage of our modelling approach is the fact that we don't make strong structural assumptions for the transmission rate, and subsequently for the case reproduction number, concerning the way they change in time. Changes in  $R_t$  are only driven by variations in the observed data on deaths for each country. We consider that control measures, different public responses on these measures based on cultural characteristics, adaptive human behaviour during a pandemic and any other time-varying factor is reflected in the trends in the numbers of deaths resulting from the respective infections. Therefore, our framework can be easily adapted to other countries, without making further developments. On the contrary, [Flaxman et al. \(2020\)](#) model  $R_t$  as a piecewise constant function that changes only when an intervention occurs, while also assuming that each intervention has the same effect on  $R_t$  across countries and over time. However, in reality, interventions can have a different impact on individuals' behaviour in each country and individuals change their response even to the same measures between subsequent outbreaks.

Several limitations need to be considered when using death counts as the main source of information. Early in the pandemic, in the absence of European and international standards, some deaths due to COVID-19 may have not be recorded, leading to underestimation of infections, therefore our initial estimates must be viewed with caution. Also reporting procedures

may differ significantly between countries both in terms of the timing of the report as well as the definition of the COVID-19 related death. We are aware of possible discrepancies in the data, therefore we try to access the most integrated data source, comparing the data reported by each national health authority to the data maintained by CSSE at JHU. In addition, data on deaths can only offer a window to the past, so they are not suitable for real-time analysis without making further assumptions. However, in the absence of large seroprevalence studies in many countries, death counts are the only way to evaluate the actual burden of the pandemic, in terms of people infected.

The initial objective of this work was to provide a flexible framework, offering an accurate representation of what has happened in the pandemic so far, without making predictions for the future. Extending our analysis for 16 months increased the computational cost and introduced several time-dependent factors which should be taken into account. Our findings rely on estimates of the *ifr* which have large uncertainty especially after partial immunity is induced through vaccination. We allow only for deterministic changes in *ifr* at specific time points based on our empirical observations on the emergence of more lethal and transmissible variants, the introduction of vaccines and the improvement or extreme pressure on health infrastructures. However, seroprevalence studies can give us a more accurate insight into the time-varying *ifr*. Unfortunately, we don't have many such surveys available for all countries. Finally, during the time course of the pandemic, factors such as the limiting capacity of hospitals during periods of high transmission, the age distribution of fatal infections and the efficacy of the available vaccines, affect the infection to death distribution, which we assume constant throughout our analysis. In this case, detailed data on hospitalizations can improve our estimates.

**Acknowledgments.** Anastasia Chatzilena states that: "Part of this research is co-financed by Greece and the European Union (European Social Fund- ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the project "Strengthening Human Resources Research Potential via Doctorate Research" (MIS-5000432), implemented by the State Scholarships Foundation (IKY)"

## REFERENCES

- ABBOTT, S., HELLEWELL, J., THOMPSON, R., SHERRATT, K., GIBBS, H., BOSSE, N., MUNDAY, J., MEAKIN, S., DOUGHTY, E., CHUN, J., CHAN, Y., FINGER, F., CAMPBELL, P., ENDO, A., PEARSON, C., GIMMA, A., RUSSELL, T., NULL, N., FLASCHE, S., KUCHARSKI, A., EGGO, R. and FUNK, S. (2020). Estimating the time-varying reproduction number of SARS-CoV-2 using national and subnational case counts [version 2; peer review: 1 approved with reservations]. *Wellcome Open Research* **5**.
- ANDERSON, R. M. and MAY, R. M. (1992). *Infectious diseases of humans: dynamics and control*. Oxford university press.
- BIRRELL, P., BLAKE, J. and VAN LEEUWEN, E. (2020). COVID-19: nowcast and forecast.
- BIRRELL, P., BLAKE, J., VAN LEEUWEN, E., GENT, N. and DE ANGELIS, D. (2021). Real-time nowcasting and forecasting of COVID-19 dynamics in England: the first wave. *Philosophical Transactions of the Royal Society B* **376** 20200279.
- DAVIES, N. G., ABBOTT, S., BARNARD, R. C., JARVIS, C. I., KUCHARSKI, A. J., MUNDAY, J. D., PEARSON, C. A., RUSSELL, T. W., TULLY, D. C., WASHBURNE, A. D. et al. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. *Science* **372**.
- DEHNING, J., ZIERENBERG, J., SPITZNER, F. P., WIBRAL, M., NETO, J. P., WILCZEK, M. and PRIESEMANN, V. (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* **369**.
- DONG, E., DU, H. and GARDNER, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* **20** 533–534.
- DUREAU, J., KALOGEROPOULOS, K. and BAGUELIN, M. (2013). Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics* **14** 541–555.

- FLAXMAN, S., MISHRA, S., GANDY, A., UNWIN, H. J. T., MELLAN, T. A., COUPLAND, H., WHITTAKER, C., ZHU, H., BERAH, T., EATON, J. W. et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584** 257–261.
- GOOGLE (2021). COVID-19 Community Mobility Reports. <https://www.google.com/covid19/mobility/>.
- GRINSZTAJN, L., SEMENOVA, E., MARGOSSIAN, C. C. and RIOU, J. (2020). Bayesian workflow for disease transmission modeling in Stan. *arXiv preprint arXiv:2006.02985*.
- HALE, T., ANGRIST, N., GOLDSZMIDT, R., KIRA, B., PETHERICK, A., PHILLIPS, T., WEBSTER, S., CAMERON-BLAKE, E., HALLAS, L., MAJUMDAR, S. et al. (2021). A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour* **5** 529–538.
- HANNAH RITCHIE, D. B. E. M. J. H. B. M. C. G. C. A. L. R.-G. ESTEBAN ORTIZ-OSPIÑA and ROSER, M. (2020). Coronavirus Pandemic (COVID-19). *Our World in Data*. <https://ourworldindata.org/coronavirus>.
- JAMES, G., WITTEN, D., HASTIE, T. and TIBSHIRANI, R. (2013). *An introduction to statistical learning* **112**. Springer.
- JOMBART, T., VAN ZANDVOORT, K., RUSSELL, T. W., JARVIS, C. I., GIMMA, A., ABBOTT, S., CLIFFORD, S., FUNK, S., GIBBS, H., LIU, Y. et al. (2020). Inferring the number of COVID-19 cases from recently reported deaths. *Wellcome Open Research* **5**.
- KRAEMER, M. U., YANG, C.-H., GUTIERREZ, B., WU, C.-H., KLEIN, B., PIGOTT, D. M., GROUP†, O. C.-D. W., DU PLESSIS, L., FARIA, N. R., LI, R. et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368** 493–497.
- LAVEZZO, E., FRANCHIN, E., CIAVARELLA, C., CUOMO-DANNENBURG, G., BARZON, L., DEL VECCHIO, C., ROSSI, L., MANGANELLI, R., LOREGIAN, A., NAVARIN, N. et al. (2020). Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* **584** 425–429.
- LEMAITRE, J. C., PEREZ-SAEZ, J., AZMAN, A., RINALDO, A. and FELLAY, J. (2020). Assessing the impact of non-pharmaceutical interventions on SARS-CoV-2 transmission in Switzerland. *medRxiv*.
- LEVIN, A. T., HANAGE, W. P., OWUSU-BOAITEY, N., COCHRAN, K. B., WALSH, S. P. and MEYEROWITZ-KATZ, G. (2020). Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *European journal of epidemiology* 1–16.
- LEWANDOWSKI, D., KUROWICKA, D. and JOE, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis* **100** 1989–2001.
- LI, R., PEI, S., CHEN, B., SONG, Y., ZHANG, T., YANG, W. and SHAMAN, J. (2020a). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368** 489–493.
- LI, Q., GUAN, X., WU, P., WANG, X., ZHOU, L., TONG, Y., REN, R., LEUNG, K. S., LAU, E. H., WONG, J. Y. et al. (2020b). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England journal of medicine*.
- LIU, T., HU, J., KANG, M., LIN, L., ZHONG, H., XIAO, J., HE, G., SONG, T., HUANG, Q., RONG, Z. et al. (2020). Transmission dynamics of 2019 novel coronavirus (2019-nCoV).
- PERRAKIS, K. and NTZOUFRAS, I. (2018). Bayesian variable selection using the hyper-g prior in WinBUGS. *Wiley Interdisciplinary Reviews: Computational Statistics* **10** e1442.
- POLLÁN, M., PÉREZ-GÓMEZ, B., PASTOR-BARRIUSO, R., OTEO, J., HERNÁN, M. A., PÉREZ-OLMEDA, M., SANMARTÍN, J. L., FERNÁNDEZ-GARCÍA, A., CRUZ, I., DE LARREA, N. F. et al. (2020). Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a nationwide, population-based seroepidemiological study. *The Lancet* **396** 535–544.
- ROBERTS, G. O. and STRAMER, O. (2001). On Inference for Partially Observed Nonlinear Diffusion Models Using the Metropolis-Hastings Algorithm. *Biometrika* **88** 603–621.
- STAN DEVELOPMENT TEAM (2018). Stan Modeling Language User's Guide and Reference Manual, Version 2.18.0. <http://mc-stan.org/>.
- VERITY, R., OKELL, L. C., DORIGATTI, I., WINSKILL, P., WHITTAKER, C., IMAI, N., CUOMO-DANNENBURG, G., THOMPSON, H., WALKER, P. G., FU, H. et al. (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet infectious diseases* **20** 669–677.
- VOLZ, E., MISHRA, S., CHAND, M., BARRETT, J. C., JOHNSON, R., GEIDELBERG, L., HINSLEY, W. R., LAYDON, D. J., DABRERA, G., O'TOOLE, Á. et al. (2021). Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593** 266–269.
- WARD, H., ATCHISON, C., WHITAKER, M., AINSLIE, K. E., ELLIOTT, J., OKELL, L., REDD, R., ASHBY, D., DONNELLY, C. A., BARCLAY, W. et al. (2021). SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. *Nature communications* **12** 1–8.
- WOOD, S. N. (2021). Inferring UK COVID-19 fatal infection trajectories from daily mortality data: Were infections already in decline before the UK lockdowns? *Biometrics*.
- WOOD, S. N., PYA, N. and SÄFKEN, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* **111** 1548–1563.

- WOOD, S. and WOOD, M. S. (2015). Package ‘mgcv’. *R package version 1 29*.
- ZELLNER, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques*.

## SUPPLEMENTARY MATERIAL

### Addition Tables and Figures

Supplemental tables, figures and diagnostics for the implemented SEIR model, as well as indicative figures for the SEIR model without vaccinations and the analogous SIR model with vaccinations. Analytical results on the posterior estimates of the multivariate regression coefficients.

### **SEIR - Estimated number of daily cases, Estimated number of cumulative cases, Goodness of fit to the data, Estimated number of daily unreported cases**

Figures 7a-12b offer a thorough graphic presentation of the time course of estimated infections in parallel with some general control measures implemented in each country, for the time period between March and mid-October 2020 and between mid-October 2020 and June 2021. For ease of presentation, uncertainty in our results on new daily infections is expressed through 50% credible intervals (CI) derived from the 25% and 75% quantiles. Our initial estimates of the infections are uncertain, due to possible under-ascertainment in deaths, and likewise, when it comes to the estimated infections of the latest period under study, uncertainty exists due to the fact that deaths can only offer an accurate representation of infections during the past 2 – 3 weeks. The estimated aggregate infections in each country from March 2020 to June 2021 are presented in Figure 13 offering an indication of the actual burden of the disease. Finally, Figure 14 presents a visual inspection of the model fit to the data on death counts for each country and Figure 15 sets out additional model's estimates on the unreported daily number of cases for each country.

In all countries, during the first and second pandemic wave, the estimated daily new cases are significantly higher than the laboratory-confirmed cases. Also, both posterior medians and credible intervals are skewed to the left in comparison to the reported cases, which can be explained by possible reporting delays and by the fact that transmission also occurs through asymptomatic individuals ([Lavezzo et al., 2020](#)).

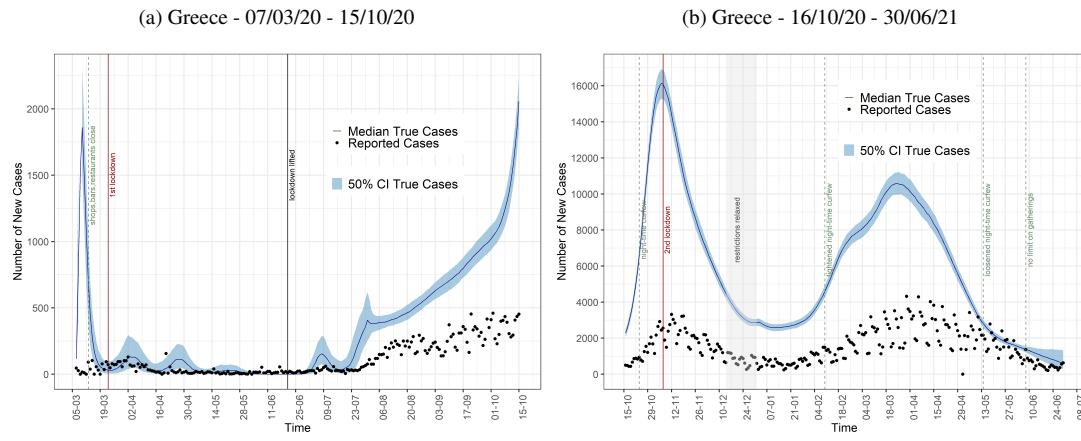


Fig 7: Daily estimated number of total cases, medians(line) and 50% CI(shaded areas) along with reported cases data(dots).

We estimate that during the first pandemic wave, Greece and Portugal having introduced several control measures well short of lockdown, were able to suppress infections early (Fig.

7a and 8a). With regard, however, to the resurgence of infections in fall 2020, stricter measures were necessary to stem the increase of infections. The notable difference between estimated and reported cases in Portugal in January of 2021 (Fig. 8b), indicates a possible change in the infection to death distribution, which we considered constant throughout our analysis. A shorter infection to death distribution is consistent with the extreme pressure in hospitals at that time in Portugal.

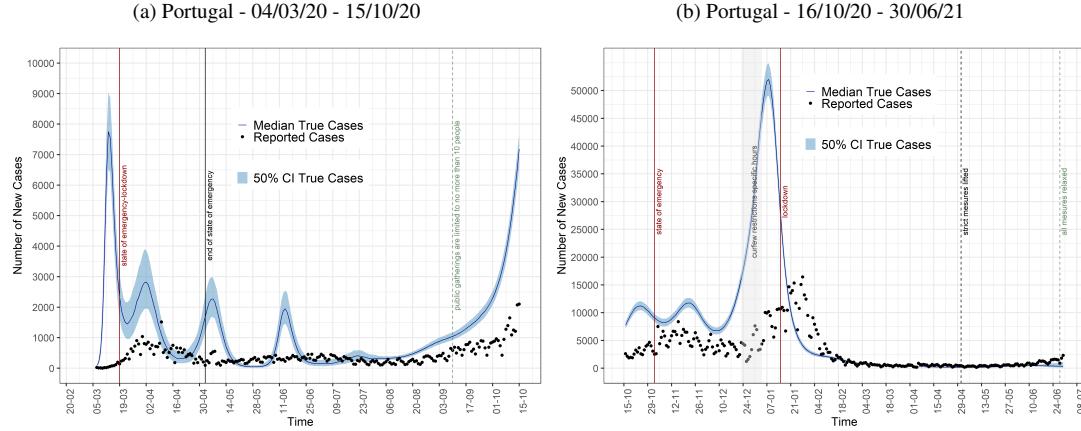


Fig 8: Daily estimated number of total cases, medians(line) and 50% CI(shaded areas) along with reported cases data(dots).

Regarding the estimated infections in Germany and the United Kingdom, during the first wave, a decrease in daily new cases appeared some days before the nationwide lockdown in both countries (Fig. 9a) and 10a). However, during the subsequent pandemic waves infections are estimated to increase steadily although both countries had imposed nationwide lockdowns in November. The dominance of the alpha variant led to extended lockdowns which eventually reduced the number of infections as estimated.

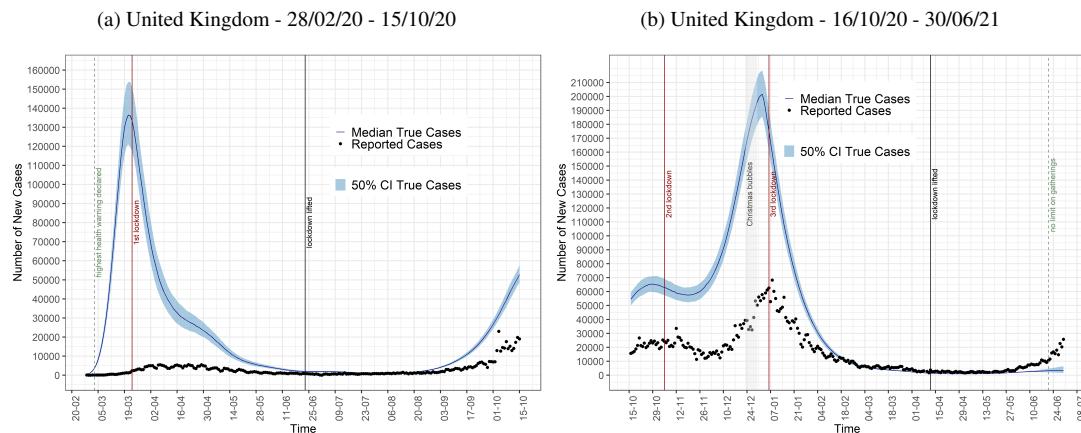


Fig 9: Daily estimated number of total cases, medians(line) and 50% CI(shaded areas) along with reported cases data(dots).

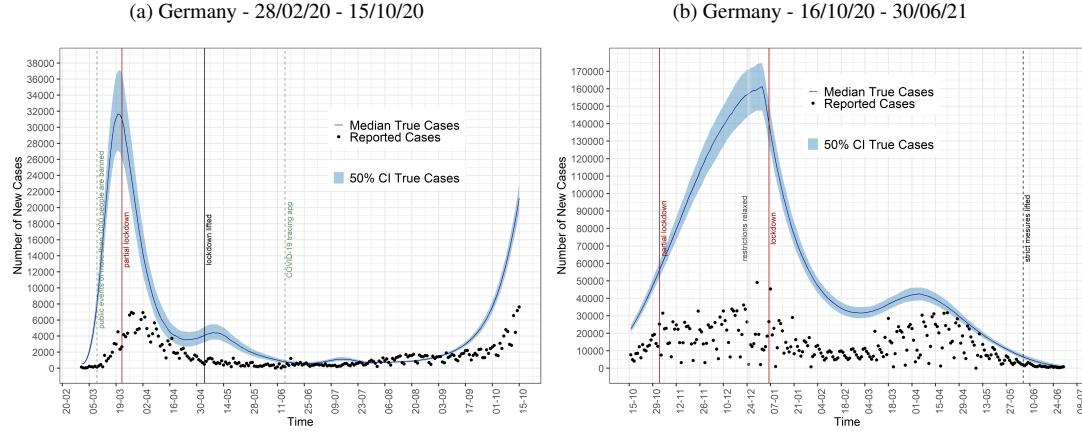


Fig 10: Daily estimated number of total cases, medians(line) and 50% CI(shaded areas) along with reported cases data(dots).

In Sweden, we estimate that a large proportion of total infections were undocumented during the first wave (Fig. 11a). Also, a substantial amount of these infections appear much earlier in the pandemic than what is reported. Poor monitoring procedures and uncontrolled outbreaks in care homes in Sweden can partly explain this significant mismatch, since older infected individuals may have a smaller infection to death distribution. The assessment of the subsequent waves, especially after January 2021 requires further investigation. The apparent inconsistency between estimated and reported cases may be the result of discrepancies in the data as well as several time-dependent factors not accounted for in our model. Estimates on the true cases in Norway must also be viewed with caution, given the sparsity of the reported data on deaths (Fig. 14f). Days with zero reported deaths were followed by days with high death counts particularly after September 2020. Norway is a case in point, indicating the strong dependence of our model on the quality of reported data. Regarding our estimates on the total infected population, Norway's increased testing capacity compared to the other countries under study, can explain the small difference between the estimated and reported number of individuals who have been infected, as presented in Figure 13f.

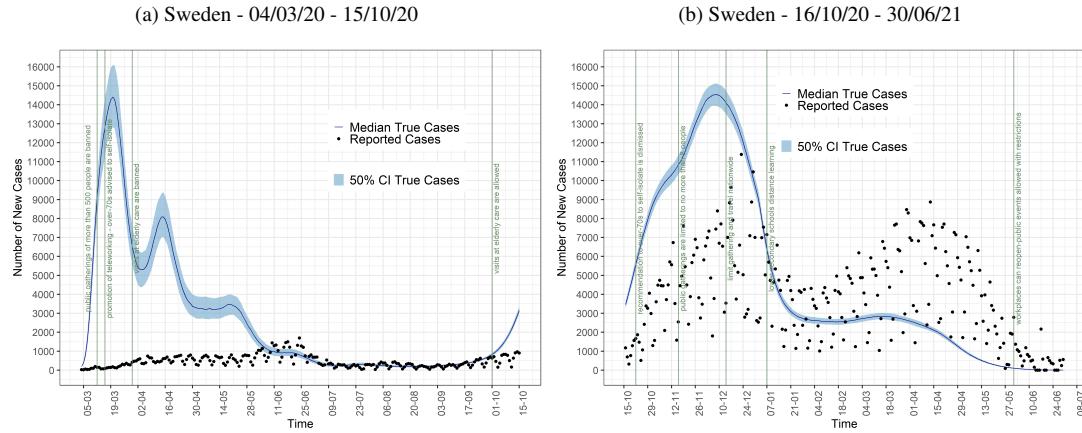


Fig 11: Daily estimated number of total cases, medians(line) and 50% CI(shaded areas) along with reported cases data(dots).

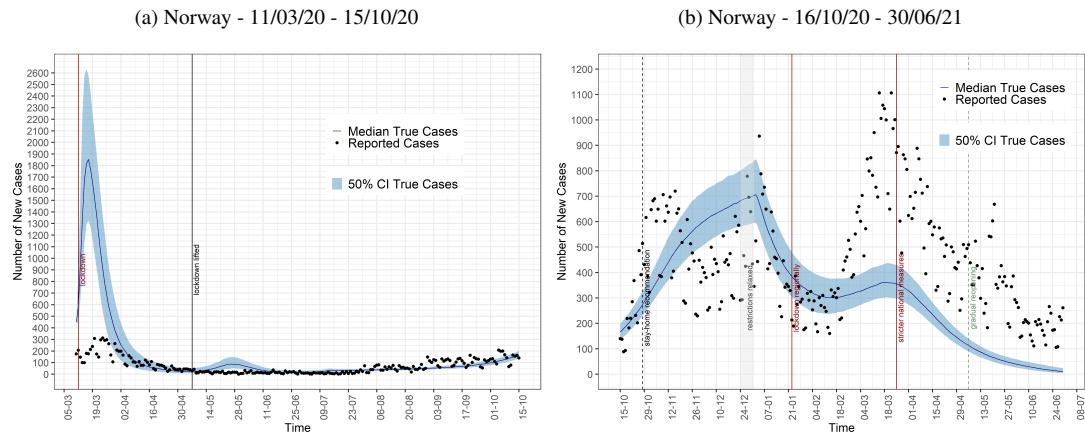


Fig 12: Daily estimated number of total cases, medians(line) and 50% CI(shaded areas) along with reported cases data(dots).

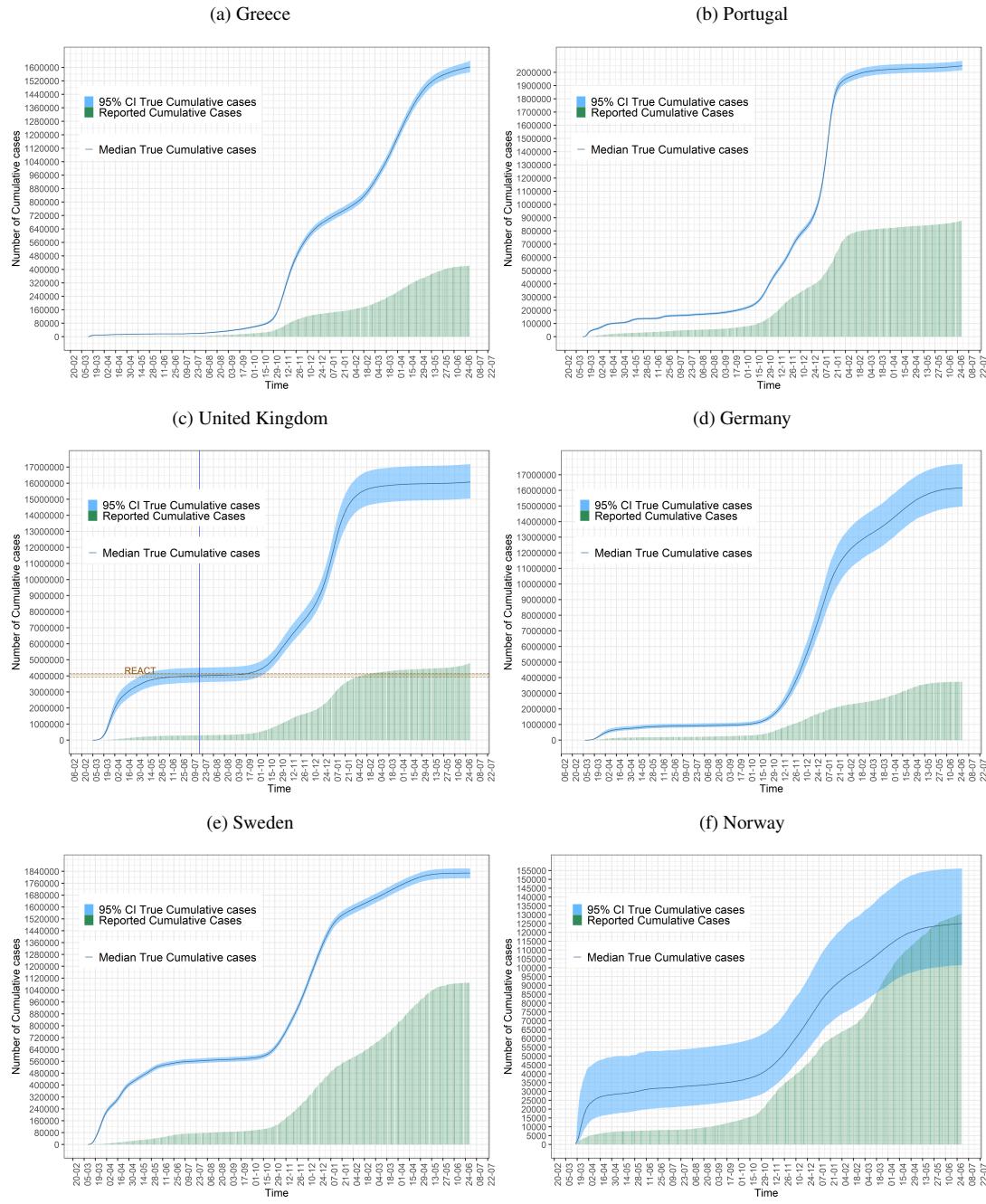


Fig 13: Total population infected, data(bars), median(lines) and 95% CI(shaded areas).

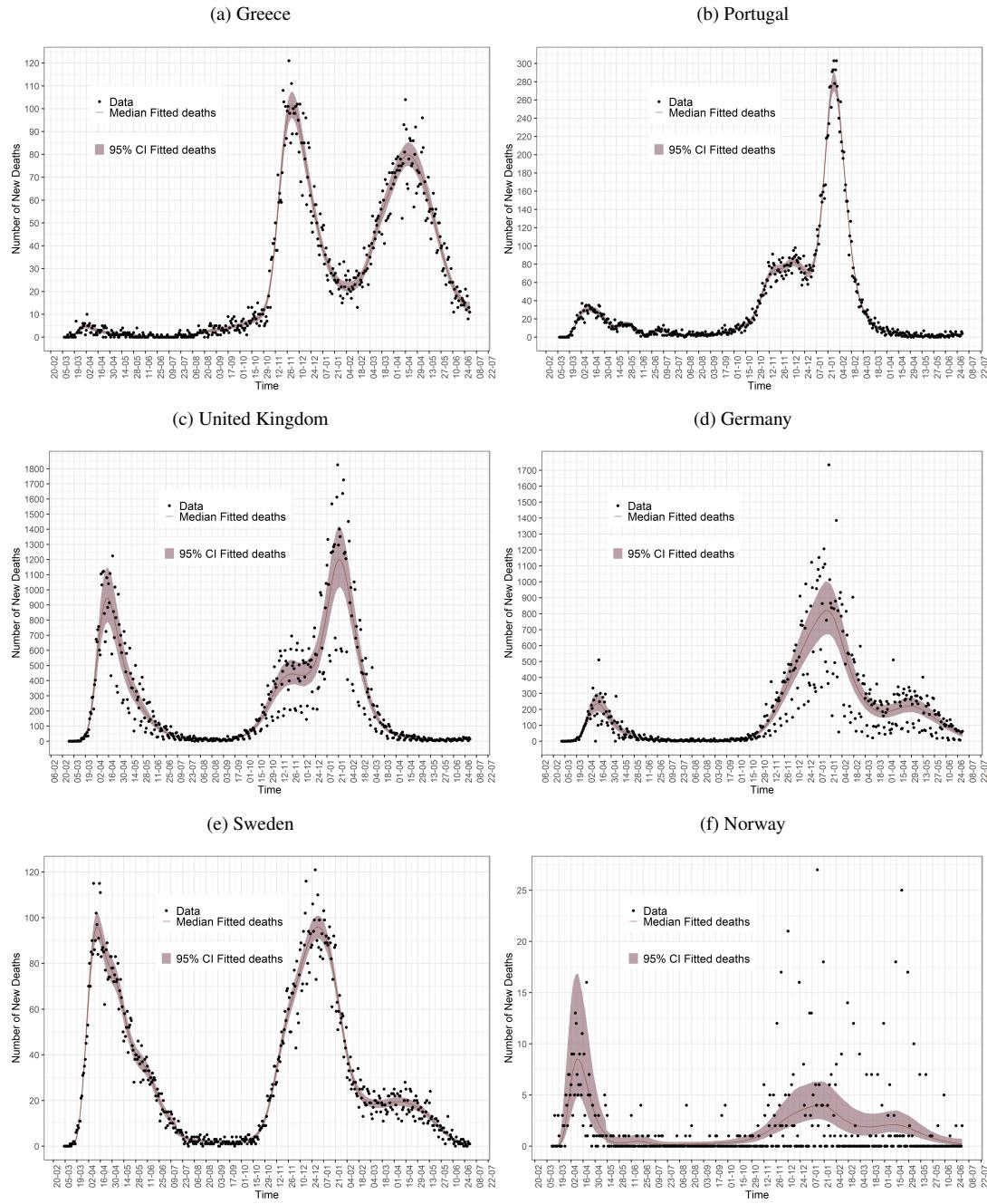


Fig 14: Goodness of fit to data on death counts. Medians(lines) and 95% CI(shaded area) along with reported deaths data(dots).

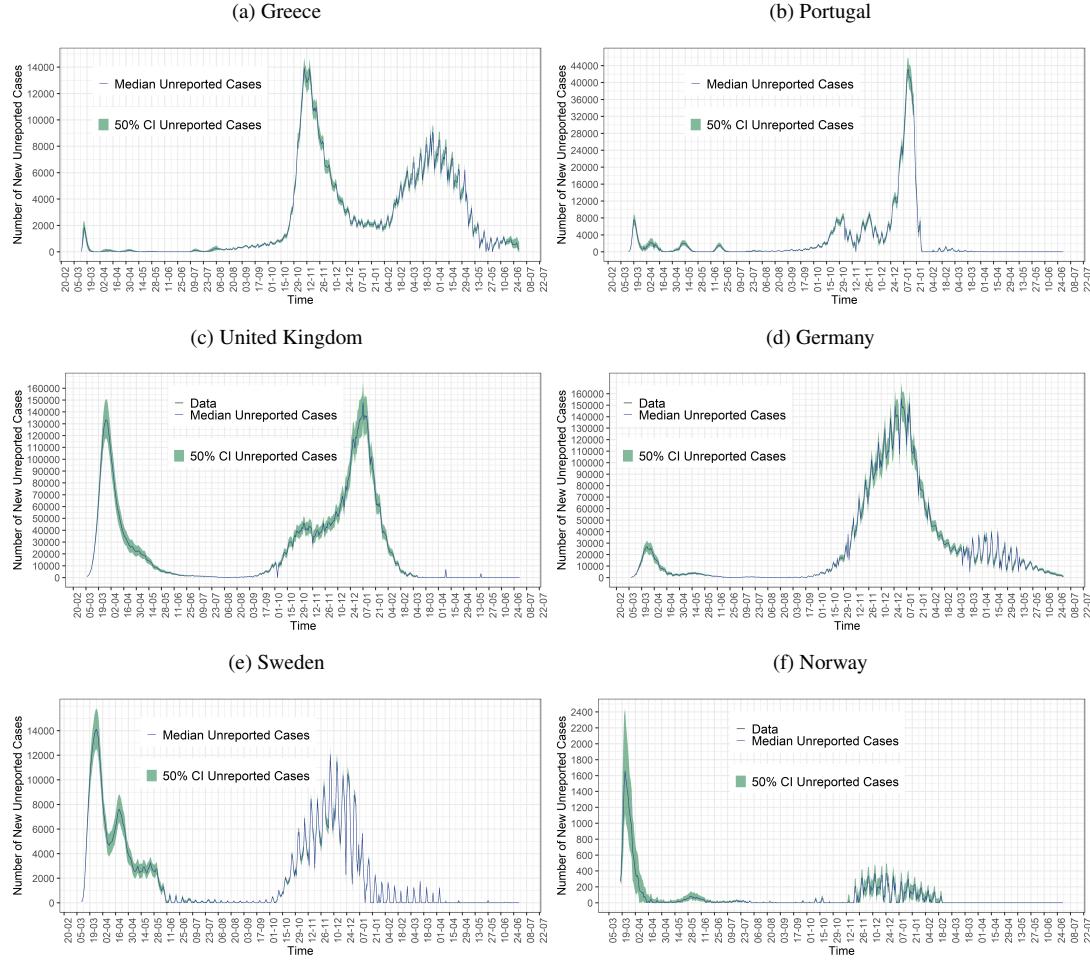


Fig 15: Estimated number of daily unreported cases, medians(lines) and 50% CI(shaded area).  $\text{Unreported}_t = \text{fit\_true\_cases}_{t-6} - \text{reported\_cases}_t$

## Indicative figures for the SEIR model without vaccinations, the case of United Kingdom

We fit the SEIR model without vaccination as described by

$$\begin{aligned}
 \frac{dS_t}{dt} &= -\beta_t S_t \frac{(I_{1t} + I_{2t})}{N} \\
 \frac{dE_{1t}}{dt} &= \beta_t S_t \frac{(I_{1t} + I_{2t})}{N} - \gamma_1 E_{1t} \\
 \frac{dE_{2t}}{dt} &= \gamma_1 E_{1t} - \gamma_1 E_{2t} \\
 \frac{dI_{1t}}{dt} &= \gamma_1 E_{2t} - \gamma_2 I_{1t} \\
 \frac{dI_{2t}}{dt} &= \gamma_2 I_{1t} - \gamma_2 I_{2t} \\
 \frac{dR_t}{dt} &= \gamma_2 I_{2t} \\
 \eta_t &= \mu(\eta_t, \theta_\eta) + \sigma(\eta_t, \theta_\eta) dB_t \\
 \eta_t &= g(\beta_t).
 \end{aligned} \tag{10}$$

using HMC-NUTS with 3 chains, each with 1000 iterations of which the first 500 are warm-up to automatically tune the sampler, leading to a total of 1500 posterior samples. We graphically present our results on  $R_t$  and the daily number of estimated cases for the case of the United Kingdom.

Our results indicate that excluding the significant vaccine rollout, affects  $R_t$  by affecting the relationship between susceptible and infected. The observation process remains the same, therefore, according to our model, observed deaths are the result of the same past incidence. Failing to adopt the effect of vaccinations on the available pool of susceptible and consequently infected individuals, leads  $\beta_t$  to adjust accordingly in order to result to the same number of cases.

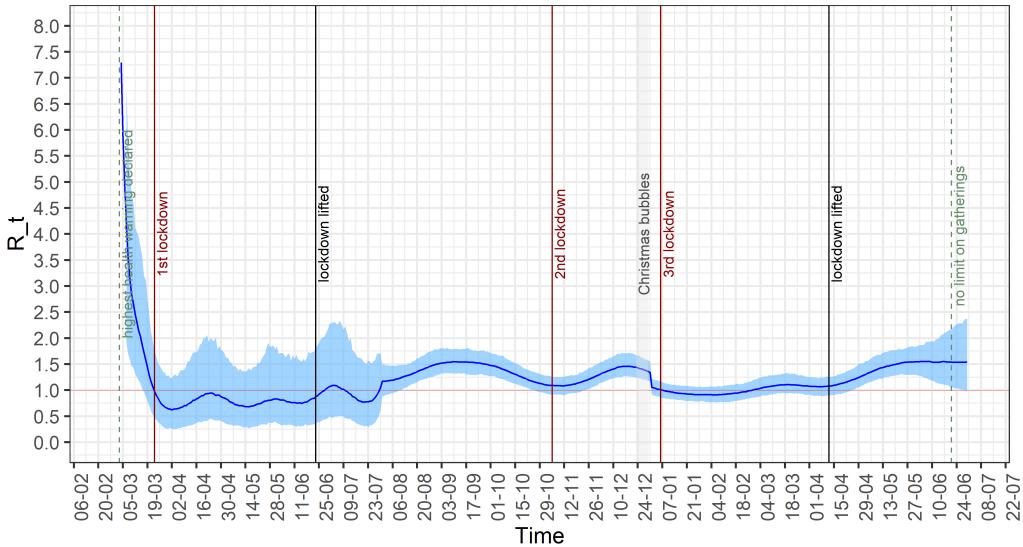


Fig 16: United Kingdom - SEIR without vaccinations. Time-varying reproduction number.

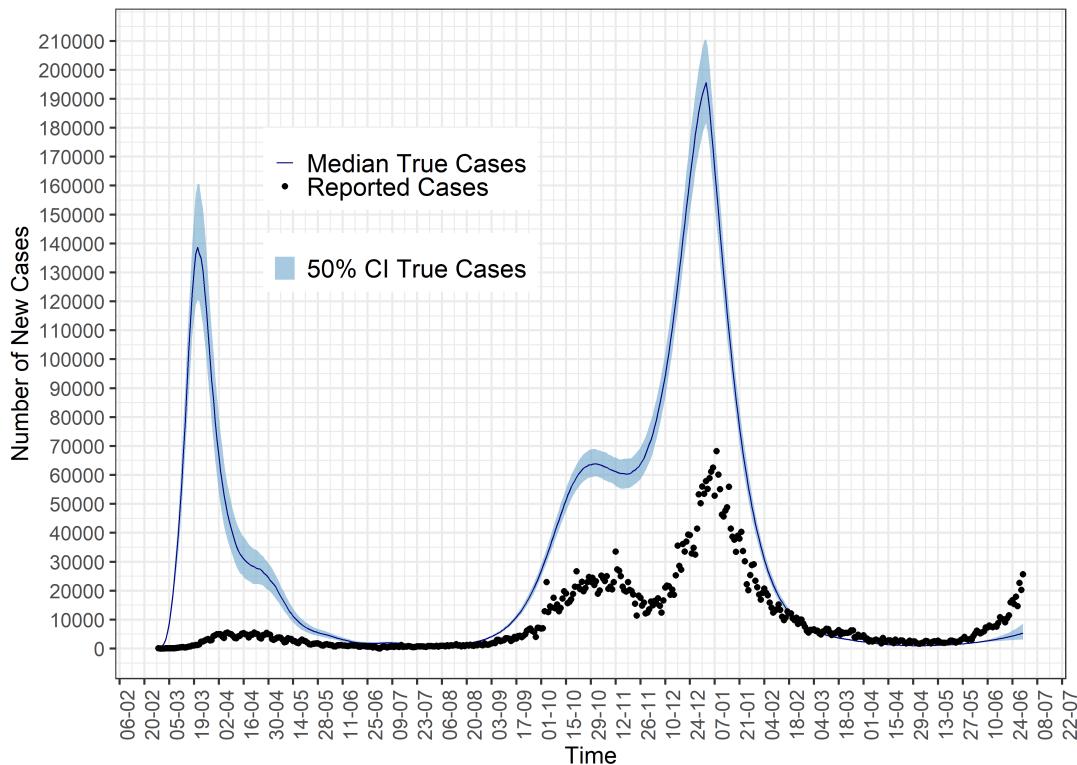


Fig 17: United Kingdom - SEIR without vaccinations. Daily estimated number of total cases, medians(lines) and 50% CI(shaded area).

### Indicative figures for the SIR model with vaccinations, the case of Greece

We consider that the transmission dynamics in the population are addressed using a compartmental SIR model which can be described by a system of ODEs where we model the transmission rate  $\beta_t$  as a diffusion process as in the SEIR model

$$(12) \quad \begin{aligned} \frac{dS}{dt} &= -\beta_t S_t \frac{(I_{1,t} + I_{2,t})}{N} - \rho \nu_{t-35} \\ \frac{dI_1}{dt} &= \beta_t S_t \frac{(I_{1,t} + I_{2,t})}{N} - \gamma I_{1,t} \\ \frac{dI_2}{dt} &= \gamma(I_{1,t} - I_{2,t}) \\ \frac{dR}{dt} &= \gamma I_{2,t} + \rho \nu_{t-35} \end{aligned}$$

where  $S_t$  represents the number of susceptible,  $I_t$  the number of infected and  $R_t$  the number of removed individuals at time  $t$ . The total population size is denoted by  $N$  (with  $N = S_t + I_t + R_t$ ), the vaccine efficacy is denoted by  $\rho$  and set equal to 95% and  $\nu_{t-45}$  is the reported number of individuals who received the first dose of a vaccine 45 weeks prior to time  $t$ . The recovery rate is denoted by  $\gamma$ , for which we consider a Gamma prior distribution, reflecting 5-6 days average infectious period (Li et al., 2020b; Liu et al., 2020). We consider an average infectious period of 6 days during the first year of the pandemic, and we adopt a shorter average infectious period of 5 days for the last several months. In order to link with the available observations, the model-implied daily new infections, denoted by  $c_t$ , are needed i.e.  $c_t = \int_{t-1}^t \beta_s S_s \frac{I_s}{N} ds$ . The rest of the model specification is the same as in the SEIR model.

We fit the model using HMC-NUTS with 3 chains, each with 1000 iterations of which the first 500 are warm-up to automatically tune the sampler, leading to a total of 1500 posterior samples. We graphically present our results on  $R_t$  and the daily number of estimated cases for the case of Greece. Our results indicate that compared to the SEIR model, using the SIR we underestimate  $R_t$ , especially during the first two pandemic waves.

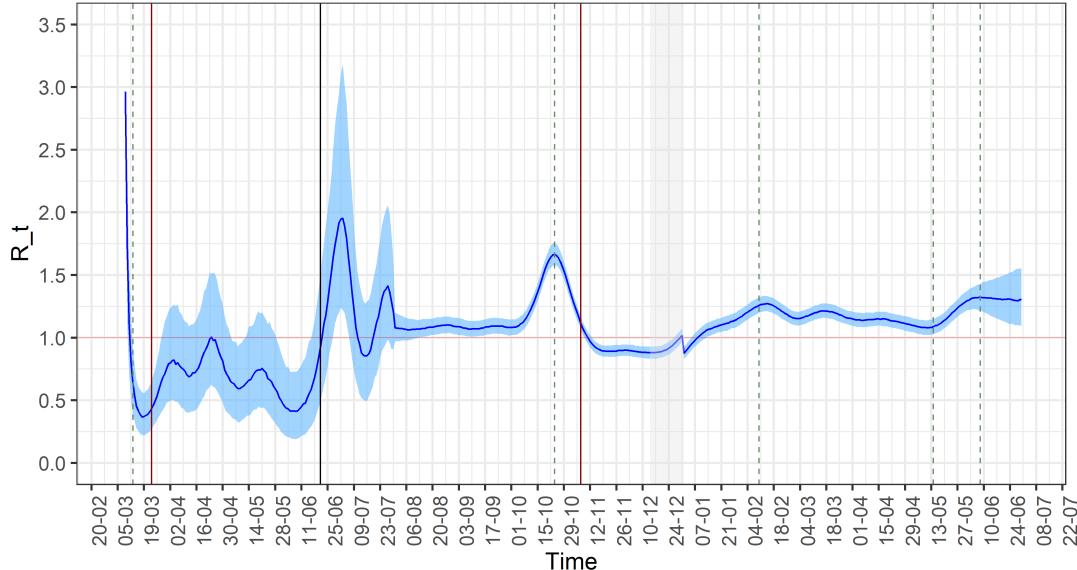


Fig 18: Greece - SIR. Time-varying reproduction number, median(line) and 95% CI(shaded area).

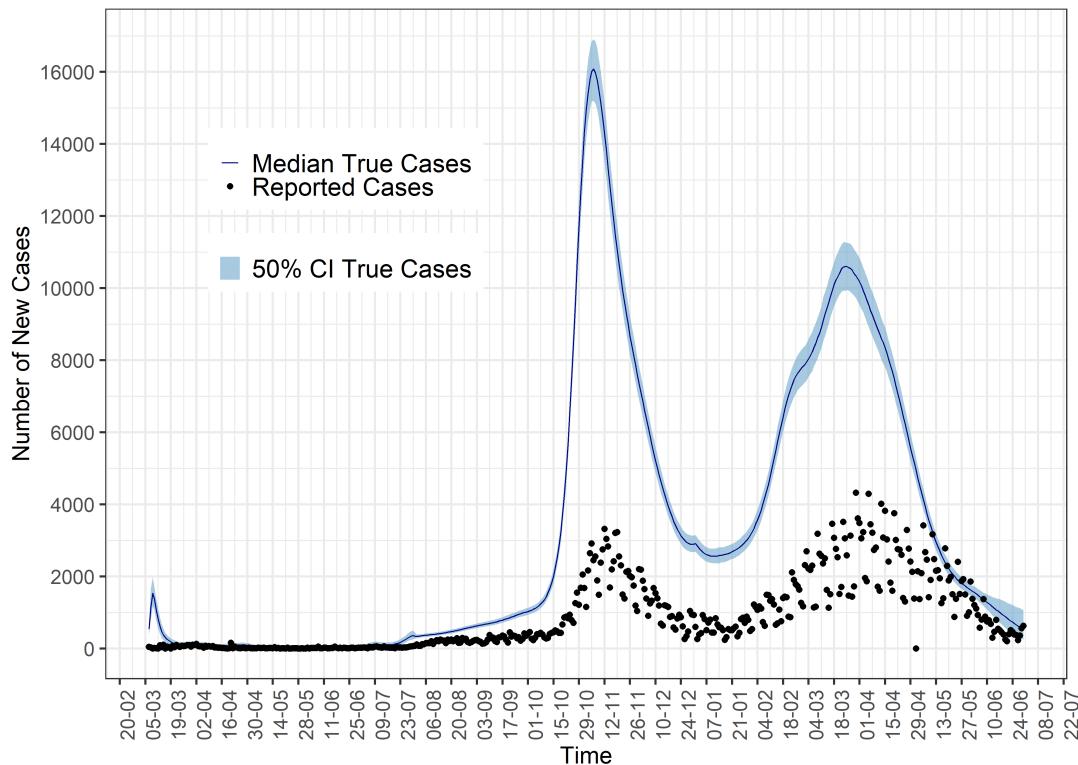


Fig 19: Greece - SIR. Daily estimated number of total cases, median(lines) and 50% CI(shaded area) along with reported cases data(dots).

## Posterior estimates of the multivariate regression coefficients

TABLE 1  
*Greece - Portugal. Posterior estimates of the regression coefficients.*

		Greece			Portugal		
		mean	Quantile (2.5%)	Quantile (97.5%)	mean	Quantile (2.5%)	Quantile (97.5%)
estimated	$m_t$	-295.4448578	-339.2611508	-251.93007122	574.16149938	506.4564514	645.7242736
cases	$testst_{-3}$	0.0333482	0.0281855	0.03835433	0.15112542	0.1342634	0.1672737
cases $c_t$	$testst_{-4}$	0.0298824	0.0243965	0.03555068	0.00246604	-0.0210502	0.0259008
	$testst_{-5}$	0.0295974	0.0241365	0.03504364	0.04776517	0.0238878	0.0717301
	$testst_{-6}$	0.0389645	0.0336866	0.04400839	0.04168455	0.0250962	0.0592665
$\log(\beta_t)$	$m_t$	0.2549224	0.1451619	0.36138759	0.20788732	0.1452496	0.2719134
	$testst_{-3}$	-0.0000105	-0.0000230	0.00000260	-0.00002060	-0.0000354	-0.0000055
	$testst_{-4}$	-0.0000078	-0.0000222	0.00000663	-0.00000343	-0.0000251	0.0000185
	$testst_{-5}$	-0.0000076	-0.0000216	0.00000632	-0.00000657	-0.0000142	0.0000151
	$testst_{-6}$	-0.0000088	-0.0000213	0.00000428	-0.00001310	-0.0000284	0.0000023
$\logit(r_t)$	$m_t$	0.1591968	0.1295273	0.18901434	-0.10138035	-0.1394672	-0.0634241
	$testst_{-3}$	-0.0000058	-0.0000094	-0.00000240	0.00000704	-0.0000022	0.0000160
	$testst_{-4}$	-0.0000046	-0.0000086	-0.00000080	-0.00000081	-0.0000134	0.0000119
	$testst_{-5}$	-0.0000047	-0.0000087	-0.00000099	0.00000102	-0.0000122	0.0000140
	$testst_{-6}$	-0.0000060	-0.0000096	-0.00000255	0.00000134	-0.0000077	0.0000102

TABLE 2  
*United Kingdom - Germany. Posterior estimates of the regression coefficients.*

		United Kingdom			Germany		
		mean	Quantile (2.5%)	Quantile (97.5%)	mean	Quantile (2.5%)	Quantile (97.5%)
estimated	$m_t$	0.1909569	-0.3312650	0.6755159	-144.02662497	-417.4149935	-14.3824210
cases	$testst_{-3}$	-0.00000006	-0.0000106	0.0000101	-0.3894321	-0.4262515	-0.3535754
cases $c_t$	$testst_{-4}$	0.00000081	-0.000010949	0.000010975	-0.0062332	-0.0556456	0.0416999
	$testst_{-5}$	-0.000000016	-0.000010608	0.000011104	-0.0094670	-0.0597152	0.0402142
	$testst_{-6}$	0.000000292	-0.000009458	0.000010404	0.7114286	0.6748734	0.7482453
$\log(\beta_t)$	$m_t$	0.304375006	0.243972270	0.378316687	0.2035188	0.14980518	0.2565617
	$testst_{-3}$	-0.000000333	-0.000000970	0.000000289	-0.0000117	-0.0000223	-0.0000008
	$testst_{-4}$	-0.000000318	-0.000000967	0.000000323	-0.0000003	-0.00001537	0.0000146
	$testst_{-5}$	-0.000000030	-0.000000960	0.000000357	-0.0000005	-0.00001564	0.0000155
	$testst_{-6}$	-0.000000368	-0.000001002	0.000000265	0.0000043	-0.00000661	0.0000154
$\logit(r_t)$	$m_t$	0.4096307	0.3597061	0.4594726	0.3856904	0.3415282	0.4277465
	$testst_{-3}$	-0.000000831	-0.000001344	-0.000000304	0.000003434	-0.000006063	0.000012726
	$testst_{-4}$	0.000000444	-0.000000116	0.000001002	-0.000000210	-0.000012631	0.000013038
	$testst_{-5}$	0.000000745	0.000000162	0.0000001338	-0.000000571	-0.000013950	0.000012685
	$testst_{-6}$	0.000002207	0.000001678	0.000002728	-0.000005157	-0.000014417	0.000004312

TABLE 3  
*Sweden - Norway. Posterior estimates of the regression coefficients.*

		Sweden			Norway		
		mean	Quantile (2.5%)	Quantile (97.5%)	mean	Quantile (2.5%)	Quantile (97.5%)
estimated	$m_t$	-0.00515295	-1.6758185	1.5175613	-53.074619	-60.6580347	-45.7226692
cases	$testst_{-3}$	-0.00000946	-0.0017866	0.0017169	0.005092	0.0033033	0.0067980
cases $c_t$	$testst_{-4}$	0.000003193	-0.0023371	0.0024826	0.003066	0.0009381	0.0052173
	$testst_{-5}$	-0.000004805	-0.0026163	0.0024664	0.002865	0.0006670	0.0050890
	$testst_{-6}$	0.000002729	-0.0018085	0.0018595	0.005179	0.0033629	0.0069498
$\log(\beta_t)$	$m_t$	-0.26216474	-0.3108732	-0.2132896	-0.062477	-0.1398509	0.0132023
	$testst_{-3}$	0.00000970	-0.0000455	0.0000650	-0.000028	-0.0000472	-0.0000091
	$testst_{-4}$	-0.00000066	-0.0000790	0.0000768	-0.000016	-0.0000403	0.0000081
	$testst_{-5}$	0.000000281	-0.0000769	0.0000825	-0.000016	-0.0000390	0.0000090
	$testst_{-6}$	-0.000006618	-0.0001236	-0.0000091	-0.000027	-0.0000465	-0.0000072
$\logit(r_t)$	$m_t$	0.19390353	0.1518186	0.2349748	0.295694	0.2042354	0.3862325
	$testst_{-3}$	0.000002234	-0.0000246	0.0000696	0.000046	0.0000236	0.0000686
	$testst_{-4}$	0.000000523	-0.0000636	0.0000757	0.000025	-0.0000029	0.0000524
	$testst_{-5}$	0.000000281	-0.0000628	0.0000692	0.000028	0.0000004	0.0000562
	$testst_{-6}$	0.000005752	0.0000086	0.0001079	0.000047	0.0000243	0.0000697

Tables 4, 5, 6 present posterior means and credible intervals for some indicative model parameters for each country, as well as effective sample sizes and  $\hat{R}$ s. Trace plots showing the evolution of parameter vector over the iterations of multiple Markov chains starting from distinct initial values are illustrated in Figures 20. Trace plots offer a visual way to inspect sampling behaviour and assess mixing across chains and convergence.

TABLE 4  
*Greece-Portugal - Inference results using HMC-NUTS, 3 chains, each with iter=1000; warmup=500; thin=1; post-warmup draws per chain=500, total post-warmup draws=1500*

	Greece				Portugal			
	mean	95% CI	ESS	$\hat{R}$	mean	95% CI	ESS	$\hat{R}$
$\beta_0$	4.615802	0.29894 - 21.18123	917	1	3.851096	0.50425969 - 11.940254	907	1
$\beta_1$	34.49046	2.038184 - 93.808547	1027	1	6.295344	0.87133962 - 18.313026	1388	1
$\beta_{27}$	1.042928	0.00000199 - 6.70228461	872	1	0.232046	0.01785919 - 0.905857	1835	1
$\beta_{69}$	0.727328	0.00000002 - 3.58338859	904	1	0.059500	0.00089268 - 0.273725	1396	1
$\beta_{141}$	1.377333	0.00000305 - 10.17028789	949	1	0.232494	0.01080004 - 0.932645	1331	1
$\beta_{180}$	0.226725	0.1747956 - 0.2978135	1028	1	0.247898	0.16725222 - 0.352656	1413	1
$\beta_{220}$	0.290101	0.2290207 - 0.3539788	975	1	0.330537	0.23669781 - 0.445904	1524	1
$\beta_{257}$	0.169151	0.1343898 - 0.2073608	1035	1	0.254199	0.18419250 - 0.354344	907	1
$\beta_{292}$	0.178706	0.1383344 - 0.2207352	1038	1	0.345976	0.24599117 - 0.468424	1540	1
$\beta_{365}$	0.295586	0.2380134 - 0.3667774	947	1	0.254697	0.16815267 - 0.380040	1463	1
$\beta_{440}$	0.276727	0.2176132 - 0.3488697	1035	1	0.484066	0.30432596 - 0.727334	609	1
$\beta_{481}$	0.338702	0.1567588 - 0.6457428	1002	1	0.480271	0.09747981 - 1.299428	830	1
$\gamma_1$	1.000496	0.9809176 - 1.0202894	932	1	1.000063	0.97973082 - 1.020083	1848	1
$\gamma_2^1$	0.400871	0.3816949 - 0.4211050	1108	1	0.398595	0.37949189 - 0.417800	2001	1
$\gamma_2^2$	0.498274	0.4777484 - 0.5194934	1023	1	0.500113	0.48128268 - 0.519825	2334	1
$\sigma_1$	2.174051	0.5251264 - 6.0078186	1033	1	0.864314	0.48561628 - 1.546247	34	1.1
$\sigma_2$	0.07393	0.038995 - 0.097686	34	1	0.130114	0.08695116 - 0.204222	39	1.1
$\sigma_3$	25.89806	0.1960101 - 145.7326258	861	1	19.446863	0.19918285 - 132.961406	970	1
$ifr_1$	0.01142	0.0114178 - 0.0114219	1066	1	0.011577	0.01157447 - 0.011579	1112	1
$ifr_2$	0.00799	0.0079923 - 0.0079955	977	1	0.008165	0.00816375 - 0.008167	1513	1
$ifr_3$	0.01142	0.0114180 - 0.0114221	990	1	0.011577	0.01157466 - 0.011579	1610	1
$ifr_4$	0.0115	0.0114977 - 0.0115020	913	1	0.00001	0.00000994 - 0.00001	1551	1
$ifr_5$	0.0002	0.0001997 - 0.0002003	1013	1	0.000001	0.00000098 - 0.000001	1744	1
$\phi_d$	144.2365	73.18127 - 277.98176	873	1	987.564554	191.69415749 - 5122.065664	811	1

TABLE 5

**United Kingdom-Germany - Inference results using HMC-NUTS, 3 chains, each with iter=1000; warmup=500; thin=1; post-warmup draws per chain=500, total post-warmup draws=1500**

	United Kingdom				Germany			
	mean	95% CI	ESS	$\hat{R}$	mean	95% CI	ESS	$\hat{R}$
$\beta_0$	1.685376	0.896673 - 2.855327	1158	1	1.58484	0.73729 - 3.09	1680	1
$\beta_1$	1.663571	1.010392 - 2.554535	1178	1	1.57477	0.82737 - 2.77919	875	1
$\beta_{27}$	0.173172	0.096332 - 0.279496	1765	1	0.14706	0.06608 - 0.25969	1849	1
$\beta_{69}$	0.164774	0.085399 - 0.271964	1479	1	0.22110	0.09871 - 0.42800	1688	1
$\beta_{141}$	0.181026	0.098053 - 0.306026	1507	1	0.18137	0.07878 - 0.34586	1403	1
$\beta_{180}$	0.292794	0.235633 - 0.368226	1666	1	0.24388	0.21275 - 0.28120	1217	1
$\beta_{220}$	0.294394	0.240203 - 0.359728	1565	1	0.30053	0.26304 - 0.34464	924	1
$\beta_{257}$	0.211272	0.166239 - 0.264423	1574	1	0.24610	0.21354 - 0.28108	1407	1
$\beta_{292}$	0.298911	0.244495 - 0.365559	1284	1	0.23105	0.20299 - 0.26464	1204	1
$\beta_{365}$	0.259699	0.200892 - 0.320179	1439	1	0.29677	0.25588 - 0.33717	1113	1
$\beta_{440}$	0.752368	0.600701 - 0.945950	1126	1	0.29476	0.25113 - 0.34098	935	1
$\beta_{487}$	0.928370	0.462613 - 1.629142	714	1	0.30556	0.19446 - 0.45061	545	1
$\gamma_1$	0.999954	0.980884 - 1.020071	1486	1	1.000053	0.98129 - 1.02031	1740	1
$\gamma_1^1$	0.399102	0.378741 - 0.418424	1052	1	0.40008	0.38148 - 0.42038	772	1
$\gamma_2^2$	0.500156	0.493429 - 0.519969	1441	1	0.49989	0.48033 - 0.51888	704	1
$\sigma_1$	0.187932	0.114836 - 0.305300	46	1.1	0.24665	0.13962 - 0.40548	35	1.1
$\sigma_2$	0.059774	0.038995 - 0.097686	27	1.1	0.03292	0.01933 - 0.05720	16	1.1
$\sigma_3$	18.113676	0.261822 - 114.199836	1070	1	38.86222	0.19314 - 122.77661	966	1
ifr <sub>1</sub>	0.010350	0.010348 - 0.010352	1492	1	0.01104	0.01104 - 0.01104	1440	1
ifr <sub>2</sub>	0.007245	0.007243 - 0.007247	1499	1	0.00543	0.00543 - 0.00543	1608	1
ifr <sub>3</sub>	0.009500	0.009498 - 0.009502	1514	1	0.01142	0.01142 - 0.01142	1433	1
ifr <sub>4</sub>	0.00004	0.00004 - 0.00004	1489	1	0.005	0.005 - 0.005	1550	1
ifr <sub>5</sub>	0.000018	0.000018 - 0.000018	1414	1	0.00015	0.00015 - 0.00015	1735	1
$\phi_d$	4.674407	4.016512 - 5.374142	1808	1	2.79597	2.41610 - 3.22119	1720	1

TABLE 6

**Norway-Sweden - Inference results using HMC-NUTS, 3 chains, each with iter=1000; warmup=500; thin=1; post-warmup draws per chain=500, total post-warmup draws=1500**

	Norway				Sweden			
	mean	95% CI	ESS	$\hat{R}$	mean	95% CI	ESS	$\hat{R}$
$\beta_0$	1.685376	0.896673 - 2.855327	1158	1	3.673774	1.520351 - 7.432186	1446	1
$\beta_1$	1.663571	1.010392 - 2.554535	1178	1	3.711570	1.876315 - 7.090244	307	1
$\beta_{27}$	0.173172	0.096332 - 0.279496	1765	1	0.124197	0.049109 - 0.225367	468	1
$\beta_{69}$	0.164774	0.085399 - 0.271964	1479	1	0.215347	0.078990 - 0.421069	1530	1
$\beta_{141}$	0.181026	0.098053 - 0.306026	1507	1	0.261902	0.094716 - 0.584609	282	1
$\beta_{180}$	0.292794	0.235633 - 0.368226	1666	1	0.242492	0.121181 - 0.414250	1654	1
$\beta_{220}$	0.294394	0.240203 - 0.359728	1565	1	0.356899	0.336094 - 0.425916	527	1
$\beta_{257}$	0.211272	0.166239 - 0.264423	1574	1	0.235605	0.202738 - 0.268980	1380	1
$\beta_{292}$	0.298911	0.244495 - 0.365559	1284	1	0.211575	0.184675 - 0.244751	906	1
$\beta_{365}$	0.259699	0.200892 - 0.320179	1439	1	0.313699	0.268676 - 0.372263	1262	1
$\beta_{440}$	0.752368	0.600701 - 0.945950	1126	1	0.234077	0.184762 - 0.286273	730	1
$\beta_{484}$	0.928370	0.462613 - 1.629142	714	1	0.246511	0.121231 - 0.457712	396	1
$\gamma_1$	0.999954	0.980884 - 1.020071	1486	1	1.000097	0.981814 - 1.019264	1785	1
$\gamma_1^1$	0.399102	0.378741 - 0.418424	1052	1	0.399872	0.380822 - 0.419252	808	1
$\gamma_2^2$	0.500156	0.493429 - 0.519969	1441	1	0.499424	0.480374 - 0.518784	1030	1
$\sigma_1$	0.187932	0.114836 - 0.305300	46	1.1	0.315119	0.185836 - 0.601009	22	1.1
$\sigma_2$	0.059774	0.038995 - 0.097686	27	1.1	0.048906	0.030847 - 0.074904	18	1.1
$\sigma_3$	18.113676	0.261822 - 114.199836	1070	1	53.621081	0.199574 - 130.509926	523	1
ifr <sub>1</sub>	0.010350	0.010348 - 0.010352	1492	1	0.0103	0.010298 - 0.010302	1439	1
ifr <sub>2</sub>	0.007245	0.007243 - 0.007247	1499	1	0.007	0.006998 - 0.007002	1377	1
ifr <sub>3</sub>	0.009500	0.009498 - 0.009502	1514	1	0.009	0.008998 - 0.009002	1325	1
ifr <sub>4</sub>	0.00004	0.00004 - 0.00004	1489	1	0.0002	0.0002 - 0.0002	1443	1
ifr <sub>5</sub>	0.000018	0.000018 - 0.000018	1414	1	0.000002	0.000002 - 0.000002	1844	1
$\phi_d$	4.674407	4.016512 - 5.374142	1808	1	676.426927	147.239174 - 2842.887983	718	1



## Detailed data

TABLE 7  
*Detailed Data References*

Data	References
<b>Population</b>	
All countries	Worldometer, <a href="https://www.worldometers.info/world-population/population-by-country/">https://www.worldometers.info/world-population/population-by-country/</a>
<b>Cumulative deaths</b>	
Greece, Portugal, United Kingdom, Germany, Norway	COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University. <a href="https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data">https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data</a>
Sweden	Folkhälsomyndigheten, Public Health Agency of Sweden. <a href="https://www.folkhälsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/statistik-och-analys/bekräftade-fall-i-sverige/">https://www.folkhälsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/statistik-och-analys/bekräftade-fall-i-sverige/</a>
<b>Cumulative cases</b>	
Greece, Portugal, United Kingdom, Norway	COVID-19 Data Repository by the Center for Systems Science and Engineering at Johns Hopkins University. <a href="https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data">https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data</a>
Sweden	Folkhälsomyndigheten, Public Health Agency of Sweden. <a href="https://www.folkhälsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/statistik-och-analys/bekräftade-fall-i-sverige/">https://www.folkhälsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19/statistik-och-analys/bekräftade-fall-i-sverige/</a>
<b>Age distribution of cases</b>	
Greece	Hellenic National Public Health Organization. <a href="https://cody.gov.gr/epidemiologika-statistika-dedomena/ekthesis-covid-19/">https://cody.gov.gr/epidemiologika-statistika-dedomena/ekthesis-covid-19/</a>
Portugal	Directorate General for Health via Data Science for Social Good. <a href="https://github.com/dssg-pt/covid19pt-data">https://github.com/dssg-pt/covid19pt-data</a>
United Kingdom	Public Health England. <a href="https://coronavirus.data.gov.uk/details/download">https://coronavirus.data.gov.uk/details/download</a>
Germany	Robert Koch Institute, Federal Ministry of Health. <a href="https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Altersverteilung.html">https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Daten/Altersverteilung.html</a>
Sweden	COVerAGE-DB. <a href="https://imrnrf.github.io/covrage/">https://imrnrf.github.io/covrage/</a> <a href="https://imrnrf.github.io/covrage/GettingStarted.html">GettingStarted.html</a>
Norway	Norwegian Institute of Public Health. <a href="https://www.fhi.no/en/fd/infectious-diseases/coronavirus/daily-reports/daily-reports-COVID19/covid19associated-deaths-by-age-and-sex">https://www.fhi.no/en/fd/infectious-diseases/coronavirus/daily-reports/daily-reports-COVID19/covid19associated-deaths-by-age-and-sex</a>
<b>Daily vaccinations with 1st dose</b>	
Greece	Hellenic National Public Health Organization. Retrieved from: <a href="https://docs.google.com/spreadsheets/d/14rK14TAM0SYWpj4u3rAkS2PKTSqYzdCeuXVMV6ptM/edit#gid=782062930">https://docs.google.com/spreadsheets/d/14rK14TAM0SYWpj4u3rAkS2PKTSqYzdCeuXVMV6ptM/edit#gid=782062930</a>
Portugal	Directorate General for Health via Data Science for Social Good. <a href="https://github.com/dssg-pt/covid19pt-data">https://github.com/dssg-pt/covid19pt-data</a>
United Kingdom, Germany, Sweden, Norway	Our World In Data. <a href="https://ourworldindata.org/grapher/people-vaccinated-covid">https://ourworldindata.org/grapher/people-vaccinated-covid</a>
<b>Daily tests</b>	
Greece, Portugal, United Kingdom, Sweden, Norway	Our World In Data. <a href="https://ourworldindata.org/grapher/full-list-covid-19-tests-per-day">https://ourworldindata.org/grapher/full-list-covid-19-tests-per-day</a>
Germany	Robert Koch Institute, Federal Ministry of Health. <a href="https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Testzahl.html;jsessionid=821BB52428F222307CBE3BA46A8A4106.internet091?nn=13490888">https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Testzahl.html;jsessionid=821BB52428F222307CBE3BA46A8A4106.internet091?nn=13490888</a>
<b>Mobility trends</b>	
All countries	COVID-19 Community Mobility Report (Google, 2021). <a href="https://www.google.com/covid19/mobility/?hl=en">https://www.google.com/covid19/mobility/?hl=en</a>
<b>Government responses</b>	
All countries	Hale et al. (2021). <a href="https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker">https://www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker</a>