RESEARCH ARTICLE

A Bayesian hierarchical approach to account for evidence and uncertainty in the modeling of infectious diseases: An application to COVID-19 📊

Raphael Rehms[1,2] 🆔 | **Nicole Ellenbach**[1,2] | **Eva Rehfuess**[1,2] | **Jacob Burns**[1,2] | **Ulrich Mansmann**[1,2,3] 🆔 | **Sabine Hoffmann**[1,2,3]

[1]Institute of Medical Data Processing, Biometrics and Epidemiology (IBE), Faculty of Medicine, Ludwig-Maximilians-University Munich, Munich, Germany

[2]Pettenkofer School of Public Health, Ludwig-Maximilians-University Munich, Munich, Germany

[3]Department of Statistics, Ludwig-Maximilians-University Munich, Munich, Germany

Correspondence
Raphael Rehms, Institute of Medical Data Processing, Biometrics and Epidemiology (IBE), Faculty of Medicine, Ludwig-Maximilians-University Munich, Marchioninistr. 15, 81377 Munich, Germany.
Email: rrehms@ibe.med.uni-muenchen.de

Abstract

Infectious disease models can serve as critical tools to predict the development of cases and associated healthcare demand and to determine the set of nonpharmaceutical interventions (NPIs) that is most effective in slowing the spread of an infectious agent. Current approaches to estimate NPI effects typically focus on relatively short time periods and either on the number of reported cases, deaths, intensive care occupancy, or hospital occupancy as a single indicator of disease transmission. In this work, we propose a Bayesian hierarchical model that integrates multiple outcomes and complementary sources of information in the estimation of the true and unknown number of infections while accounting for time-varying underreporting and weekday-specific delays in reported cases and deaths, allowing us to estimate the number of infections on a daily basis rather than having to smooth the data. To address dynamic changes occurring over long periods of time, we account for the spread of new variants, seasonality, and time-varying differences in host susceptibility. We implement a Markov chain Monte Carlo algorithm to conduct Bayesian inference and illustrate the proposed approach with data on COVID-19 from 20 European countries. The approach shows good performance on simulated data and produces posterior predictions that show a good fit to reported cases, deaths, hospital, and intensive care occupancy.

KEYWORDS
Bayesian modeling, COVID-19, hierarchical models, infectious diseases, modeling uncertainty

1 | INTRODUCTION

The experience with severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) causing coronavirus disease 2019 (COVID-19) has underlined both the importance of and the challenges in the modeling of infectious diseases. Infectious disease models can serve as critical tools to predict epidemic development and healthcare demand and to determine when and which nonpharmaceutical interventions (NPIs) should be implemented to slow the spread of an infectious agent. However, the modeling of infectious diseases is complicated by the fact that the main quantity of interest, that is, the number of daily infections, is a latent variable that cannot be observed and therefore has to be estimated by using information on other observable quantities. The number of reported cases, hospital occupancy, and deaths all provide complementary, yet incomplete and sometimes even contradictory, information on the number of infections in a given geographical region. The number of reported cases is prone to underreporting, as it depends on the reliability of tests, the testing capacity, and the employed testing strategy (May, 2020; Pullano et al., 2021). When changes in the testing strategy concur with the introduction or the relaxation of NPIs, they can create severe distortions in the estimation of NPI effects. While modeling disease mortality, which is less prone to underreporting, can avoid biases due to changes in the testing strategy, it is difficult to predict future healthcare demand based on disease mortality. Moreover, the time lag between infection and death may be highly variable, leading to a reduction of statistical power (Sharma et al., 2021) and where (more or less effective) treatment is available, differences in the availability, accessibility, and quality of healthcare may affect case fatality rates. A possible solution for this situation, in which there are several imperfect proxy variables for the number of infections, is to combine information on disease incidence, mortality, and hospital occupancy in a common framework while explicitly accounting for time-varying underreporting and reporting delays.

Bayesian hierarchical approaches can address this challenge by providing a coherent and flexible framework to integrate all available sources of information while accounting for different sources of uncertainty. By combining different submodels through conditional independence assumptions, it is possible to integrate mechanistic assumptions on disease dynamics and submodels describing the relationship between the true (and unknown) number of infections and reported cases, hospital occupancy, and deaths. Additionally, we can borrow information from other geographical regions to stabilize parameter estimates and to improve forecasts on future healthcare demand. Current approaches to assess the effect of NPIs typically either focus on the number of deaths (Flaxman et al., 2020) or the number of cases (Banholzer et al., 2021; Dehning et al., 2020; Islam et al., 2020; Li et al., 2021b). Unwin et al. (2020), Brauner et al. (2021), and Sharma et al. (2021) extend the semimechanistic Bayesian hierarchical model proposed by Flaxman et al. (2020) by including information on reported cases and deaths when inferring the number of new infections. However, these approaches typically only estimate NPI effects for short time periods because they do not explicitly account for differences in host susceptibility over time (due to vaccination or previous infection), seasonality, the prevalence of different virus variants, or time-varying underreporting.

Here, we show how a Bayesian hierarchical approach can be used to integrate the available information on the number of reported cases, the number of deaths, and hospital and intensive care unit (ICU) occupancy in the estimation of the true and unknown number of infections while accounting for underreporting and reporting delays in the number of reported cases. We account for the influence of seasons, previous infections, vaccination coverage, and the prevalence of different virus variants as these factors can have a critical influence on the number of new infections and on disease severity. By doing so, it is possible to use data over long time periods in several countries rather than focusing on short time periods in a single country during which the prevailing variant, vaccination coverage, and the testing strategy remained roughly constant. By allowing for weekday-specific delays in reported cases and deaths (which mainly arise due to reduced reporting during the weekend), we are not required to smooth the analyzed time series and we can estimate the number of infections on a daily basis. We illustrate the proposed approach using data for COVID-19 from 20 European countries and investigate its performance both on simulated data and by assessing how well the model describes reported cases, hospital and ICU occupancy, and deaths through posterior predictive checks.

The rest of the paper is organized as follows. In Section 2, we describe the proposed Bayesian hierarchical model. In Section 3, we present a simulation study to assess the performance of our proposed approach. A case study on the modeling of COVID-19 in 20 European countries is presented in Section 4. In Section 5, we summarize this article with a brief discussion.
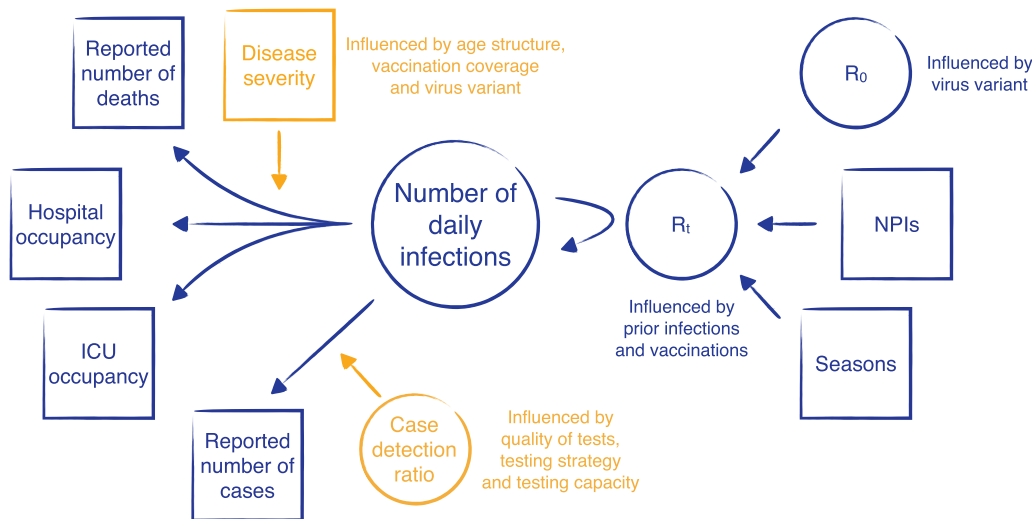
FIGURE 1 Simplified directed acyclic graph (DAG) describing how the true and unknown number of infections is estimated through the four observed time series. Parameters are shown in orange and variables are shown in blue. Unknown quantities that have to be estimated are given in circles and quantities that are either observed or assumed to be known are given in squares.

2 | THE MODEL

In this section, we describe the main elements of the Bayesian hierarchical model. At its core, the model treats the number of true and unknown infections $I_{t,m} \in \mathbb{N}_0$ at time t in geographical region m as a discrete latent variable where t and m are considered as discrete index sets (e.g., days and countries). We describe the model in two parts. First, we describe how we infer the number of true and unknown infections (i.e., the values of the latent variable) from the available information on reported cases, hospital and ICU occupancy, and deaths while accounting for time-varying underreporting, weekday-specific reporting delays, and changes in the severity of the disease due to vaccination coverage and different virus variants. Second, we describe how we estimate the effects of NPIs given the true and unknown number of infections represented through the discrete latent variable while accounting for seasonality, time-varying differences in host susceptibility and changes in the transmissibility of the virus due to different virus variants.

2.1 | Inferring the number of infections

As illustrated in Figure 1, we use information on the reported number of cases, deaths, and hospital and ICU occupancy to estimate the true and unknown number of infections $I_{t,m}$ at time t in geographical region m. Each of these observed time series is linked through a submodel to this discrete latent variable: the reporting model, the death model, and two hospitalization models (normal beds and ICU). Before linking the number of infections to the observable time series, we define a second latent variable, the number of cases $C_{t,m} \in \mathbb{N}_0$ in geographical region m with symptom onset on day t, which is simply a deterministic function of the number of infections $I_{t,m}$ occurring until time t, as described through the following *disease model*:

$$C_{t,m} = \sum_{u \leq t} I_{t,m}\big(F_{\xi C}(t - u + 1) - F_{\xi C}(t - u)\big),$$

where $F_{\xi C}$ is the cumulative distribution function of the incubation period. Note that we do not distinguish infections by strength of symptoms. Some infected individuals may even experience so weak symptoms that they are not noticed, and in this case, the incubation period is of merely technical nature. Through the disease model, the number of (symptomatic and asymptomatic) cases becomes a deterministic function of the number of infections, which is simply shifted by the incubation time distribution.

The number of cases $C_{t,m}$ is linked to the number of reported cases $C_{t,m}^R$ through the following *reporting model*:

$$C_{t,m}^R \sim NegBinom\left(\mu_{t,m}^R, \phi^R\right),$$

where

$$\mu_{t,m}^R = \rho_{t,m} \sum_{u<t} C_{u,m}\left(F_{\xi_m^{R,w}}(t-u+1) - F_{\xi_m^{R,w}}(t-u)\right),$$

where $F_{\xi_m^{R,w}}$ is the reporting delay distribution for a specific weekday w in geographical region m. In this model, the number of reported cases follows a negative binomial distribution where the expected number of reported cases on day t is described as the sum of all true cases occurring on some day u before day t weighted by their probability of being reported after $t-u$ days and multiplied by a time-specific underreporting rate $\rho_{t,m}$. We choose a negative binomial distribution rather than a Poisson distribution to allow for overdispersion (controlled by the size parameter ϕ^R). The variance of $C_{t,m}^R$ is then given by $\mu_{t,m}^R + \mu_{t,m}^{R\,2}/\phi^R$. Therefore, for high values of ϕ^R relative to $\mu_{t,m}^R$, the distribution resembles a Poisson distribution and low values indicate high overdispersion. The time-specific underreporting rates $\rho_{t,m}$ are modeled through a piece-wise constant function. By accounting for time-varying underreporting and weekday-specific reporting delays, the reporting model allows for discrepancies between the true dynamics of the disease and the number of cases that are reported by health authorities (Höhle & an der Heiden, 2014). In the reporting model, we assume that the delay between symptom onset and day of reporting can be specific to a geographical region m. Since it is very difficult to obtain this information for each region, we use information from a specific geographical region for which we can estimate these weekday-specific reporting delay distributions (in our application to COVID-19, this region is Bavaria in Germany) and adapt them for each location. The whole procedure is as follows. For each weekday of symptom onset, we use a different reporting delay distribution. This means, for instance, that an infected individual with symptom onset on Monday may have another reporting pattern and therefore another reporting delay distribution than an individual with symptom onset on Sunday (since several local authorities do not work on Sundays). Figure S6 in the Supporting Information shows the estimated reference distributions from Bavaria. Furthermore, we individualize these weekday-specific reporting delay distributions for each location with its regional reporting pattern (e.g., some countries do not report any cases on Sundays at all while others do). We achieve this by introducing location and weekday-specific parameters β_m^w to adapt these distributions for weekdays w for each geographical region m. These parameters β_m^w are multiplied with the discretized versions of the Bavarian reporting delay distributions to inflate or deflate the probability mass at the respective time spans that match the weekdays. For example, for a country A that does not report any cases at all on Sundays, the respective β_A^{Sunday} would be estimated as zero. Thus, the reporting delay distribution for symptom onset on Mondays would result in a probability mass of zero on the 6th, 13th, and so on, day, whereby the reporting delay distribution for symptom onset on Tuesday would result in a probability mass of zero on the 5th, 12th, and so on, day. After the multiplication, we renormalize the result to obtain a proper probability distribution. As a consequence, we can account for weekly reporting patterns that are specific to each geographical region m.

Following Flaxman et al. (2020), we describe the number of deaths $D_{t,m}$ occurring on day t in geographical region m as a function of the number of true cases with disease onset prior to t through the following *death model*:

$$D_{t,m} \sim NegBinom\left(\mu_{t,m}^D, \phi^D\right),$$

where

$$\mu_{t,m}^D = \pi_{t,m}^D \sum_{u\leq t} C_{u,m}\left(F_{\xi_m^{D,w}}(t-u+1) - F_{\xi_m^{D,w}}(t-u)\right).$$

In this model, $D_{t,m}$ is described by a negative binomial distribution with expected value equal to the sum of the number of true cases with disease onset at time $t-u$, weighted by the probability of dying on the uth day after the onset of symptoms. This latter probability can be obtained by discretizing the probability distribution describing the time until death for patients who died, that is, $F_{\xi_m^{D,w}}$, and multiplying by the infection fatality rate (IFR) $\pi_{t,m}^D$, that is, the probability of dying for an infected individual where this rate can depend on day t and geographical region m. Similarly to $F_{\xi_m^{R,w}}$ in the

reporting model, $F_{\xi_m^{D,w}}$ accounts for weekday effects that can be specific to geographical region m to account for differences in reporting through variables $\beta_m^{D,w}$.

The IFR is a crucial part of the model as we assume that this quantity is fixed and known, but we consider several (potentially time-varying) factors that have an influence on this quantity, namely, the age composition in a country, vaccination rollout, and the prevalence of new variants.

The proposed model operates on an aggregated level with respect to locations m (e.g., countries). However, different locations have different age compositions. As disease severity is strongly correlated with age, we obtain an aggregated IFR by weighting an age-specific IFR with the age structure in each location. We use information on the age strata of each country from O'Driscoll et al. (2021) and the aggregated IFR for four different age groups $s = 1, \ldots, 4$ from Staerk et al. (2021), that is, we use an IFR of 0.008% for the age group of 0–34 years, 0.122% for 35–59 years, 0.992% for 60–79 years and 7.274% for an age of 80 or older. The location-specific IFR is then calculated by

$$ifr_m = \sum_{s=1}^{4} w_{s,m} \cdot ifr_s,$$

where $w_{s,m}$ is the proportion of the age category (stratum) of the population of country m and ifr_s defines the age-specific IFR in each stratum s.

As vaccinations substantially reduce the probability of dying, we make the assumption that the IFR at each location m changes as a function of the time-varying vaccination coverage. With growing coverage in the population, the IFR is lowered. It is important to mention that older age groups in the population were vaccinated with a higher priority at the beginning of the vaccination rollout in most countries. We include the effect of vaccinations by reducing the IFR in the different age strata relative to their share of the population. Most countries do not provide enough information about their vaccination progress in the different age groups. We therefore use publicly available data from France (Santé Publique France, 2021) and extrapolate this information to all other countries, because the majority of European countries used a similar vaccination strategy, making it plausible to assume that the evolution of vaccination coverage over time in the different age groups was roughly comparable across different countries. We assume that after the first vaccination, the probability of dying is reduced by 80%, that is, $\beta_{\pi^D}^{vacc} = 0.8$ after a lag of 2 weeks. For example, Haas et al. (2021) found a higher effectiveness against COVID-19-related deaths of the BNT162b2 vaccine. However, since not all countries use the same type of vaccine and one can assume a reduction of the vaccine effectiveness as time goes on, we decide to use this approximation with a lower effectiveness.

Finally, new mutations of the virus can change disease severity. For COVID-19, we include the effect of alpha (B.1.1.7) and delta (B.1.617.2) in the considered time window. Fisman and Tuite (2021) provide information on the severity of these variants of concern. To account for the fact that these variants changed the overall IFR, we combine the time-varying prevalence of these variants of concern with their disease severity. For B.1.1.7, we inflate the IFR by a factor of $\beta_{\pi^D}^{alpha} = 1.51$ and for B.1.617 with $\beta_{\pi^D}^{delta} = 2.08$.

To test the robustness to the assumptions concerning the overall value of the IFR and the influence of variants of concern and of vaccinations on the IFR, we conduct a number of sensitivity analyses (see Section G and Figures S8 and S9 in the Supporting Information for more details and results). Finally, many countries provide information on hospital and ICU occupancy, and it is important to be able to integrate these two additional sources of information in the estimation of the true and unknown number of infections wherever this is possible. We integrate this information through two *hospitalization models* that have a very similar structure as the death model:

$$H_{t,m} \sim NegBinom\left(\mu_{t,m}^H, \phi^H\right),$$

where

$$\mu_{t,m}^H = \pi_{t,m}^H \sum_{u \leq t} C_{u,m}\left(F_{\xi^H}(t - u + 1) - F_{\xi^H}(t - u)\right),$$

where $\pi_{t,m}^H$ varies over time in the same way as $\pi_{t,m}^D$ to account for vaccination coverage and different virus variants. In contrast to $\pi_{t,m}^D$, however, we can estimate $\pi_{t,m}^H$ for each geographical region and do not have to consider it to be known. By doing so, we can account for differences in medical care and definitions of hospital admissions and ICU admissions that

may vary between geographical regions. To account for vaccination coverage and the influence of virus variants, we define parameters π_m^H that are specific to each geographical region, but that do not change over time, and calculate $\pi_{t,m}^H$ as the product $\pi_m^H \times g_{t,m}$ where $g_{t,m}$ is a fixed quantity representing the effect of vaccinations and virus variants, assuming that these factors modify the severity of the disease in the same time-varying manner as for $\pi_{t,m}^D$. $g_{t,m}$ can therefore constructed from $\pi_{t,m}^D$ as $g_{t,m} = \pi_{t,m}^D / \pi_{1,m}^D$.

We use exactly the same model for hospital (normal beds) and ICU occupancy. For the sake of brevity, we therefore do not present the model for ICU occupancy in detail, but it can be obtained by merely changing the superscripts from H to ICU. The two distributions, $F_{\xi H}$ and $F_{\xi ICU}$, provide the probability that a person with symptom onset on day t_{onset} occupies a hospital or an ICU unit on day $t_{onset} + t_{delay}$ with $t_{delay} = 1, 2, 3, \dots$. For the application to COVID-19, we obtain these two distributions by combining information on the time between symptom onset and hospitalization with information on the time a person occupies a bed or ICU after being hospitalized through Monte Carlo methods. See Sections D.2 of the Supporting Information for a more detailed description of the definition of $F_{\xi H}$ and $F_{\xi ICU}$.

2.2 | Modeling the effects of NPIs and seasons

As mentioned above, we model the number of true and unknown infections as a discrete latent variable. To describe the dynamics of the infectious disease, we assume that this latent variable follows the following *renewal model*:

$$I_{t,m} \sim NegBinom\left(\mu_{t,m}, \phi^I\right),$$

where

$$\mu_{t,m} = R_{t,m} \sum_{u<t} I_{u,m}\left(F_\gamma(t-u+1) - F_\gamma(t-u)\right).$$

This renewal model describes the number of infected individuals $I_{t,m}$ at each time point t in geographical region m as a function of past infections, the instantaneous reproduction number $R_{t,m}$, and the generation time distribution. To be more specific, the expected number of infections $\mu_{t,m}$ is the sum of the previous infections on the $t-1$ days before t weighted by the corresponding probability mass of the discretized generation time distribution $F_\gamma(t-u+1) - F_\gamma(t-u)$ multiplied by the instantaneous reproduction number $R_{t,m}$ at time t in geographical region m, where the generation time distribution $F_\gamma(t-u+1) - F_\gamma(t-u)$ represents the probability to transmit the infection from one infected individual to another between time $t-u$ and $t-u+1$. Applying the renewal equation to past infections yields the current number of infections $I_{t,m}$ (see, e.g., Fraser et al., 2009), and it can be seen as a more flexible version of the disease dynamics described in classical compartmental models for infectious diseases (Wallinga & Lipsitch, 2007). We assume that the latent variable follows a negative binomial distribution. We set a prior on the size parameter with $\tilde{\phi}^I \sim N^+(0, 0.015)$ where $\tilde{\phi}^I = 1/\phi^{I^2}$. Through this prior assumption, the dispersion is pushed toward smaller values (see Section 2.3 for an explanation). The same prior is also used for the size parameter for the observed time series (i.e., $C_{t,m}^R, D_{t,m}, H_{t,m}, ICU_{t,m}$).

We seed the model for the first day $I_{1,m}$ in each geographical region m through a negative binomial distribution with mean parameter τ_m:

$$I_{1,m} \sim NegBinom\left(\tau_m, \phi^I\right), \tag{1}$$

where we assume a hierarchical model for τ_m, that is, each τ_m follows a truncated Normal distribution around a common parameter τ that follows a Gamma distribution with shape a_τ and scale b_τ.

$$\tau_m \sim N^+(\tau, \sigma_\tau),$$

$$\tau \sim Ga(a_\tau, b_\tau),$$

$$\sigma_\tau \sim N^+(\mu_{\sigma_\tau}, \sigma_{\sigma_\tau}).$$

For a graphical display of the prior on $I_{1,m}$ that this hierarchical structure implies, see Figure S2 in the Supporting Information.

Following Flaxman et al. (2020), Brauner et al. (2021), and Sharma et al. (2021), we describe the effect of K NPIs $\alpha_{k,m}$ through the following model:

$$R_{t,m} = R_{t,m}^0 \exp\left(-\sum_{k=1}^{K+3} \alpha_{k,m} \cdot \mathbb{1}_{k,m}(t) \right) \cdot (1 - c_{t,m}^1 - c_{t,m}^2 \cdot (1 - c_{t,m}^1))$$

with

$$c_{t,m}^1 = \frac{\sum_{u<t} I_{u,m}}{N_m} \cdot (1 - \beta^{reinf}),$$

$$c_{t,m}^2 = \frac{\sum_{u<t}(Vacc_{u,m}^1 \cdot \beta^{vacc1} + Vacc_{u,m}^2 \cdot \beta^{vacc2})}{N_m},$$

where $\mathbb{1}_{k,m}(t)$ are indicator variables taking the value of 1 if the kth NPI is active at time t in geographical region m and 0 otherwise. The correction factors $c_{t,m}^1$ and $c_{t,m}^2$ reduce the transmissibility of the virus in the population: $c_{t,m}^1$ corrects for previously infected individuals. Since an infection may not guarantee protection against the infectious agent, we include a parameter β^{reinf} giving the probability of reinfection. The term $c_{t,m}^2$ corrects for vaccination coverage where $\sum_{u<t} Vacc_{u,m}^1$ and $\sum_{u<t} Vacc_{u,m}^2$ are the number of vaccinated individuals in the population at time t and geographical region m and β^{vacc1} and β^{vacc2} represent the probability of infection after a first and second vaccine dose.

Besides NPIs, we also include the effect of seasons (choosing summer as reference category, resulting in $K + 3$ indicator variables in total) where each indicator variable is 1, if the current t corresponds to the according season. Since it is reasonable to assume variations in the effect of NPIs and seasons between different geographical regions, we allow for country-specific effects that are linked through a hierarchical structure. This hierarchical structure makes it possible to share information between regions to infer an overall effect of the NPIs while allowing to estimate individual effects that are specific to each geographical region:

$$\alpha_{k,m} \sim N\left(\alpha_k, \sigma_{\alpha_k}^2\right).$$

The basic reproduction number R_m^0 may vary over time due to the occurrence of different variants that modify the transmissibility of the virus. We propose a convex combination to construct a time-dependent basic reproduction number R_m^0. For the application to COVID-19, we account for two variants of concern yielding the following formula:

$$R_{t,m}^0 = R_m^0 \cdot (1 - p_{t,m}^{alpha} - p_{t,m}^{delta}) +$$
$$(1 + \beta^{alpha}) \cdot R_m^0 \cdot p_{t,m}^{alpha} +$$
$$(1 + \beta^{delta}) \cdot R_m^0 \cdot p_{t,m}^{delta},$$

where $p_{t,m}^{alpha}$ and $p_{t,m}^{delta}$ are the prevalence of the alpha (B.1.1.7) and delta (B.1.617.2) variants, respectively, at each time t in geographical region m. The two unknown parameters β^{alpha} and β^{delta} represent the increased transmissibility of these variants compared to the wild type. We obtain a time variant reproduction number by taking the reproduction number of the original wild type as basis and multiplying it with $(1 + \beta^{alpha})$ and $(1 + \beta^{delta})$ which accounts for the effect of these subsequent variants.

Finally, we allow for variation in the basic reproduction number among geographical regions. We therefore assume reproduction numbers R_m^0 that are specific to geographical region m that are again modeled in a hierarchical manner with common mean R_0:

$$R_m^0 \sim N\left(R^0, \sigma_R^2\right).$$

The proposed model requires a large set of parameters that are either estimable (and possibly with a prior) or have to be specified as a fixed quantity. Table 1 provides an overview of all parameters of the model with their specifications.

TABLE 1 Summary of all parameters used in the model.

Parameters to model disease dynamics			
Parameter	**Description**	**Additional information**	**Prior**
$I_{1,m}$	Initial number of infected individuals	Location-specific (hierarchical)	$NegBinom(\tau_m, \phi^I)$
F_γ	Generation time distribution	$Ga(5, 0.45)$	Fixed
τ_m	Expected mean of $I_{1,m}$	Location-specific (hierarchical)	$N^+(\tau, \sigma_\tau)$
τ	Mean over all τ_m	Shared mean	$Ga(10, 1)$
σ_τ	Variation of τ_m across all m		$N^+(0, 10)$
ϕ^I	Size parameter of infections	Prior on $\tilde{\phi}^I = 1/\phi^{I^2}$	$N^+(0, 0.015)$
$I_{t,m}$	Number of infected individuals at t, m	Location- & time-specific	$NegBinom(\tau_m, \phi)$
R^0_m	Basic reproduction number for each m	Location-specific (hierarchical)	$N(R^0, \sigma_R)$
R^0	Mean R^0 over all m	Shared mean	$N(3.25, 0.05)$
σ_R	Variation of R^0 over all m		$N^+(0, 0.01)$
β^{alpha}	Increased transmissibility of variant	For B1.1.7	$N(0.6, 0.01)$
β^{delta}	Increased transmissibility of variant	For B.1.617.2	$N(1.5, 0.01)$
$p^{alpha}_{t,m}$	Prevalence at time t in location m	For B1.1.7	See supp. info (fixed)
$p^{delta}_{t,m}$	Prevalence at time t in location m	For B.1.617.2	See supp. info (fixed)
$\beta^{vacc1}_{t,m}$	Effectiveness of vaccination	With one dose	0.5 (fixed)
$\beta^{vacc2}_{t,m}$	Effectiveness of vaccination	With two doses	0.35 (fixed)
$\alpha_{k,m}$	Effect of the kth NPI at location m	Location-specific (hierarchical)	$N(\alpha_k, \sigma_{\alpha_k})$
α_k	Mean of the kth NPI over all m	Shared mean	$N(0, 0.3)$
σ_{α_k}	Variation of α_k over all m		$N^+(0, 0.015)$
Parameters to infer the infections			
Parameter	**Description**	**Additional information**	**Prior**
F_{ξ^C}	Incubation time distribution	$Ga(5.68, 0.08)$	Fixed
$F_{\xi^{R,w}_m}$	Reporting delay distribution	See Supp. Info	Fixed
β^w_m	Seasonal reporting at w in m	Weekday-specific	$U(0, 10)$
ϕ^R	Size for reported cases	Prior on $\tilde{\phi}^R = 1/\phi^{R^2}$	$N^+(0, 0.015)$
$\rho_{t,m}$	Reporting ratio at t & m	Piece-wise constant	$U(0, 3)$
$\pi^D_{t,m}$	Infection fatality rate for location m	See Section 2.1	Fixed
ifr_g	IFR for age stratum g	See Section 2.1	Fixed
w_{s_m}	Share of age stratum in location m	See Section 2.1	Fixed
$\beta^{vacc}_{\pi^D}$	Effect of vaccination on IFR		0.8 (fixed)
$\beta^{alpha}_{\pi^D}$	Severity of B.1.1.7		1.51 (fixed)
$\beta^{delta}_{\pi^D}$	Severity of B.1.617.2		2.08 (fixed)
$F_{\xi^{D,w}_m}$	Symptoms-to-death distribution	$Ga(15.93, 0.1)$	Fixed
$\beta^{D,w}_m$	Seasonal reporting at w in m	One for each weekday w	$U(0, 10)$
ϕ^D	Size for reported deaths	Prior on $\tilde{\phi}^D = 1/\phi^{D^2}$	$N^+(0, 0.015)$
π^H	Depends on π^D via $g_{t,m}$		$U(0, 10)$
F_{ξ^H}	Symptoms-to-hospital occupancy	See Supp. Info D.2	Fixed
ϕ^H	Size of hospital occupancy	Prior on $\tilde{\phi}^H = 1/\phi^{H^2}$	$N^+(0, 0.015)$
π^{ICU}	Depends on π^D via $g_{t,m}$		$U(0, 10)$
$F_{\xi^{ICU}}$	Symptoms-to-ICU occupancy	See Supp. Info D.2	Fixed
ϕ^{ICU}	Size of ICU occupancy	Prior on $\tilde{\phi}^{ICU} = 1/\phi^{ICU^2}$	$N^+(0, 0.015)$
$g_{t,m}$	Correction to π^H and π^{ICU}	Derived from $\pi^D_{t,m}$ (Section 2.1)	Fixed

Furthermore, we provide a full directed acyclic graph (Figure 1), a summary of the model, and the full expression of the joint posterior in Section A of the Supporting Information.

2.3 | Inference, identifiability, and implementation

The flexibility of the proposed model can come at the cost of nonidentifiability issues. The first obvious problem of identifiability occurs if we try to estimate the IFR $\pi_{t,m}^D$, the probability of being hospitalized π_m^H or being treated in ICU π_m^{ICU}, and the case detection ratios $\rho_{t,m}$ simultaneously. This problem is easily circumvented by considering one of the four parameters as known. As the IFR can be reliably estimated in seroprevalence studies and modified by accounting for factors like the age structure of the population, vaccination coverage, and the prevalence of different variants, we consider this factor known to be able to estimate the three remaining factors. The second identifiability issue arises in the estimation of the number of true and unknown infections. Since we assume that this variable follows a negative binomial distribution where the expected value is a function of the effects of NPIs (that are to be estimated), the model can in theory describe the data through any set of values for these parameters if the dispersion is high (i.e., the size parameter is small). Moreover, the dimension of the latent variable can be rather high depending on the length of the observation window and number of geographical regions (the dimension in the latent variable grows with t and m). We address this issue by assuming an informative prior for the different size parameters in the renewal model, the hospitalization models, and the death model and by splitting $I_{t,m}$ in blocks of 10 for each geographical region m to be able to update each block one at a time. Finally, assuming a hierarchical model structure on $\alpha_{k,m}, R_m^0$, and τ_m has the advantage of stabilizing parameter estimates by using information across countries. This effect is particularly important for the estimation of NPI. Since such interventions are often implemented or relaxed as multicomponent interventions on the same or subsequent days in a country, it is difficult to disentangle their effects if we assume country-specific effects that do not follow a hierarchical structure because the estimated effects would be highly correlated. Using a hierarchical model allows us to account for variation in the effect of these interventions while using the information across countries to reduce the correlation between effect estimates. However, it is difficult to determine the exact amount of shrinkage that should be applied (expressed through the prior distributions on the variance parameters). The choice needs to be transparently reported and tested in sensitivity analyses.

Due to the complexity of the hierarchical model, there is no analytical solution and we use a Metropolis–Hastings algorithm (Hastings, 1970) to sample from the joint posterior distribution. For the simulation study and the application, we fine-tune acceptance rates by using an adaptive phase (Brooks et al., 2011; Roberts & Rosenthal, 2009) and discard a defined number of iterations as burn-in. We apply thinning to reduce the autocorrelation in the generated Markov chains. For more details on the implementation, we refer to Section E of the Supporting Information.

3 | SIMULATION STUDY

3.1 | Data generation and study design

We carry out a simulation study with the aims (1) to assess the correctness of the implemented algorithm, (2) to investigate potential problems concerning the identifiability of model parameters, and (3) to assess the impact of model misspecification concerning age stratification. We simulate date according to the model with prespecified parameters. Afterwards, we apply the proposed model to the generated data sets and compare the results with the known parameters and the latent variable. We generate 100 data sets with 10 geographical regions and an observation period of 600 days for each of them. Thus, each data set contains 6000 rows of data. We specify five artificial interventions with mean effects $\alpha_1 = 0.22, \alpha_2 = 0.25, \alpha_3 = 0.3, \alpha_4 = 0.4, \alpha_5 = 0.45$. This allows the reproduction number to be reduced by roughly 80% when all NPIs are active. To obtain region-specific effects of NPIs, we sample from a Gaussian distribution with the corresponding mean α_k and a standard deviation $\sigma_{\alpha_k} = 0.01$. The basic reproduction number is sampled in the same way using a mean $R^0 = 3.25$ and a standard deviation of $\sigma_R^2 = 0.1$. We seed the first day of the pandemic in each region by sampling from a negative binomial distribution with a mean τ_m that is generated from a Gaussian distribution with mean $\tau = 10$ and $\sigma_\tau = 2$. All size parameters of the negative binomial distributions are set to 1000 to obtain stable disease dynamics. To obtain realistic time points at which the NPIs are set to active, we generate data in which the decision on whether an NPI is set to active depends on ICU occupancy: To do so, we generate Bernoulli variables for currently inactive NPIs at

TABLE 2 Average estimated effects of nonpharmaceutical interventions (NPIs) on simulated data. All values (except the coverage) are taken as the mean over all simulated data sets and generated Markov chains for all $\alpha_{k,m}$'s.

Intervention	True value	Estimate	Relative bias (%)	Coverage (%)
NPI1	0.22	0.223	1.312	97.4
NPI2	0.25	0.253	1.140	97.0
NPI3	0.30	0.304	1.189	97.2
NPI4	0.40	0.406	1.462	95.0
NPI5	0.45	0.456	1.534	91.7
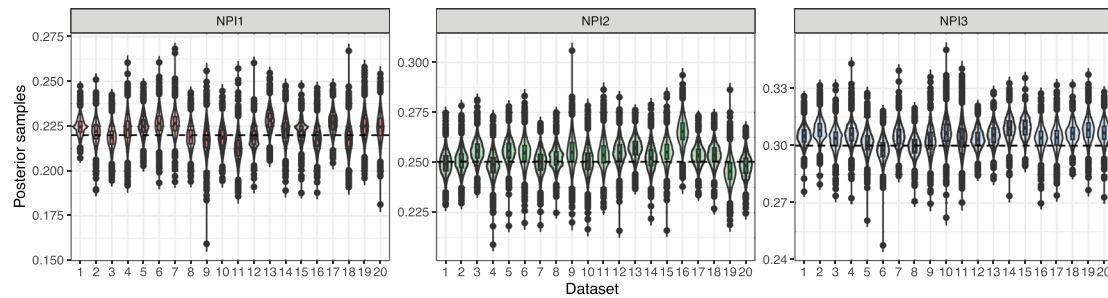
FIGURE 2 Results for the first three NPIs and 20 data sets. The horizontal line is the true mean value.

each t with probability $p_{k,t,m}$ depending on ICU occupancy on $t - 1$. In the case an NPI is activated, it remains active for a random time period between 60 and 120 days.

We carry out a second simulation scenario to test the impact of misspecifying the mixing between different age groups (third aim). The proposed model assumes homogeneous mixing by aggregating all time series over the different age groups and only reflects different age compositions via the IFR. We test the impact of this potential misspecification by simulating age-stratified data with diffusion between the age groups and fit the model on the aggregated data. In Section F of the Supporting Information, we provide further details on how the diffusion between age groups is performed. Table S1 and Figure S7 in the Supporting Information shows that aggregating over age strata has only a negligible impact on the estimation of NPI effects.

The data generation is carried out in R version 4.0.4 (R Core Team, 2021). For further details, see the provided R scripts that we used for the data generation.
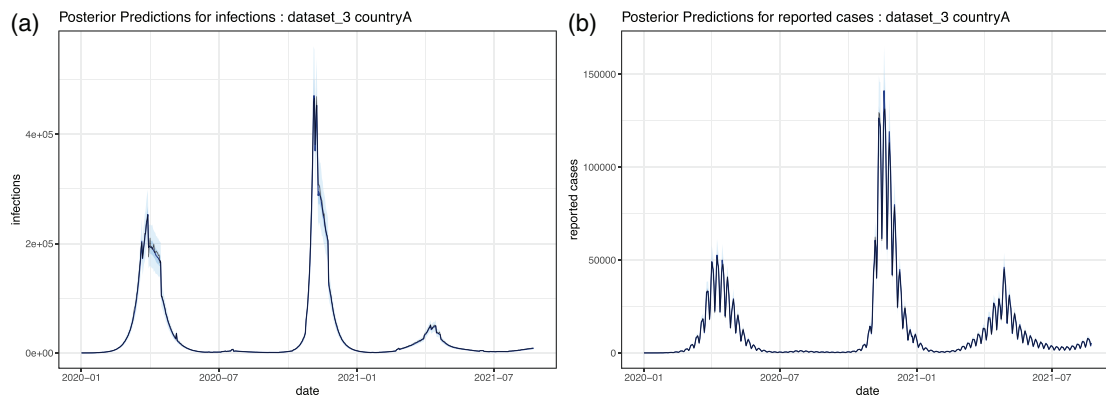
3.2 | Results on simulated data

We fit the model to each of the 100 data sets where we run two chains with 100,000 iterations and a burn-in of 50,000. We apply thinning by keeping only every 60th iteration. We check convergence by analyzing traceplots and potential scale reduction factors that are always < 1.01 (Gelman & Rubin, 1992). As can be seen in Table 2, the algorithm produces estimates that are very close to the true NPI effects (with a mean relative bias of at most 1.534%) and very high coverage rates. For illustration purposes, we present in Figure 2 the samples from the posterior as violin plot for the first three NPIs and 20 data sets. Figure 3 shows the posterior predictions of the number of (unknown) daily infections (A) and reported cases (B) for one of the 10 regions for one data set. For the simulated data, the model fits very well with a low uncertainty. The results for the missspecified model can be found in Table S1 and Figure S7 of the Supporting Information.

4 | CASE STUDY: MODELING COVID-19 IN 20 EUROPEAN COUNTRIES

4.1 | Data sources

In our case study on COVID-19, we analyze data from 20 European countries (Austria, Belgium, Czechia, Denmark, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Netherlands, Norway, Poland, Portugal, Slovenia, Spain, Sweden,

FIGURE 3 Posterior predictions for two time series on simulated data. The (unobserved) infections are shown in A and reported cases in B. Black encodes the true underlying simulated time series. The blue color represents the mean predictions with 95% credible interval.

Switzerland, United Kingdom). Following Flaxman et al. (2020), we define the start of the observation period in each country as 30 days before 10 cumulated deaths were reported. We include data on the entire course of the pandemic until the 31st of October 2021 resulting in a median length of 620 days. We use data on reported cases and deaths from the *Johns Hopkins CSSE COVID-19 Dataset* (Dong et al., 2020). Data on the prevalence of variants of concern and hospital and ICU occupancy are obtained from the *European Centre for Disease Prevention and Control* (European Centre for Disease Prevention and Control, 2022), except for hospital and ICU occupancy for the United Kingdom, which is obtained from the *COVID-19 in the UK dashboard* provided by the UK Health Security Agency (UK Health Security Agency, 2022). As the data on the prevalence of different variants are only available on a weekly basis, we fit a sigmoid function with a squared loss to obtain smooth daily data. More details on this procedure and the resulting time series are presented in Section C.2 with Figures S4 and S5 of the Supporting Information. Data on vaccinations are obtained from *Our World in Data* (Mathieu et al., 2021). Since we use a weighted IFR by age strata, we need information on the number of vaccinations in different age groups. However, very few countries provide information on the age structure of currently vaccinated individuals. We therefore use publicly available data from France and map the relative age-specific vaccination progress to other countries, making the assumption that the prioritization of vaccinations for different age groups evolved roughly in the same manner across different European countries (see Section 2.1). We define the following interventions using information from the *COVID-19 Government Response Tracker* (OxCGRT; Hale et al., 2021) resulting in five NPIs: school closure, gatherings, lockdown, subsequent lockdown, and general behavioral changes. The NPI "school closure" is active when at least some levels of schools and universities (e.g., just high schools) are required to close, and "gatherings" captures the restriction of gatherings to 10 or fewer people. We use two different NPIs depending on whether it was forbidden to leave the house (with possible exceptions such as grocery shopping, and "essential" trips) for the first time ("lockdown") or further times ("subsequent lockdown") because subsequent lockdowns often followed a much more detailed protocol. The last NPI "general behavioral changes" is active from the first time an NPI was implemented in a country and remains active until the end of the observation period. It subsumes many behavioral adaptations that were taken since the beginning of the pandemic and that were respected by a large part of the population in many countries until the end of 2021. These include, for instance, restricting physical contact, working from home wherever possible, higher alertness in case of any respiratory disease symptoms, and the wearing of face masks in some countries. We give a more detailed overview of how we derived these NPIs with the OxCGRT variable coding and the resulting time series (Figure S3) in Section C.1 of the Supporting Information.

4.2 | Challenges in the analysis of the observed time-series

Figure 4 illustrates the challenges in the estimation of the number of daily new infections in a given country by showing the number of reported cases, hospital and ICU occupancy and deaths in the United Kingdom and various factors that have an influence on disease transmission and severity. While the four observed time series show a similar trend during the time between October 2020 and March 2021, they provide rather contradictory information in early-2020 and in late-2021. In particular, the growth rates for the four time series are very different for specific time points (see, e.g., the steep
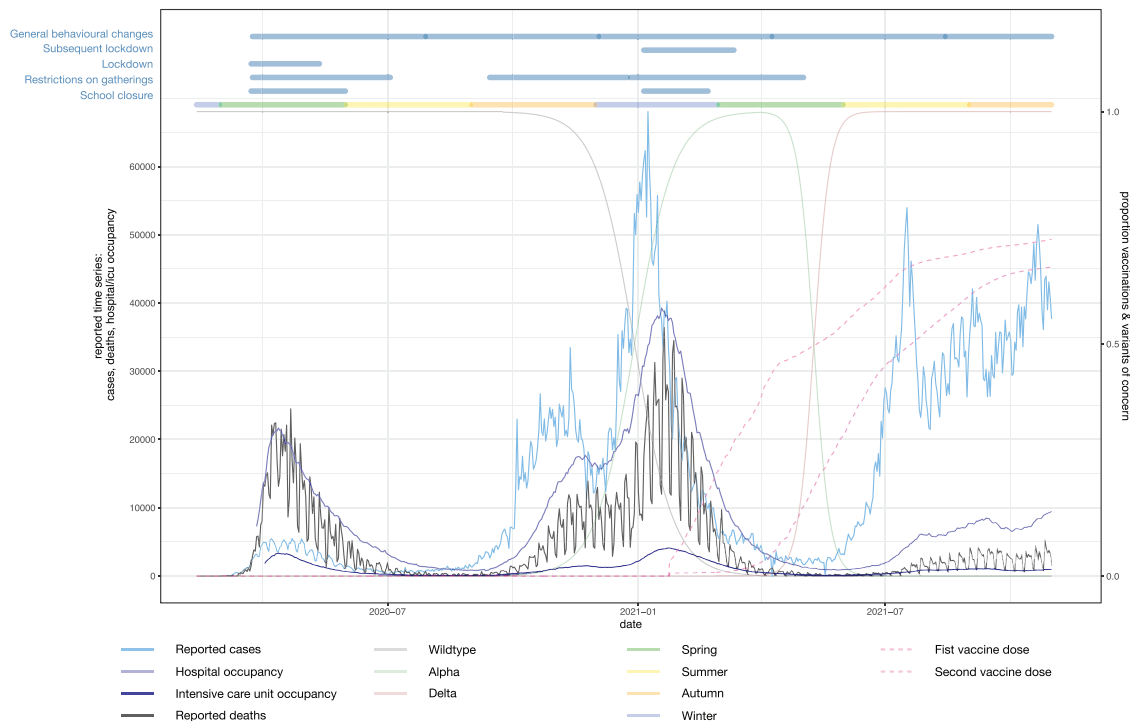
FIGURE 4 Reported cases, hospital occupancy, intensive care unit occupancy, and reported deaths (multiplied by a factor of 20) in the United Kingdom between early-2020 and late-2021. The four observed time series are influenced by the set of nonpharmaceutical interventions that were active at each time point (shown at the top), the season (shown just below), and the number of persons having received a first and second vaccine dose and the prevalence of different virus variants.

increase in reported deaths and hospital occupancy in the early-2020 vs. the more gradual increase in reported cases or the very steep increase in reported cases in July 2021 and the comparably gradual increase in reported deaths). In early-2020, it is obvious that, as in many other countries, only a small proportion of cases were reported because of limited testing capacities. In late-2021, on the other hand, previous infections and vaccinations are likely to have led to fewer severe cases of infections in the population. As a consequence, the numbers of reported deaths and hospital and ICU occupancy are low compared to the number of reported cases. Moreover, the reported time series are not only influenced by the set of NPIs that is active at each time point, but also by the current season with higher infections observed in autumn and winter than in spring and summer and by the prevalence of different virus variants that influence both the transmissibility of the virus and the severity of the disease. Overly simplistic analyses of these time series that only focus on a single indicator of disease transmission and that ignore one or several of the various influencing factors and the weekly patterns in reported cases and deaths may obtain very different answers concerning the same research question, leading to contradictory results that are difficult to communicate to the general public and decision-makers.

4.3 | Results

We run eight chains with a burn-in of 20,000 followed by 50,000 iterations per chain. We apply a thin of 100 resulting in 4000 (i.e., 500×8) samples from the posterior distribution for each parameter. We run a longer adaptive phase with 200 adaptive steps (each with 100 iterations) to get good initial proposal standard deviations. For the final sampling procedure, we again fine-tune these proposals by running 10 adaptive phases (with 50 iterations each). Information about the convergence diagnostic for the parameters of major interest (NPIs and seasonal effects) and further results are presented in Section E of the Supporting Information.

Estimated effects of NPIs Figure 5 provides information on the estimated relative reduction in the reproduction number for NPIs and seasons, respectively. For NPIs, the smallest effect is "school closure" with a credibility interval that includes zero. The most effective NPI is "general behavioral changes," which we defined with the aim to capture several protective measures that were respected by a large portion of the population between the beginning of 2020 and the end
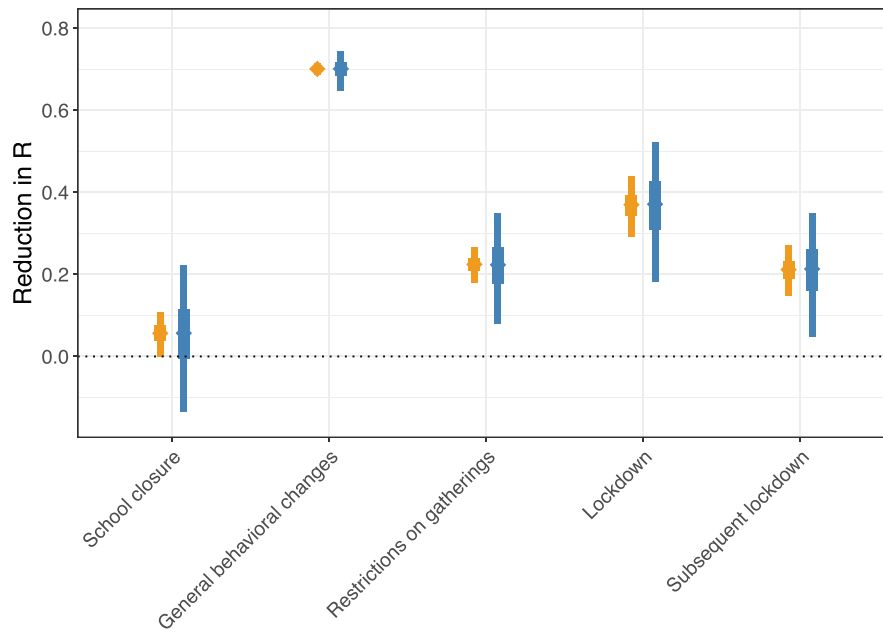
FIGURE 5 Reduction in the reproduction number for NPIs estimated across 20 European countries. Posterior distributions for the mean effects α_k are given in orange. Posterior predictive distribution for $\alpha_{k,m}$ reflecting effect heterogeneity across countries is shown in blue. They are obtained by sampling from a normal distribution with mean α_k and standard deviation σ_{α_k} for each iteration. The 50% and 95% credible intervals are given as bold and normal lines, respectively.

of 2021 including, for instance, working from home wherever possible, higher alertness in case of any respiratory disease symptoms, complying with hygiene recommendations, social distancing, and the wearing of face masks in some countries.

When comparing the effects for the first lockdown with one or several subsequent lockdowns, we can see that the first lockdown is estimated to have a larger effect than subsequent lockdowns, reflecting the fact that the first lockdown was characterized by stronger restrictions and probably better adherence to these than subsequent lockdowns. Figure 6 shows
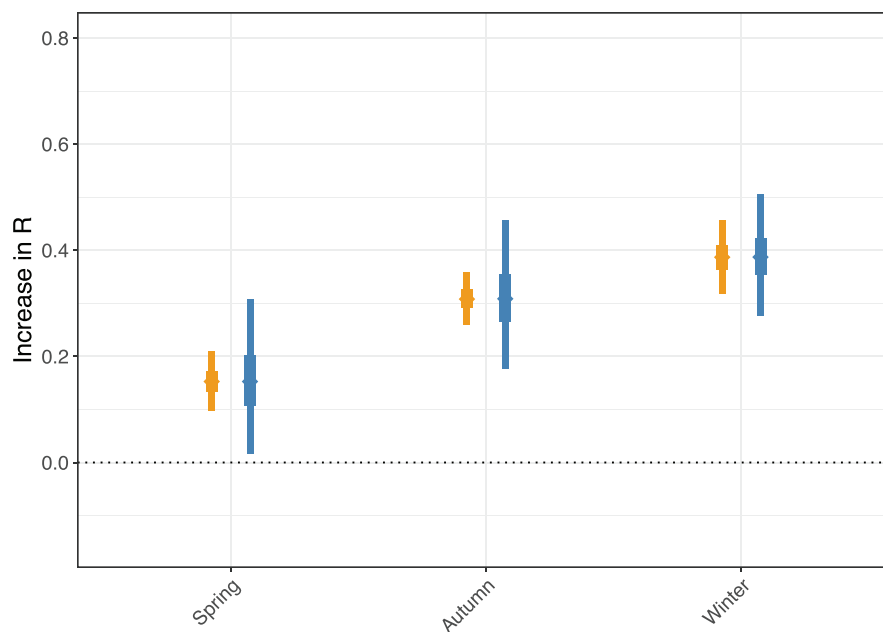
FIGURE 6 Increase in the reproduction number for seasons estimated across 20 European countries. Posterior distributions for the mean effects are given in orange. Posterior predictive distributions are shown in blue. The 50% and 95% credible intervals are given as bold and normal lines, respectively.
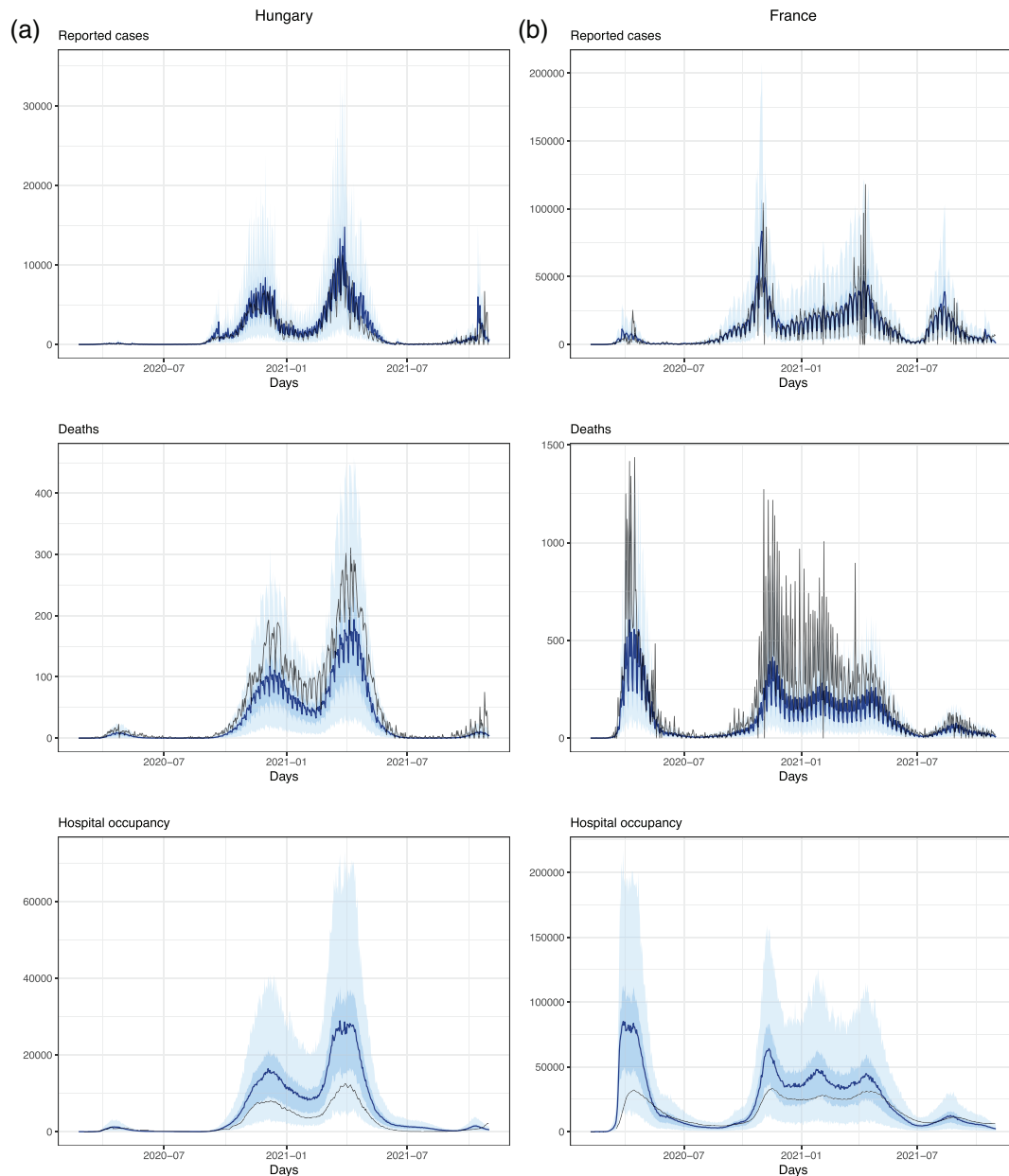
FIGURE 7 Posterior predictions of the reported cases for two countries, Hungary (A) and France (B). The observed time series are given in black and the estimated mean with 50% and 95% credible intervals are shown in blue.

the results for the seasons. As expected, one can observe a strong seasonal influence with an estimated increase in the reproduction number of about 14%, 30%, and 37% for spring, autumn, and winter, respectively.

Model fit and case detection ratios Figure 7 shows posterior predictive checks comparing the observed time series and the posterior predictions for reported cases, hospital occupancy, and deaths for two selected countries, Hungary and France. We chose these two countries, because they represent very different geographical regions in Europe, they differ in their size, they present very different disease dynamics, and data on hospital occupancy were available for both countries. The approach captures the weekly variation in reported cases and deaths that are specific to the two countries. Moreover, it is capable of reproducing the three complementary time series, even though they provide quite contrasting information, in particular, for the first wave. The model shows a tendency to overestimate hospital occupancy during the peaks of the first and second waves of the pandemic for many countries, including Hungary and France. This overestimation might be linked to increases in hospital mortality during the peaks of the first and second waves that are well documented for several countries and have been linked to increasing strain on services that may have led to changes in the case-mix and illness severity of admitted patients (Docherty et al., 2021; Gray et al., 2022; Jassat et al., 2021)
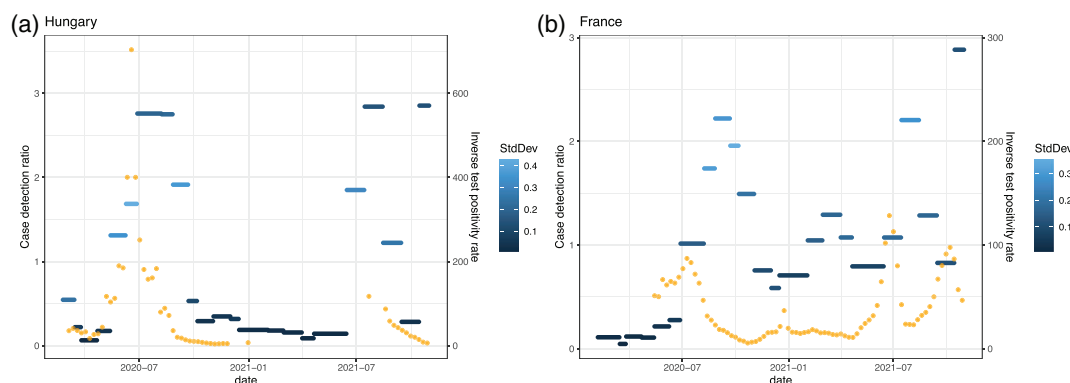
FIGURE 8 Estimated case detection ratios between early-2020 and late-2021 for Hungary (A) and France (B). The shades of blue represent the standard deviation of the posterior with light blue indicating more uncertainty in the estimation. The inverse test positivity rate, which can be interpreted as the number of tests that are performed to detect a case, is given in orange.

Integrating information on reported cases, hospital and ICU occupancy and deaths while accounting for time-varying underreporting in the number of reported cases allow us to estimate variations in case detection ratios that occurred over time. Figure 8 compares these estimated case detection ratios for Hungary and France between early-2020 and late-2021 with the inverse test positivity rate, that is, the number of tests that are performed to obtain a positive test. In general, we observe high underreporting (i.e., very small detection ratios) for the first wave of the pandemic indicating that the true number of infections by far exceeded the reported number of cases. Subsequently, the case detection ratios increased during the summer months and even reached values of over 100%, that is, there were more cases being reported than there were estimated true infections. This might be explained by the fact that the prevalence of the virus was very low during this period and the number of performed tests was very high. In this situation, there may be a nonnegligible proportion of false positive results and we can therefore expect the number of reported cases to exceed the number of true infections due to the imperfect specificity of the tests (Bisoffi et al., 2020; Brownstein & Chen, 2021; Cohen & Kessel, 2020). However, these results must be interpreted with caution as the estimated case detection ratios critically depend on the assumed value of the IFR and on assumptions about how this case fatality rate changes as a function of virus variants and vaccination coverage. In Section H of the Supporting Information, we present in Figure S10 the individual NPI estimations for each country, in Figure S11 the estimated overcontagiousness of the variants of concern, in Figures S12–S14 the posterior predictions for the observed time series, in Figure S15 the estimated infections (latent variable), in Figure S16 the estimated case detection ratios, and in Figure S17 the trace plots for the mean NPIs.

5 | DISCUSSION

We presented a Bayesian hierarchical approach for the modeling of infectious diseases that allows to integrate information on the number of reported cases, hospital and ICU occupancy, and deaths in the estimation of the number of daily new infections. As mentioned in Section 1, previous studies have used various modeling approaches to assess the effect of NPIs on COVID-19 transmission, hospitalizations, and deaths (Banholzer et al., 2021; Brauner et al., 2021; Dehning et al., 2020; Flaxman et al., 2020; Islam et al., 2020; Li et al., 2021a; Sharma et al., 2021; Unwin et al., 2020). Some of these report different findings related to, for example, the magnitude of effect for a specific NPI, as well as the ordering of the relative effectiveness of multiple NPIs. There are numerous other such studies, and systematically identifying and reviewing each of them to compare their findings with those of our study is beyond the scope of this study. Indeed, the principal aim of our study was to show how many of the shortcomings of previous modeling approaches can be overcome by adopting a Bayesian hierarchical approach. Owing to its modular nature, it is possible to model the dynamics of infectious diseases while allowing proper statistical inference and an evaluation of the fit to the observed time series. Moreover, we can integrate the available information while accounting for various sources of uncertainty in this information. By explicitly accounting for time-varying underreporting, seasonality, the spread of different virus variants, vaccination coverage, and previous infections, it is possible to use information over long time periods rather than focusing on short time periods during which these factors remain roughly constant. Using this approach allows for the transparent reporting of model and parameter assumptions and is very flexible: It is thus straightforward to adapt the model to account for additional factors

that might have an influence on disease dynamics. In contrast to most other modeling approaches, our approach explicitly accounts for weekly patterns in the reporting delay distribution for reported cases and deaths, and it is therefore not necessary to smooth these time series. The explicit estimation of new infections on a daily basis is a major advantage if one is interested in the effect of influencing factors that may show variations on a daily scale, for instance, weather conditions (Tosepu et al., 2020), air pollution (Cole et al., 2020), or pollen (Damialis et al., 2021). Due to the flexible combination of different submodels, it would also be straightforward to integrate further information, for example, on the number of tests, on measurements from wastewater, from seroprevalence surveys, or from randomized surveillance testing. Nicholson et al. (2022) combine the latter information with targeted test counts using a stochastic SIR model on weekly aggregated data to obtain fine-scale spatiotemporal prevalence estimates for the United Kingdom.

While the flexible modeling of the observed time series allows to account for different sources of uncertainty, it also comes at the cost of making a number of model and parameter assumptions. Since our modeling approach implicitly gives more weight to reported deaths and ICU and hospital occupancy than to reported cases (by compensating deviations between the true and unknown number of infections and reported cases by differences in case detection ratios for time periods at which the testing strategy changed), our assumptions on how disease severity (and therefore the IFR) is influenced by vaccinations and virus variants play a crucial role in the model. To test the robustness to these assumptions, we conducted extensive sensitivity analyses in which we assessed the degree to which NPI estimates are influenced by variations in the assumptions concerning the overall IFR value and the effect of vaccinations and variants of concern on the IFR. As can be seen in Section G of the Supporting Information, variations in the assumptions concerning these factors has a negligible effect on the estimates of NPI effects.

In our application to COVID-19, we only accounted for the wild type, the alpha, and the delta variant. In principle, it would also be possible to account for the omicron (B.1.1.529) variant using the proposed approach, but there is evidence that this new variant did not only increase the transmissibility of the virus and decrease the severity of the disease, but also entailed changes in the generation time distribution and vaccine efficacy. As a consequence, accounting for omicron would have required a great number of additional assumptions, and it was not in the scope of this work to find reliable information to be able to make all these additional assumptions.

Despite evidence on the importance of asymptomatic infections in the transmission of COVID-19, we did not explicitly distinguish symptomatic and asymptomatic cases in our case study concerning COVID-19. Indeed, it is not clear whether this distinction would necessarily improve the model. This distinction is typically neither made by health authorities in the reporting of cases nor in seroprevalence studies when estimating IFRs. Distinguishing symptomatic and asymptomatic infections would therefore require additional assumptions, in particular, on IFRs that apply only to symptomatic cases, without a clear benefit concerning the insights that we gain from the observable quantities.

Similarly, we assume homogeneous mixing between the different age groups and do not account for age stratification in our model, but it is not clear whether accounting for age stratification would improve our estimates. Indeed, accounting for age stratification would require a great number of additional assumptions, including on the interaction patterns between the age strata in the different countries and age-specific information on the number of reported cases, hospital occupancy, ICU occupancy, and deaths, and this information is only available for a small proportion of the countries that we considered in our application to COVID-19. Even if we had reliable information on mixing patterns between different age groups and age-specific time series, it is not clear whether ignoring these age groups will strongly affect the estimation of NPI effects. In accordance, the results on simulated data presented in Section F of the Supporting Information show that violations of the homogeneous mixing assumption only have a minor influence on NPI estimates. While ignoring age stratification might in general not have a large impact on our NPI estimates, it may lead to an underestimation of the effect of school closures: Since our model relies more on reported deaths and ICU and hospital occupancy than on reported cases, it might not be able to detect an increase or decrease in the number of new infections among children because disease severity in this group is very low. While this reasoning is consistent with some empirical evidence (Fukumoto et al., 2021), others have found differing findings. Studies using various modeling approaches, for example, have reported meaningful decreases in transmission, hospitalization, and deaths due to school closures (Haug et al., 2020; Li et al., 2021a; Liu et al., 2021). Similarly, an overview of systematic reviews, which included and described mainly observational studies, also found that most systematic reviews reported benefits of school closures (Talic et al., 2021). However, each of these studies, as well as many of the underlying studies included in the systematic review, emphasize concerns related to their internal and external validity. Indeed, our model is designed to improve upon multiple assumptions made and approaches taken by such evidence. Nevertheless, our finding that school closures only have a negligible impact on disease dynamics has to be interpreted with caution.

In many countries, the question of whether NPIs should be implemented as a function of reported cases or hospital occupancy was widely debated as both quantities are to some extent unreliable. The proposed Bayesian hierarchical approach provides a framework in which information on both quantities (and on reported deaths and ICU occupancy) can be integrated to predict epidemic development and health care demand in the near future to be able to weigh costs, benefits, and uncertainties in a more robust manner in evidence-informed decision making. While we have developed the approach with reference to COVID-19, the model could easily be adapted to any other known or presently unknown infectious agent.

CONFLICT OF INTEREST STATEMENT
The authors have declared no conflict of interest

DATA AVAILABILITY STATEMENT
We provide the code of the Bayesian hierarchical model along with all scripts to generate the data sets for the simulation study and application on 20 European countries. Also, we provide the final processed data sets (simulated and application) directly to run the model and reproduce all results. The code is available at https://github.com/RaphaelRe/BayesModelCOVID or in achieved form at Rehms (2023).

OPEN RESEARCH BADGES
This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID
Raphael Rehms https://orcid.org/0000-0002-3996-0194
Ulrich Mansmann https://orcid.org/0000-0002-9955-8906

REFERENCES
Banholzer, N., Van Weenen, E., Lison, A., Cenedese, A., Seeliger, A., Kratzwald, B., Tschernutter, D., Salles, J. P., Bottrighi, P., Lehtinen, S., Feuerriegel, S., & Vach, W. (2021). Estimating the effects of non-pharmaceutical interventions on the number of new infections with Covid-19 during the first epidemic wave. *PLoS One*, *16*(6), e0252827.

Bisoffi, Z., Pomari, E., Deiana, M., Piubelli, C., Ronzoni, N., Beltrame, A., Bertoli, G., Riccardi, N., Perandin, F., Formenti, F., Gobbi, F., Buonfrate, D., & Silva, R. (2020). Sensitivity, specificity and predictive values of molecular and serological tests for Covid-19: A longitudinal study in emergency room. *Diagnostics*, *10*(9), 669.

Brauner, J. M., Mindermann, S., Sharma, M., Johnston, D., Salvatier, J., Gavenčiak, T., Stephenson, A. B., Leech, G., Altman, G., Mikulik, V., Norman, A. J., Monrad, J. T., Besiroglu, T., Ge, H., Hartwick, M. A., Teh, Y. W., Chindelevitch, L., Gal, Y., & Kulveit, J. (2021). Inferring the effectiveness of government interventions against Covid-19. *Science*, *371*(6531), eabd9338.

Brooks, S., Gelman, A., Jones, G., & Meng, X.-L. (2011). *Handbook of Markov chain Monte Carlo*. CRC Press.

Brownstein, N. C., & Chen, Y. A. (2021). Predictive values, uncertainty, and interpretation of serology tests for the novel coronavirus. *Scientific Reports*, *11*(1), 1–12.

Cohen, A. N., & Kessel, B. (2020). False positives in reverse transcription pcr testing for sars-cov-2. *medRxiv*.

Cole, M. A., Ozgen, C., & Strobl, E. (2020). Air pollution exposure and Covid-19 in Dutch municipalities. *Environmental and Resource Economics*, *76*(4), 581–610.

Damialis, A., Gilles, S., Sofiev, M., Sofieva, V., Kolek, F., Bayr, D., Plaza, M. P., Leier-Wirtz, V., Kaschuba, S., Ziska, L. H., Bielory, L., Makra, L., Del Mar Trigo, M., & Traidl-Hoffmann, C., COVID-19/POLLEN study group. (2021). Higher airborne pollen concentrations correlated with increased sars-cov-2 infection rates, as evidenced from 31 countries across the globe. *Proceedings of the National Academy of Sciences*, *118*(12), e2019034118.

Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., & Priesemann, V. (2020). Inferring change points in the spread of Covid-19 reveals the effectiveness of interventions. *Science*, *369*(6500), eabb9789.

Docherty, A. B., Mulholland, R. H., Lone, N. I., Cheyne, C. P., De Angelis, D., Diaz-Ordaz, K., Donegan, C., Drake, T. M., Dunning, J., Funk, S., García-Fiñana, M., Girvan, M., Hardwick, H. E., Harrison, J., Ho, A., Hughes, D. M., Keogh, R. H., Kirwan, P. D., Leeming, G., … ISARIC4C Investigators. (2021). Changes in in-hospital mortality in the first wave of Covid-19: A multicentre prospective observational cohort study using the who clinical characterisation protocol UK. *The Lancet Respiratory Medicine*, *9*(7), 773–785.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track Covid-19 in real time. *The Lancet infectious diseases*, *20*(5), 533–534.

European Centre for Disease Prevention and Control. (2022). Covid-19. https://www.ecdc.europa.eu/en/covid-19

Fisman, D. N., & Tuite, A. R. (2021). Evaluation of the relative virulence of novel sars-cov-2 variants: A retrospective cohort study in Ontario, Canada. *CMAJ: Canadian Medical Association Journal*, *193*(42), E1619.

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., Vollmer, M. A. C., Ferguson, N. M., Okell, L. C., & Bhatt, S., Imperial College COVID-19 Response Team. (2020). Estimating the effects of non-pharmaceutical interventions on Covid-19 in Europe. *Nature*, *584*(7820), 257–261.

Fraser, C., Donnelly, C. A., Cauchemez, S., Hanage, W. P., Van Kerkhove, M. D., Hollingsworth, T. D., Griffin, J., Baggaley, R. F., Jenkins, H. E., Lyons, E. J., Jombart, T., Hinsley, W. R., Grassly, N. C., Balloux, F., Ghani, A. C., Ferguson, N. M., Rambaut, A., Pybus, O. G., Lopez-Gatell, H., … WHO Rapid Pandemic Assessment Collaboration. (2009). Pandemic potential of a strain of influenza a (h1n1): Early findings. *Science*, *324*(5934), 1557–1561.

Fukumoto, K., McClean, C. T., & Nakagawa, K. (2021). No causal effect of school closures in Japan on the spread of Covid-19 in Spring 2020. *Nature Medicine*, *27*(12), 2111–2119.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.

Gray, W. K., Navaratnam, A. V., Day, J., Wendon, J., & Briggs, T. W. (2022). Covid-19 hospital activity and in-hospital mortality during the first and second waves of the pandemic in England: An observational study. *Thorax*, *77*(11), 1113–1120.

Haas, E. J., Angulo, F. J., McLaughlin, J. M., Anis, E., Singer, S. R., Khan, F., Brooks, N., Smaja, M., Mircus, G., Pan, K., Southern, J., Swerdlow, D. L., Jodar, L., Levy, Y., & Alroy-Preis, S. (2021). Impact and effectiveness of mrna bnt162b2 vaccine against sars-cov-2 infections and Covid-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: An observational study using national surveillance data. *The Lancet*, *397*(10287), 1819–1829.

Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., & Tatlow, H. (2021). A global panel database of pandemic policies (Oxford Covid-19 government response tracker). *Nature Human Behaviour*, *5*(4), 529–538.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, *57*(1), 97–109.

Haug, N., Geyrhofer, L., Londei, A., Dervic, E., Desvars-Larrive, A., Loreto, V., Pinior, B., Thurner, S., & Klimek, P. (2020). Ranking the effectiveness of worldwide Covid-19 government interventions. *Nature Human Behaviour*, *4*(12), 1303–1312.

Höhle, M., & an der Heiden, M. (2014). Bayesian nowcasting during the stec o104: H4 outbreak in Germany, 2011. *Biometrics*, *70*(4), 993–1002.

Islam, N., Sharp, S. J., Chowell, G., Shabnam, S., Kawachi, I., Lacey, B., Massaro, J. M., D'Agostino, R. B., & White, M. (2020). Physical distancing interventions and incidence of coronavirus disease 2019: Natural experiment in 149 countries. *BMJ*, *370*, m2743.

Jassat, W., Mudara, C., Ozougwu, L., Tempia, S., Blumberg, L., Davies, M.-A., Pillay, Y., Carter, T., Morewane, R., Wolmarans, M., von Gottberg, A., Bhiman, J. N., Walaza, S., Cohen, C., & DATCOV author group. (2021). Difference in mortality among individuals admitted to hospital with Covid-19 during the first and second waves in South Africa: A cohort study. *The Lancet Global Health*, *9*(9), e1216–e1225.

Li, Y., Campbell, H., Kulkarni, D., Harpur, A., Nundy, M., Wang, X., & Nair, H. (2021a). The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (r) of sars-cov-2: A modelling study across 131 countries. *The Lancet Infectious Diseases*, *21*(2), 193–202.

Li, Y., Campbell, H., Kulkarni, D., Harpur, A., Nundy, M., Wang, X., Nair, H., & Usher Network for COVID. (2021b). The temporal association of introducing and lifting non-pharmaceutical interventions with the time-varying reproduction number (r) of sars-cov-2: A modelling study across 131 countries. *The Lancet Infectious Diseases*, *21*(2), 193–202.

Liu, Y., Morgenstern, C., Kelly, J., Lowe, R., & Jit, M. (2021). The impact of non-pharmaceutical interventions on sars-cov-2 transmission across 130 countries and territories. *BMC Medicine*, *19*(1), 1–12.

Mathieu, E., Ritchie, H., Ortiz-Ospina, E., Roser, M., Hasell, J., Appel, C., Giattino, C., & Rodés-Guirao, L. (2021). A global database of Covid-19 vaccinations. *Nature Human Behaviour*, *5*, 947.

May, T. (2020). Lockdown-type measures look effective against Covid-19. *BMJ*, *370*, m2809.

Nicholson, G., Lehmann, B., Padellini, T., Pouwels, K. B., Jersakova, R., Lomax, J., King, R. E., Mallon, A.-M., Diggle, P. J., Richardson, S., Blangiardo, M., & Holmes, C. (2022). Improving local prevalence estimates of sars-cov-2 infections using a causal debiasing framework. *Nature Microbiology*, *7*(1), 97–107.

O'Driscoll, M., Dos Santos, G. R., Wang, L., Cummings, D. A., Azman, A. S., Paireau, J., Fontanet, A., Cauchemez, S., & Salje, H. (2021). Age-specific mortality and immunity patterns of sars-cov-2. *Nature*, *590*(7844), 140–145.

Pullano, G., Di Domenico, L., Sabbatini, C. E., Valdano, E., Turbelin, C., Debin, M., Guerrisi, C., Kengne-Kuetche, C., Souty, C., Hanslik, T., Blanchon, T., Boëlle, P. Y., Figoni, J., Vaux, S., Campèse, C., Bernard-Stoecklin, S., & Colizza, V. (2021). Underdetection of cases of Covid-19 in france threatens epidemic control. *Nature*, *590*(7844), 134–139.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Rehms, R. (2023). Raphaelre/bayesmodelcovid: Release. https://doi.org/10.5281/zenodo.8047631

Roberts, G. O., & Rosenthal, J. S. (2009). Examples of adaptive mcmc. *Journal of Computational and Graphical Statistics*, *18*(2), 349–367.

Santé Publique France. (2021). Données relatives aux personnes vaccinées contre la covid-19. https://www.data.gouv.fr/fr/datasets/donnees-relatives-aux-personnes-vaccinees-contre-la-covid-19-1/

Sharma, M., Mindermann, S., Rogers-Smith, C., Leech, G., Snodin, B., Ahuja, J., Sandbrink, J. B., Monrad, J. T., Altman, G., Dhaliwal, G., Finnveden, L., Norman, A. J., Oehm, S. B., Sandkühler, J. F., Aitchison, L., Gavenciak, T., Mellan, T., Kulveit, J., Chindelevitch, L., … Brauner, J. M. (2021). Understanding the effectiveness of government interventions against the resurgence of Covid-19 in Europe. *Nature Communications*, *12*(1), 1–13.

Staerk, C., Wistuba, T., & Mayr, A. (2021). Estimating effective infection fatality rates during the course of the Covid-19 pandemic in Germany. *BMC Public Health*, *21*(1), 1–9.

Talic, S., Shah, S., Wild, H., Gasevic, D., Maharaj, A., Ademi, Z., Li, X., Xu, W., Mesa-Eguiagaray, I., Rostron, J., Theodoratou, E., Zhang, X., Motee, A., Liew, D., & Ilic, D. (2021). Effectiveness of public health measures in reducing the incidence of Covid-19, sars-cov-2 transmission, and Covid-19 mortality: Systematic review and meta-analysis. *BMJ*, *375*, e068302.

Tosepu, R., Gunawan, J., Effendy, D. S., Lestari, H., Bahar, H., Asfian, P., & Ahmad, O. A. I. (2020). Correlation between weather and Covid-19 pandemic in Jakarta, Indonesia. *Science of the Total Environment*, *725*, 138436.

UK Health Security Agency. (2022). Coronavirus (Covid-19) in the UK dashboard. https://coronavirus.data.gov.uk/

Unwin, H. J. T., Mishra, S., Bradley, V. C., Gandy, A., Mellan, T. A., Coupland, H., Ish-Horowicz, J., Vollmer, M. A., Whittaker, C., Filippi, S. L., Xi, X., Monod, M., Ratmann, O., Hutchinson, M., Valka, F., Zhu, H., Hawryluk, I., Milton, P., Ainslie, K. E. C., … Flaxman, S. (2020). State-level tracking of Covid-19 in the United States. *Nature Communications*, *11*(1), 1–9.

Wallinga, J., & Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1609), 599–604.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.