BMC Public Health

RESEARCH

Open Access

Capturing the spatiotemporal spread of COVID-19 in 30 European countries during 2020 – 2022

Thi Huyen Trang Nguyen[1*], Niel Hens[1,2] and Christel Faes[1]

Abstract

Background While the COVID-19 pandemic has been burdensome globally, it has fostered extensive data collection at various spatiotemporal resolutions. These data heightened researchers' interest in investigating multiple facets of the pandemic. In Europe, key factors shaping disease transmission vary among countries, leading to a gap in understanding how the epidemic evolved and spread across countries as a whole. We endeavor to understand the similarities and differences in the spatiotemporal spread of the COVID-19 pandemic across 27 European Union (EU) countries and 3 European Economic Area (EEA) countries between March 2020 and December 2022.

Method We utilized a multivariate endemic-epidemic model to conduct a space-time analysis across 30 countries, using weekly aggregated COVID-19 case counts from week 13-2020 to week 50-2022. Our analysis considered the discrepancies in population size, the primary course and three booster vaccine doses - taking into account waning immunity, the Stringency Index as a surrogate for non-pharmaceutical interventions adopted in each country, and the circulation of various viral variants. We employed a power law approximation for spatial interactions between countries.

Results We found that within-country transmission was dominant across all countries over almost three years of observation. This work also underscored a basic transmission mechanism, whereby infections introduced by between-country transmission could be of great importance in subsequent local transmission. Furthermore, there were indications of the transition to endemicity since the beginning of 2022, particularly in light of the evolving variants of concern.

Conclusion Our study highlighted the benefit of the endemic-epidemic framework to elucidate the COVID-19 disease spread over a large spatial and temporal scale, using a wide range of epidemiological information. Insights derived from this study are beneficial for those interested in seeking an overview of the emergency phase of the COVID-19 pandemic in the EU/EEA region.

Keywords Spatiotemporal, COVID-19, Europe, Endemic-epidemic

*Correspondence:
Thi Huyen Trang Nguyen
thihuyentrang.nguyen@uhasselt.be
[1]Data Science Institute, I-BioStat, Hasselt University, Martelarenlaan 42, 3500 Hasselt, Limburg, Belgium
[2]Centre for Health Economic Research and Modeling Infectious Diseases, Vaccine and Infectious Disease Institute, University of Antwerp, Prinsstraat 13, 2000 Antwerp, Belgium

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Introduction

SARS-CoV-2 is unlikely to disappear in the near future. The evolving landscape of the COVID-19 pandemic continues to pose a significant threat to public health in the European Region. Over the last five years, this region has reported an excess of 280 million cases and 2.2 million deaths, representing approximately one-third of the total number of reported cases worldwide [1]. The rapid accumulation of reported cases, with a caveat that SARS-CoV-2 infections could manifest in a-/pre-symptomatic state, led to an urgent call for interventions to suppress viral transmission and sustain healthcare capacities. To confront the pandemic, European countries implemented multiple non-pharmaceutical interventions (NPIs) with varying degrees of stringency within and between countries, ranging from social distancing and small gathering cancellations to large-scale lockdowns and international travel bans [2]. In addition, since the COVID-19 vaccine rollout in December 2020, approximately one billion vaccine doses have been administered in the European Union/European Economic Area (EU/EEA) region [3]. It is important to note, however, that these countermeasures underwent modifications and adaptations at different stages of the emergency, particularly in light of new variants of concerns (VOCs). Over time, immunity to the SARS-CoV-2 virus has been developed among the general population, whether through infection, vaccination, or a combination of both. While COVID-19 is no longer a public health emergency of international concern since May 2023, attention has been shifted towards understanding the situation in the monitoring phase [4]. Given that extensive (health) data became publicly available, it is important to undertake a comprehensive retrospective examination of how the disease did spread and what was the effect of interventions throughout the emergency phase of the pandemic [5]. To our knowledge, no studies have been conducted to elucidate the disease transmission mechanism in Europe, which was one of the world's epidemic centers [6], as a whole. It can be reasonably presumed that the epidemiological characteristics of SARS-CoV-2 are analogous across European countries, while these countries exhibit varying degrees of population heterogeneity, including social background, human behaviours, disease susceptibilities, and the implemented countermeasures and policies to the pandemic. We hypothesize that these factors would constitute the disparities in the spatiotemporal transmission mechanisms among these countries.

In Europe, certain works on spatial and spatiotemporal aspects of the COVID-19 disease have been conducted at multi-country level or continent scale [7–11]. For instance, a study by Davis et al. (2021) highlighted that local transmission probably occurred in several areas of Europe and the United States during January and February 2020, using a global metapopulation epidemic model to investigate the space-time heterogeneity in the early stage of the pandemic [10]. However, most of these studies considered data before the implementation of stringent containment measures or before the COVID-19 vaccine became widely available. Fajgenblat et al. (2024), on the other hand, provided a comprehensive picture of the negative relationships between the confirmed COVID-19 cases with the NPIs and vaccination level, extensively oriented on the first three years of the pandemic in 38 European countries, using a Bayesian hierarchical distributed lag model [11].

In advancing the investigation of various aspects of a disease, including spatiotemporal analysis, the coalescence of statistical and mathematical modeling with innovative data sources has become a prominent area of focus [5]. During the COVID-19 emergency, Nunes et al. (2020) strongly demonstrated the crucial role of mathematical methods such as Susceptibles-Infectious-Recovered-like models to mechanistically reflect the disease spread, using epidemiological surveillance data [12]. However, information regarding the number of susceptibles, which is of importance in such models, is not frequently obtained from routine public health data [13]. Alternatively, the Endemic-Epidemic (EE) framework, which was first introduced by Held, Höhle, and Hofmann in 2005, also known as the `hhh4` model [14], is more pragmatic than the full mechanistic ones [13, 15]. Essentially, the EE model is a time-series model of disease incidence that this model can be extended for spatiotemporal multivariate analysis and does not require the number of susceptibles to be available [13]. The model decomposes the expected case counts into an endemic and an epidemic component. While the epidemic component represents an autoregression on the historical counts, i.e., "infectiousness", in the same and other regions, the endemic component represents the background disease risk associated with socio-demographic variables, and/or environmental factors. Although the model can be complex by flexibly incorporating dependencies such as countermeasures and vaccines [16, 17], human social contacts [18–20], and spatial human movements [21–23], the model parameters can be estimated via maximum likelihood estimation and the complexities can be deduced from epidemiological aspects. Besides, the model is practically implemented in the R package `surveillance` [24] and its extensions such as `hhh4addon` [25].

Using the EE model, the objective of our study is to understand the similarities and differences in the spatiotemporal spread of COVID-19 epidemics in 30 EU/EEA countries between March 2020 to December 2022. We conducted an additional simulation study to investigate how well the contribution of each model component from the additive model approach can be estimated in

different counterfactual trajectories of the pandemic. We targeted the EU/EEA countries due to their shared competencies, including public health, from which they benefit from European organizations such as the European Center for Disease Control and Prevention (ECDC). We leveraged the weekly-aggregated COVID-19 case counts in 30 countries from Week 13-2020 to Week 50-2022. We also considered the discrepancies in the population size, vaccination coverage taking into account waning immunity of the primary course and three booster doses, the Stringency Index from the Oxford COVID-19 Government Response Tracker (OxCGRT) [26], and the circulation of various virus variants of concerns (VOCs) while the power law approximation for spatial interaction between countries was applied. From this study, we want to provide the authorities and scientists with a comprehensive review of the pandemic in the European region during the emergency phase.

The paper is organized as follows. We begin by describing our various data sources and their adaptation to be used in our study in Study materials section. The endemic-epidemic spatiotemporal model section describes the proposed modeling approach, and Simulation study section demonstrates how we conducted the simulation study. While Results section shows results of the fitted models to the data and results from the simulation study, Discussion and Conclusion sections discuss and conclude.

Study materials

This study was designed as a spatiotemporal time series analysis with country and week as the main space-time resolution. All data from twenty-seven EU countries and three EEA countries were collected and transformed into weekly intervals from Week 13-2020 (23 March 2020) to Week 50-2022 (18 December 2022) (143 weeks in total). Table 1 summarizes the datasets used in our study.

Table 1 Overview of the different datasets used in the study with details on the available time resolution at the country level

Data	Source	Time resolution	Time range	Date obtained
COVID-19 cases	Our World in Data [27]	Daily	23 March 2020 – 18 December 2022	20 February 2023
The Stringency Index	OxCGRT[1] [26]	Daily	23 March 2020 – 18 December 2022	20 February 2023
Data of variants	ECDC[2] [28]	Weekly	Week 13-2020 to Week 20-2022	09 October 2024
Data of vaccination	ECDC[2] [3]	Weekly	Week 50-2020 to Week 50-2022	09 October 2024
Population data	World Bank [29]	Yearly	2020 – 2022	04 July 2024

[1]The Oxford COVID-19 Government Response Tracker [2]European Centre for Disease Prevention and Control

The COVID-19 reported cases

The daily reported COVID-19 case counts across 30 EU/EEA countries were retrieved from the Our World in Data website [27] on 20 February 2023. We analyzed a total of 182,342,684 cases, ranging from 21,128 cases in Liechtenstein to 39,278,544 cases in France. Figure 1A presents the temporal evolution of the total number of reported cases from all countries. Overall, the epidemic curve demonstrates numerous fluctuations in its trajectory. After the initial spread in March - April 2020, there was a decline in new cases during the summer months of the same year. Thereafter, the number of cases started to increase in October 2020, followed by subsequent waves with peaks and troughs, including significant spikes between late 2021 and early 2022. The highest peak was observed in Week 4-2022 (January 2022) with the total number of reported cases approaching 9 million cases in that week. Figure 2 shows the variability in the distribution of cases per 1,000 population across countries by year. In 2022, countries such as Austria, Germany, and France exhibited the highest number of cases per one 1,000 population.

The Stringency Index

The Stringency Index was extracted from OxCGRT, a dataset that documented the government policies against the COVID-19 pandemic [26]. The Stringency Index provides a summary picture of the NPIs at the national level. It comprises nine metrics, including school and workplace closures, the cancellation of public events, restrictions on public gatherings, closure of public transport, stay-at-home requirements, public awareness campaigns, and restrictions on internal and international mobility. The index was originally calculated on a daily basis and ranges from 0 to 100, with higher values indicating a greater level of stringency of NPIs imposed at certain points in time. In each included country, we computed the weekly mean of this index. Figure 1B summarizes the weekly means across 30 countries from Week 11-2020 to Week 50-2022, which encompasses two weeks before our study period. While the solid line represents the median of weekly means, light-colored ribbons show 2.5% to 97.5% quantiles across countries, depicting cross-sectional heterogeneity. Overall, this index demonstrated a rapid increase in March - April 2020, when more stringent measures were implemented in most countries. Subsequently, it exhibited significant fluctuations in late 2020 and 2021 before decreasing in 2022.

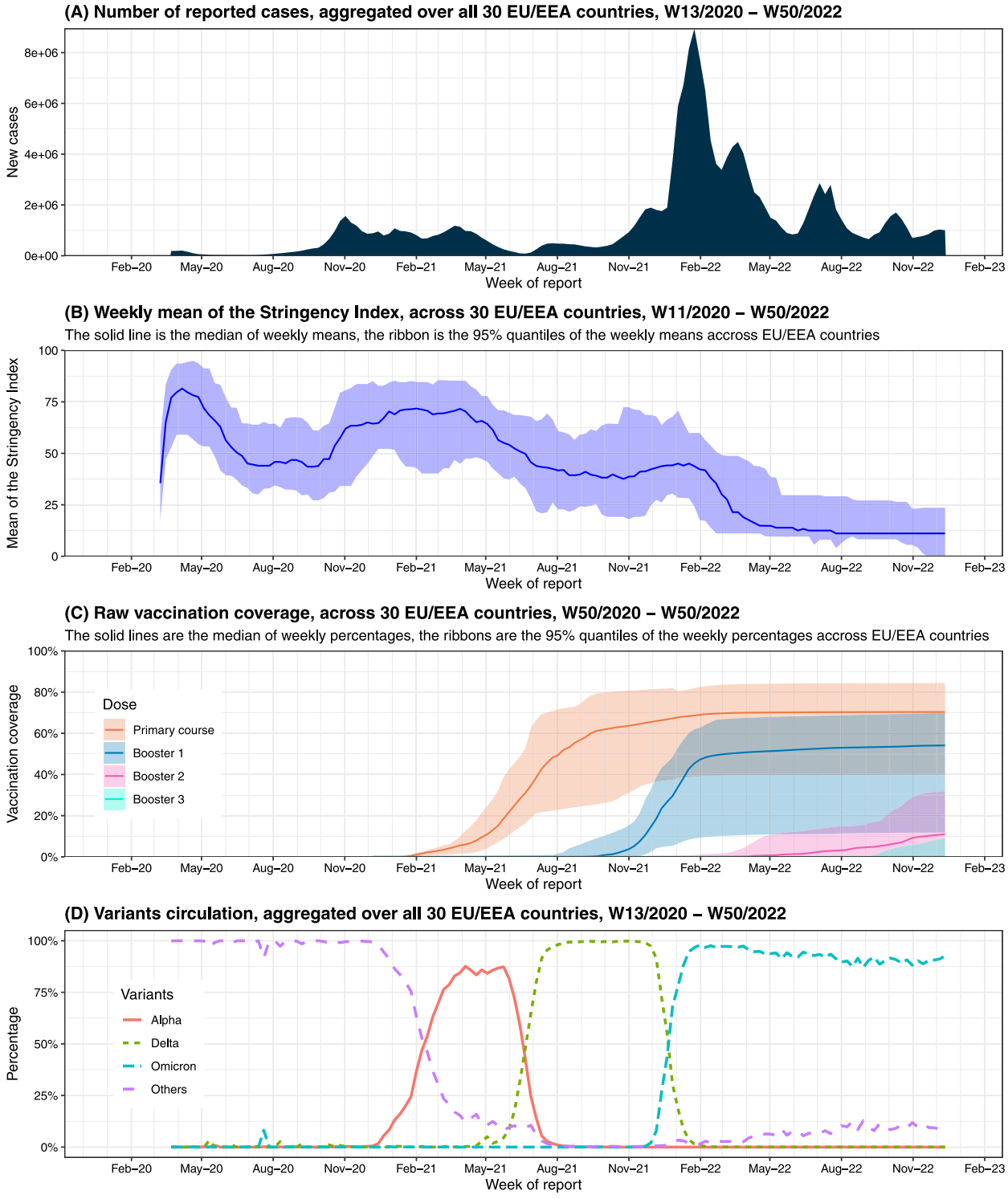

Fig. 1 Overview of the data used in our study: The total number of cases aggregated over all 30 EU/EEA countries (**A**), the weekly mean of the Stringency Index (**B**) and the raw vaccination coverage (**C**) across countries, and the circulation of the Alpha, Delta, Omicron, and other variants aggregated over all countries (**D**), between Week 13-2020 to Week 50-2022. Note that in (**B**) and (**C**), the solid lines present the median and bands are the 95% quantiles of the weekly means of the Stringency Index and the vaccination coverage, respectively

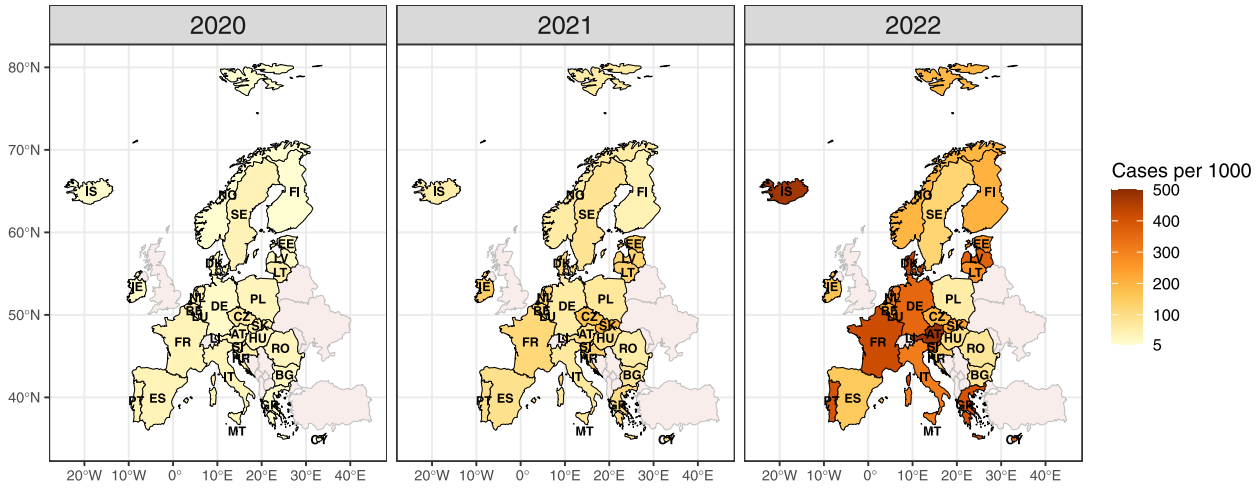

Fig. 2 Number of reported COVID-19 cases per 1,000 population by year and by country within the EU/EEA region. Country names have been assigned a two-letter country code: Austria (AT), Belgium (BE), Bulgaria (BG), Croatia (HR), Cyprus (CY), Czechia (CZ), Denmark (DK), Estonia (EE), Finland (FI), France (FR), Germany (DE), Greece (EL), Hungary (HU), Iceland (IS), Ireland (IE), Italy (IT), Latvia (LV), Liechtenstein (LI), Lithuania (LT), Luxembourg (LU), Malta (MT), Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Slovakia (SK), Slovenia (SI), Spain (ES), Sweden (SE). Countries not part of the EU/EEA are shown in grey as background

COVID-19 vaccination data

The number of vaccine doses administered in each EU/EEA country, including the primary course and three booster doses, were extracted from the ECDC Dashboard for COVID-19 Vaccine Tracker from Week 50-2020 to Week 50-2022 [3]. The raw vaccination coverages of the primary course and booster doses in a given country were calculated as the cumulative number of people receiving the respective vaccine dose at week t relative to that country-specific population (Fig. 1C).

During our study period, most of the COVID-19 vaccine products authorized for use in the EU/EEA countries were the Comirnaty – Pfizer BioNTech (73% of the total doses administered), Spikevax – Moderna (17.3%), and Vaxzervria – AstraZeneca (7.3%) vaccines (see Appendix Figure A4, sourced from Our World in Data [27]). Given that the immunity against COVID-19 induced by vaccines wanes over time, accounting for waning immunity is essential in the analysis. In our study, several assumptions were made: (i) the immunity begins after the completion of the primary course; (ii) full establishment of the immunity starts at two weeks after administration of any dose; and (iii) the waning occurs after six months post-vaccination [30, 31]. Following the approach proposed by Dunbar et al., (2024) [19], we determined the overall vaccination coverage with waning immunity, vac_{it}, at each week for each country as follows:

$$\begin{aligned} vac_{it} &= \sum_{(\cdot)} cov_{it}^{(\cdot)}, \text{ where } cov_{it}^{(\cdot)} \\ &= \frac{\sum_{d \le t}(p_{t-d} \times x_{id}^{(\cdot)})}{pop_{it}}. \end{aligned} \tag{1}$$

Specifically, the overall coverage vac_{it} for each country at each time point is the sum of the coverage with waning of different dose types, $cov_{it}^{(\cdot)}$, where $(\cdot)$ can be the primary course, the first, second, or third booster, in country i at week t. $cov_{it}^{(\cdot)}$ is expressed as a fraction, of which the numerator is the product of the number of new doses $x_{id}^{(\cdot)}$ of the dose type $(\cdot)$ administered at week $d \le t$ in country i and the waning rate p at time $t-d$ that takes two weeks to establish immunity and wanes after six months (see Appendix Figure A5, adapted from [19]), and the denominator is the population size pop_{it} of country i in week t. Figure 3 depicts the vaccination coverage across countries, where the solid lines are the median, and the ribbons are the 95% quantiles of the weekly percentages. For country-specific coverages, see Appendix Figure A7. Among 3,180 data points (106 weeks $\times$ 30 countries), 16 values were greater than one, and thus, we truncated to 0.9999 for numerical reasons.

COVID-19 variants data

The SARS-CoV-2 variants circulating during our study period were obtained from the GISAID database [28] and classified into four categories: the Alpha, Delta, Omicron, and other variants. In each week of observation, the percentage of each variant group was determined by dividing the number of positive sequences associated with that group by the total number of sequenced samples with known variants in that week. In countries with missing data, the percentage of a certain variant was imputed by the overall percentage of that variant in all 30 EU/EEA countries. Figure 1D demonstrates the significant shifts of each variant group: the Alpha variant was the

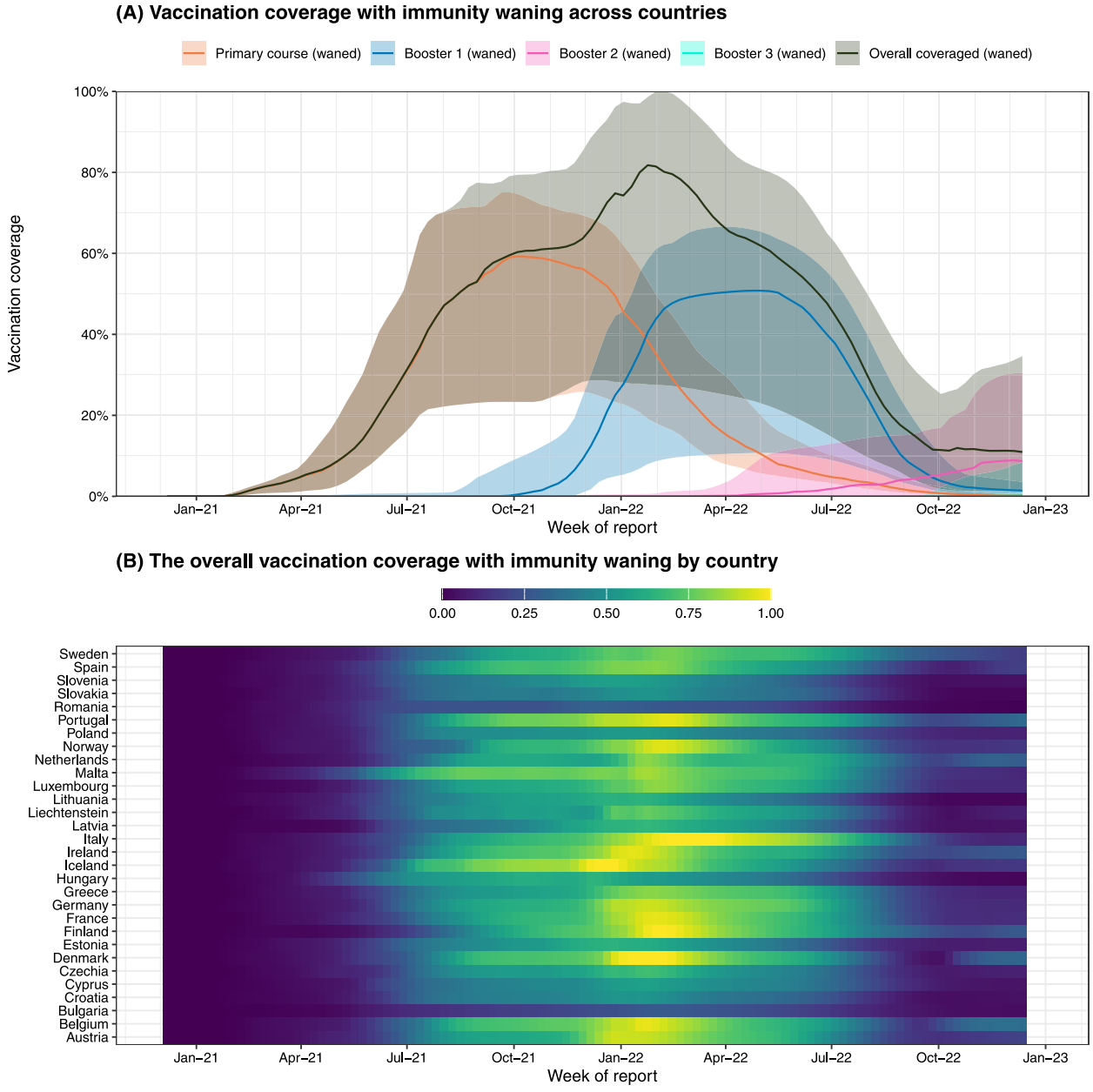

Fig. 3 The COVID-19 vaccination coverage taking into account waning immunity across 30 EU/EEA countries. In panel (**A**), the solid line is the median of weekly percentages, and band is 95% quantiles across 30 EU/EEA countries. Panel (**B**) is the overall coverage by country

predominant strain in the first half of 2021, followed by the Delta variant, which became prevalent until November 2021, and subsequently the Omicron variant, which surged globally since December 2021.

The population data

To ensure that our models accounted for demographic changes over the study period, we referenced the annual population data from the World Bank for the years 2020 to 2022 [29].

For details of the data included, please see Appendix Figure A1 to Appendix Figure A10.

The neighbourhood structure

The connectivity between countries is presented by a neighbourhood matrix, where countries are defined as neighbours if they share a boundary. Formally, we denote o_{ji} as the neighbourhood order from country j to country $i \neq j$, such that $o_{ji} = o_{ij} = r$, where $r > 0$ is defined as the shortest path between each pair of countries, with r steps taken on each journey [21]. Conventionally, this features a symmetric neighbourhood matrix with zeroes on the diagonal.

From the original neighbourhood matrix (Appendix Figure A11), several countries (Iceland, Cyprus, Malta,

Ireland) or groups of countries (Norway, Sweden, Finland) are not connected with other countries because they do not share any boundaries with the other countries. Since this does not realistically reflect the social connectivity amongst countries, the original matrix was adjusted to form a new matrix (aka the maritime neighbourhood matrix, Fig. 4B). Specifically, with regard to the maritime boundaries between countries, first-order connections were established between Finland and Estonia; Sweden and Denmark, Norway, and Estonia; Iceland and Norway; Cyprus and Greece; Malta and Italy. In addition, second-order connections were added between Ireland and France, Belgium, and the Netherlands, given that these countries share maritime neighbours in common with the United Kingdom. These new connections are shown in Fig. 4A's red lines. The adjusted matrix will be used in our spatiotemporal model, which will be described in detail in the following Section.

The endemic-epidemic spatiotemporal model

We generated a multivariate time-series model for the COVID-19 reported cases Y_{it} in country $i = 1, \ldots, I$ during week $t = 1, \ldots, T$. Depending on past observations $Y_{i,t-1}, \ldots, Y_{i,t-K}$, we assumed that Y_{it} has a negative binomial distribution with conditional mean μ_{it}, an overdispersion parameter $\psi > 0$, and conditional variance $\mu_{it}(1 + \psi\mu_{it})$ [14, 25]:

$$Y_{it}|Y_{i,t-1}, \ldots, Y_{i,t-K} \sim \text{NegBin}(\mu_{it}, \psi). \tag{2}$$

Note that when $\psi \equiv 0$, Y_{it} is Poisson distributed. The conditional mean μ_{it} in (2) is given by:

$$\begin{aligned}\mu_{it} =& \lambda_{it}\sum_{k=1}^{K} u_k Y_{i(t-k)} \\ &+ \phi_{it}\left(\sum_{k=1}^{K}\sum_{j\neq i} u_k w_{ji} Y_{j(t-k)}\right) \\ &+ \text{N}_{it}\nu_{it}.\end{aligned} \tag{3}$$

In Eq. (3), the first and the second terms are the *autoregressive* and the *neighbourhood* components (also called as the *spatiotemporal* component), respectively. These are sometimes referred to as the observation-driven epidemic component, as they are modeled based on historical counts. The autoregressive component captures local disease dynamics when new cases arise from infectious individuals in the same country, whereas the neighbourhood component describes the link between the cases in country i being infected by the previous cases in country $j \neq i$. The remaining term in Eq. (3) is the *endemic* component to reflect the background disease incidence in country i. The endemic component was modeled proportionally to an offset of expected counts, which is typically considered as an approximation of the population at risk N_{it}. The three non-negative unknown predictors λ_{it}, ϕ_{it}, and ν_{it} in the three components were modeled as log-linear structures as follows:

$$\begin{aligned}\log(\lambda_{it}) =& \alpha^{(\lambda)} + b_i^{(\lambda)} + \beta_{pop}^{(\lambda)}\log(\text{N}_{it}) \\ &+ \beta_{sus}^{(\lambda)}\log(1 - vac_{it}) + \beta_{\text{SI}}^{(\lambda)}\text{SI}_{t-2} \\ &+ \beta_{\text{VOC}_{[x]}}^{(\lambda)}\text{VOC}_{[x],it},\end{aligned} \tag{4}$$

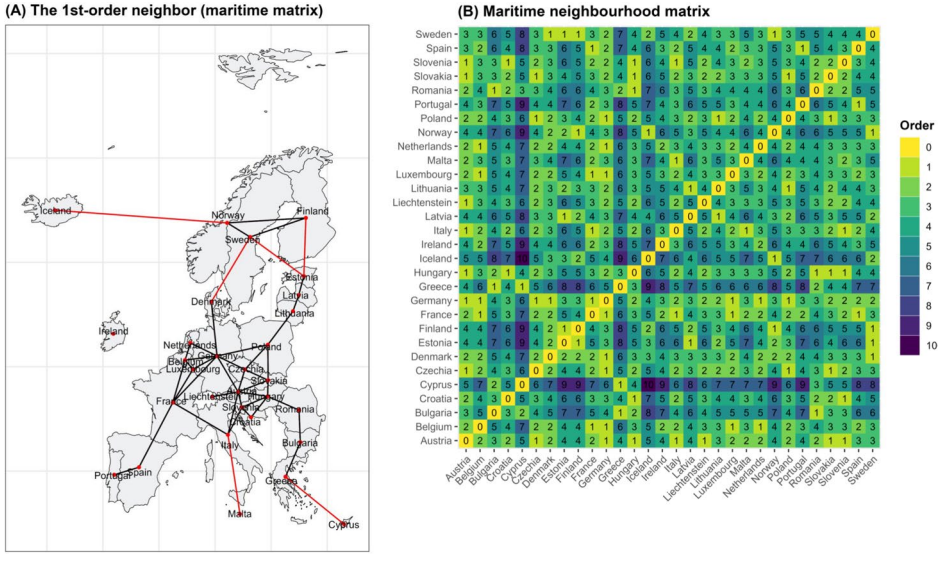

Fig. 4 The first order neighbours (**A**) and the maritime neighbourhood matrix (**B**)

$$\log(\phi_{it}) = \alpha^{(\phi)} + b_i^{(\phi)} + \beta_{pop}^{(\phi)} \log(\mathrm{N}_{it}) + \beta_{\mathrm{SI}}^{(\phi)} \mathrm{SI}_{t-2} + \beta_{\mathrm{VOC}_{[x]}}^{(\phi)} \mathrm{VOC}_{[x],it}, \quad (5)$$

$$\log(\nu_{it}) = \alpha^{(\nu)} + b_i^{(\nu)} + \beta_{sus}^{(\nu)} \log(1 - vac_{it}) + \beta_{\mathrm{VOC}_{[x]}}^{(\nu)} \mathrm{VOC}_{[x],it}. \quad (6)$$

In Eqs. (4), (5), (6), $\alpha^{(\lambda)}$, $\alpha^{(\phi)}$, $\alpha^{(\nu)}$ are the component-specific fixed intercepts. We also introduced the country-specific random intercepts, $b_i^{(\lambda)}$, $b_i^{(\phi)}$, and $b_i^{(\nu)}$ to the model to account for the heterogeneity in incidence level that is not explained by the covariates [32]. These random effects were assumed to be independent and normally distributed with mean zero and variance σ_ν^2, σ_λ^2, and σ_ϕ^2, respectively, to be estimated from the data. In addition, multiple dynamic covariates were incorporated in the model. First, the two epidemic components were scaled with the population size by including the logarithm of the population counts, $\log(\mathrm{N}_{it})$, to reflect that populous countries have higher number of people thus higher level of infectivity (the autoregressive component), and a greater potential to import cases from their neighbouring countries (the spatiotemporal component) [21]. Next, it is assumed that the proportion of the population that had not been vaccinated ($1 - vac_{it}$) can be used as a surrogate for the susceptible population (*sus*) [16, 19, 22]. In our model, $\log(1 - vac_{it})$ was embedded in the endemic and autoregressive components with the corresponding parameters $\beta_{sus}^{(\nu)}$ and $\beta_{sus}^{(\lambda)}$. Moreover, the NPIs were found to have the greatest influence on the change in daily incidence at a lag of 14 days after implementation [11]. Accordingly, we integrated the two-week lag of the Stringency Index (SI_{t-2}) in the two epidemic components to reflect the delayed impact of NPIs on disease transmission dynamics. While different VOCs may suggest different transmission risks [33, 34], the proportion of the $\mathrm{VOC}_{[x]}$, where [*x*] represents the Alpha, Delta, and Omicron variants circulating in each country during the study period, were integrated into the model.

When studying daily case counts, it is reasonable to posit that the number of cases recorded on a given day is not only dependent upon relevant observations on the previous day but also observations further back in time. In the EE models, the lag weight u_k assigned in the two observation-driven components can be perceived as the probability of a discrete-time serial interval distribution (i.e., the distribution of the average number of days between the symptom onset of two consecutive cases) [25]. Some EE studies indicated that the lags for COVID-19 disease were likely to be in the range of seven days [17, 22]. However, as we modeled weekly aggregated COVID-19 case counts, we considered the maximum length of *k* to be $K = 2$ weeks. This was done to minimize potential biases in the assumption of serial intervals due to, for example, the use of aggregated data [25], or the impact of the VOCs [35] or the control measures [36]. The lag weight u_k, which was constrained to be positive, was normalized and defined in accordance with a shifted Poisson weighting scheme [25].

In order to describe the disease spread from country *j* to country *i*, a spatial weight w_{ji}, which is a function of the neighbourhood order o_{ji} with decay parameter $d > 0$, was entered in the spatiotemporal component in Eq. (3). This is referred to as a power law model, which is inspired by human movement behaviour [21]. The spatial weight is expressed as follows :

$$w_{ji} = o_{ji}^{-d}, \quad (7)$$

for ($j \neq i$) and $w_{jj} = 0$. The weight w_{ji} was row-normalized $w_{ji} = \frac{o_{ji}^{-d}}{\sum_{m=1}^{I} o_{jm}^{-d}}$, such that the sum of all rows *j* is equal to one ($\sum_{m=1}^{I} w_{jm} = 1$) to make sure that the cases $Y_{j,t-k}$ are distributed among the countries in a manner that is proportional to the jth row vector of the weight matrix w_{ji} [24]. To ensure positivity, the decay parameter was estimated on a logarithmic scale.

We fitted the EE models with and without random effects to check whether or not the heterogeneity in disease incidence can be adequately captured by the random effects. In mixed models, Paul and Held (2011) showed that commonly used goodness-of-fit criteria, such as the Akaike Information Criterion, may introduce bias, particularly when deciding on the inclusion of random effects [32]. Czado, Gneiting, and Held (2009) proposed proper scoring rules, a more natural approach for model selection [37]. This score quantifies the difference between a predictive distribution from a fitted model and the observed value [32, 37]. A scoring rule is considered proper if its expected score is minimized when the prediction is ideal [38], that is, when the observation is drawn from the predictive distribution. Lower scores indicate better goodness-of-fit. In this paper, we advocated the logarithmic score (logS), and the ranked probability score (RPS) as these two scoring rules simultaneously demonstrate the calibration - *"the statistical consistency between the probabilistic forecasts and the observations"* and sharpness - *"the concentration of the predictive distributions of the predictive distribution"* [37]. The discrepancy between the mean scores was examined through the use of a permutation test with a statistical significance of the differences set at 0.05. For further details regarding the calculation of the aforementioned scores, we refer to [32, 37].

All the analysis was performed in R version 4.4.1, using the open software R packages `surveillance` version 1.23.0 [24], and `hhh4addon` version 0.0.0.0.9014 [25].

Simulation study

A simulation study was implemented to investigate how well the contribution of each of the model components can be estimated in different counterfactual trajectories of the epidemic. In particular, a total of 800 simulations were generated across eight pandemic scenarios, with 100 simulations per scenario. These epidemic scenarios were built from eight sets of parameters originating from the fitted model with component-specific fixed effects (hereinafter referred to as the original model). In Scenario 0, the simulated data were generated from the original model, with all parameters maintained at their original estimated values. In Scenarios 1 through 7, the three component-specific intercepts were adjusted while the impact of VOCs in all three components were either preserved (Scenarios 1 & 2), i.e., the estimates from the original model were upheld, or removed (Scenarios 3 to 7), i.e., set them to zeros. The values of the three intercepts were selected arbitrarily so that the mean total number of cases in 100 simulated datasets would be close to the original total number of cases (Scenarios 1 to 5), or that was half (Scenario 6), or double (Scenario 7) the original total number of cases. Subsequently, all simulated datasets were fitted using the model in Eq. (3) without the random effects. The estimated parameters were summarized by medians, means, and standard deviations, and were then compared with the true values. The coverage probabilities of the true values, i.e., the proportion of models in which the true values fell within the confidence intervals (CIs) for each parameter, were described. The contribution in proportion of each of the three components was also calculated. The eight sets of the parameters and the resulting estimates are presented in detail in Subsection 5.2.

Results

Spatio-temporal transmission of COVID-19 in 30 EU/EEA countries

Table 2 presents the parameter estimates obtained from the models fitted with and without random effects and model assessment using proper scoring rules. The model fitted with random effects demonstrates a better goodness-of-fit as evidenced by lower logS and RPS scores compared to the fixed effects version. Ascribed to the random effects, the estimated overdispersion was also lower, suggesting a reduction in residual heterogeneity. Thus, in this subsection, we concentrate on the findings of the random effects model. Figure 5A depicts the fitted values aggregated across countries. Overall, the model provided a good fit to the observed disease dynamics between 2020 - 2022. The decomposition of the contribution in proportion attributed to each of the three components over time is illustrated in Fig. 5B. Panels (C), (D), and (E) in the same figure summarize the fractions of conditional means explained by each component across 30 EU/EEA countries. The solid lines are the medians of the weekly proportions, and the ribbons illustrate the 2.5% and 97.5% quantiles. A country-specific aggregated overview of the component contribution is also presented in Fig. 6. The results for individual countries can be found in the Appendix Figures A12 & A13.

Throughout the study period, within-country transmission constituted the predominant mode of transmission and this observation was consistent across countries. Overall, the COVID-19 cases attributable to the autoregressive component accounted for 93.3% of the estimated total number of cases. In each country, this component was responsible for explaining between 68.5% (Greece) and 98.4% (Norway) of the total expected mean (Fig. 6A). At certain time points, the overall contribution of this component was estimated approximately between 70% (September & November 2022) and 99% (February - June 2021) (Fig. 5B & C). The within-country transmission persisted until the end of December 2022. To this point, we believe that the local transmission chains within the country were established at the time this study commenced (i.e., March - April 2020). However, in certain countries, such as Liechtenstein, Iceland, and Malta, the contributions of the autoregressive component were quantified to be below 50% during the summer months of 2020 (Appendix Figure A13). This was likely due to the relatively low number of reported cases in relation to the size of the population (Fig. 7).

In contrast to the significant influence of the autoregressive component, the overall impact of the neighbourhood component on the epidemic curve was limited, accounting for 1.3% of the estimated number of cases accumulated over all countries and all time points. The spatiotemporal component was also a relatively minor contributor to the expected number of new infections in each investigated country. Approximately half of the countries exhibited minimal neighbourhood effects (< 1 %), whereas only a few countries, such as Croatia, Belgium, and Luxembourg, demonstrated a higher impact of between-country transmission than the overall value (Fig. 6B). On average, the between-country transmission was significantly pronounced between June and November 2020 but it had a high degree of temporal heterogeneity across nations until the end of 2021 (Fig. 5B & D). Compared to the within-country transmission, this result indicates a basic transmission process: once a sufficient number of infected individuals accumulated via between-locality transmission, the risk of within-locality transmission was likely to increase, especially in the early periods

Table 2 The parameter estimates and 95% confidence intervals (95% CIs) from the models fitted with and without random effects and model assessment using proper scoring rules

Component	Parameter	Notation	Model with random effects			Model without random effects		
			Estimate	2.5 %	97.5 %	Estimate	2.5 %	97.5 %
Endemic	Intercept	$\alpha^{(\nu)}$	-21.384	-27.693	-15.075	-19.107	-23.779	-14.435
	Variance of random intercept	σ^2_{ν}	1.514					
	Log susceptible	$\beta^{(\nu)}_{sus}$	0.160	-0.155	0.475	-0.172	-0.352	0.008
	Variant Alpha	$\beta^{(\nu)}_{\mathrm{VOC}_{[Alpha]}}$	5.558	-4.093	15.209	-2.307	-15.129	10.514
	Variant Delta	$\beta^{(\nu)}_{\mathrm{VOC}_{[Delta]}}$	10.641	4.051	17.230	8.444	3.642	13.245
	Variant Omicron	$\beta^{(\nu)}_{\mathrm{VOC}_{[Omicron]}}$	13.252	6.898	19.606	10.882	6.179	15.585
Autoregressive	Intercept	$\alpha^{(\lambda)}$	-0.166	-0.392	0.060	-0.038	-0.223	0.146
	Variance of random intercept	σ^2_{λ}	0.0012					
	Log population	$\beta^{(\lambda)}_{pop}$	0.0011	-0.014	0.016	0.002	-0.009	0.013
	The Stringency Index	$\beta^{(\lambda)}_{\mathrm{SI}}$	0.0013	-0.0001	0.003	0.0005	-0.0009	0.002
	Log susceptible	$\beta^{(\lambda)}_{sus}$	-0.060	-0.082	-0.037	-0.053	-0.079	-0.026
	Variant Alpha	$\beta^{(\lambda)}_{\mathrm{VOC}_{[Alpha]}}$	-0.013	-0.079	0.053	-0.129	-0.197	-0.061
	Variant Delta	$\beta^{(\lambda)}_{\mathrm{VOC}_{[Delta]}}$	0.146	0.071	0.221	0.077	0.006	0.148
	Variant Omicron	$\beta^{(\lambda)}_{\mathrm{VOC}_{[Omicron]}}$	-0.083	-0.166	-0.00002	-0.180	-0.261	-0.099
Neighbourhood	Intercept	$\alpha^{(\phi)}$	-18.226	-21.870	-14.581	-13.094	-15.233	-10.955
	Variance of random intercept	σ^2_{ϕ}	0.899					
	Log population	$\beta^{(\phi)}_{pop}$	1.073	0.854	1.293	0.701	0.587	0.815
	The Stringency Index	$\beta^{(\phi)}_{\mathrm{SI}}$	-0.090	-0.104	-0.075	-0.087	-0.107	-0.066
	Variant Alpha	$\beta^{(\phi)}_{\mathrm{VOC}_{[Alpha]}}$	-4.573	-7.728	-1.418	-3.492	-5.689	-1.295
	Variant Delta	$\beta^{(\phi)}_{\mathrm{VOC}_{[Delta]}}$	-2.393	-3.024	-1.761	-1.968	-2.711	-1.225
	Variant Omicron	$\beta^{(\phi)}_{\mathrm{VOC}_{[Omicron]}}$	-7.441	-10.204	-4.679	-8.539	-12.484	-4.593
	Log decay parameter	d	-0.627	-1.731	0.477	0.018	-0.544	0.579
	Overdispersion	ψ	0.182	0.174	0.190	0.238	0.228	0.248
Proper scoring rules	The logarithmic score	logS	9.121			9.256		
	The ranked probability score	RPS	7930.063			8554.054		

(see more in Appendix Figures A14 & A15). From late 2021 onward, even though the contribution of the neighbourhood component was remarkably small, this spatio-temporal effect could not be disregarded. It is possible that sources of infections originating from outside the country could serve as a catalyst for the emergence of new community outbreaks within the country (Table 3).

In addition, the evolution of the COVID-19 pandemic in the EU/EEA region was found to have approximately 5.4% of its aggregated number of new infections attributable to the endemic component. However, a notable discrepancy was observed in the temporal and spatial contribution of this component between countries. Approximately two-thirds of the analyzed countries exhibited a lower percentage of endemic transmission compared to the overall estimate. Conversely, in Greece and Finland, for instance, the magnitude of the endemic behaviour was large: they contributed to approximately 31.3% and 19.6% of the estimated total number of cases in each country, respectively (Fig. 6C). The endemic effects, on average, became more pronounced since late 2021 and the fraction of the endemic component peaked at approximately 30% in September 2022 (Fig. 5B, Appendix Figure A16). In some countries, there were gradually elevated contributions to infections originating from the endemic component from the second half of 2021 onward (Fig. 5E and Appendix Figure A13). This suggests the preliminary indications of a shift from epidemic to endemic disease activity. It remains uncertain, however, whether these endemic waves were merely transient or if they were a start of a more prolonged, stationary phase of the disease.

From the parameter estimates in Table 2, the Stringency Index was found to be significantly associated with fewer cases in the neighbourhood component ($\beta^{\phi}_{SI} = -0.090$, 95% CI: (-0.104) – (-0.075)). More specifically, one unit increase in the Stringency Index would result in a reduction of approximately $1 - \exp(-0.090) = 8.6\%$ of the incidence in the neighbourhood component. In the

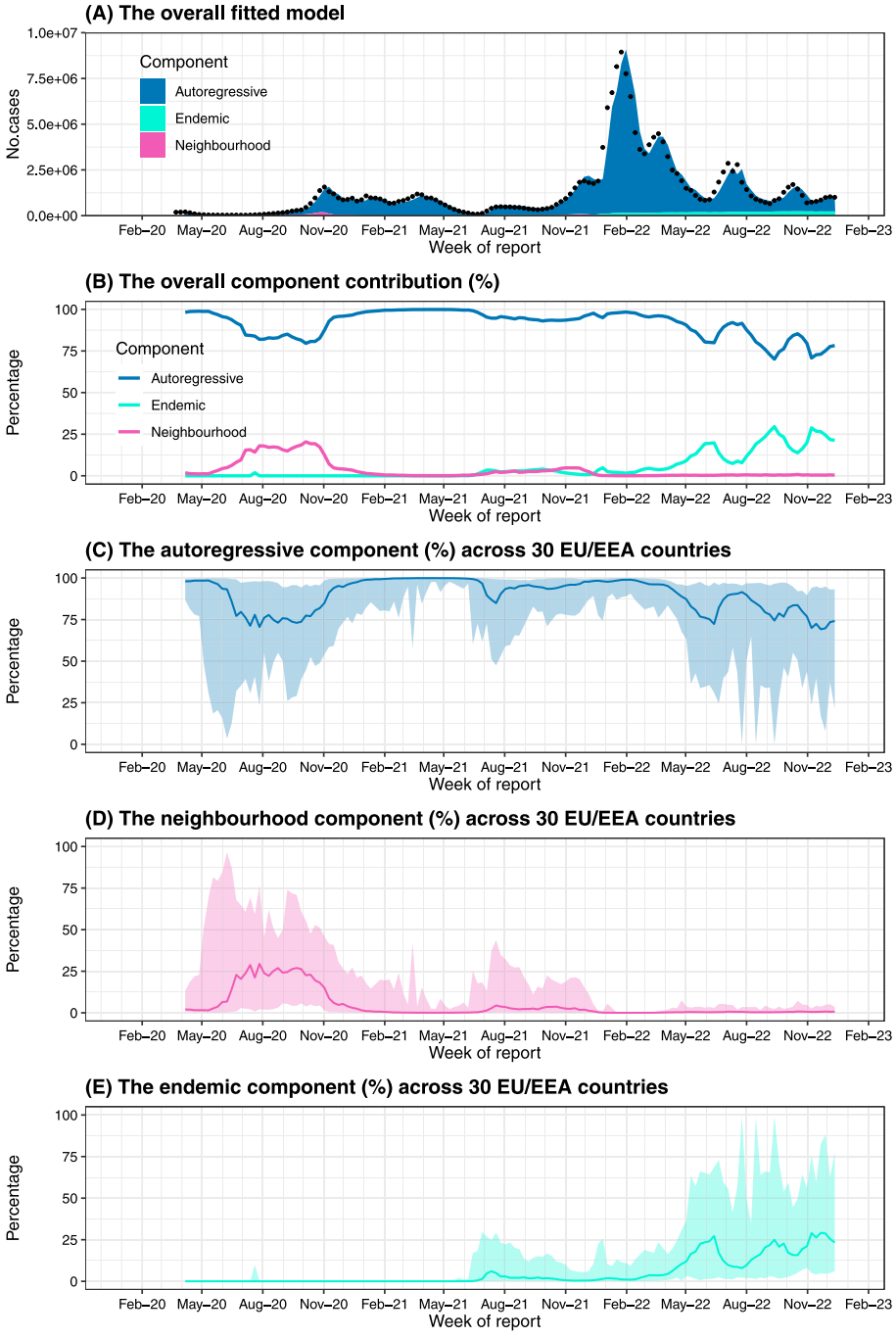

Fig. 5 Panel (**A**) is the fitted components of the random effects model, aggregated across all 30 EU/EEA countries. The dots indicate the observed number of weekly infections. Panel (**B**) shows the contribution in percentage during the observed period from Week 15-2020 to Week 50-2022 (excluding the first two weeks). Panels (**C**), (**D**), (**E**) summarize the contribution in percentage (%) of each of the three components in the weekly cases across 30 EU/EEA countries over time. The solid lines are the medians, and bands are the 95% percentile of the weekly percentages across 30 EU/EEA countries. The autoregressive, the spatiotemporal, and the endemic components are illustrated by blue, pink, and cyan, respectively

autoregressive component, the level of strictness of NPIs as measured by the Stringency Index appeared to have minimal impact on lowering incidence albeit that the corresponding parameter estimate was found to be statistically insignificant. With regard to the impact of vaccination, incorporating a log proportion of the unvaccinated population into the autoregressive component did not substantiate the hypothesis of a negative association between vaccination coverage and the disease incidence resulting from within-country transmission. Conversely, the proportions of unvaccinated individuals seemed to have a positive effect in increasing the disease incidence

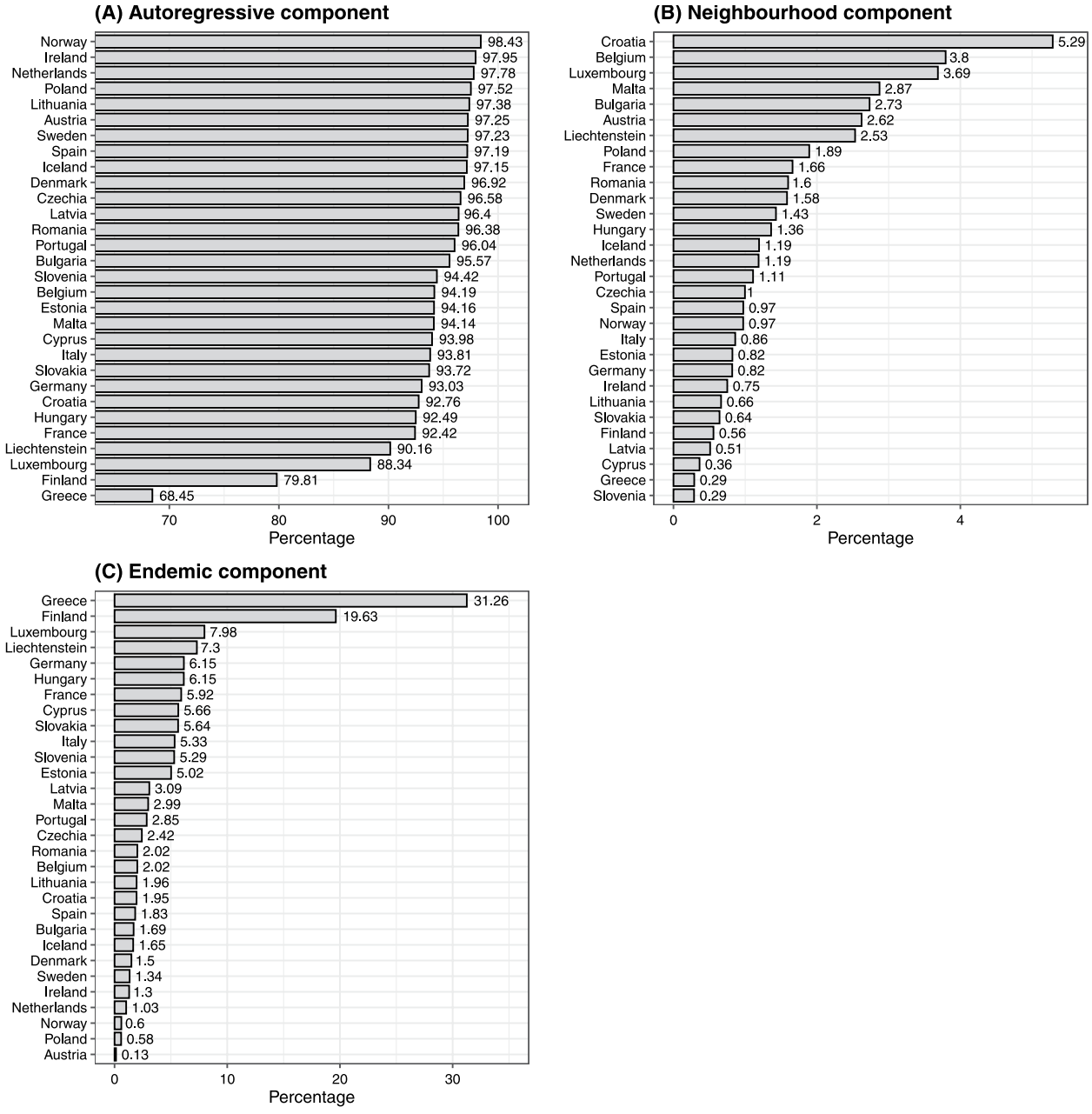

Fig. 6 The contribution in percentage (%) of each of the three components in the total cases estimated from the random effects model in each country

in the endemic component, although this effect was not statistically significant. In addition, countries characterized by larger numbers of inhabitants were more likely to have higher risks of importing a number of cases from the neighbouring countries. The corresponding estimate of this commuter-driven effect was found to be significant (i.e., $\beta_{pop}^{(\phi)} = 1.073$, 95% CI: 0.854 – 1.293), thereby substantiating this association. The spatial interaction between countries, as shaped by the power law model, displayed a relatively slow decay across neighbourhood orders (Fig. 8). The decay parameter was estimated at $d = \exp^{(-0.627)} = 0.534$ (95% CI: 0.177 – 1.611). This result revealed that all included countries may share a strong interconnectivity, with long-distance transmission events occurring beyond their immediate geographic proximity. Concerning the impact of VOCs, our results demonstrated that the endemic level of the disease was significantly influenced by the circulation of variants, particularly the Delta and Omicron variants.

Furthermore, the analysis encompassed a large number of countries, and incorporating the random intercepts in all three components appeared to be a reasonable approach to address the heterogeneity in incidence levels across countries [32]. Our results show that the

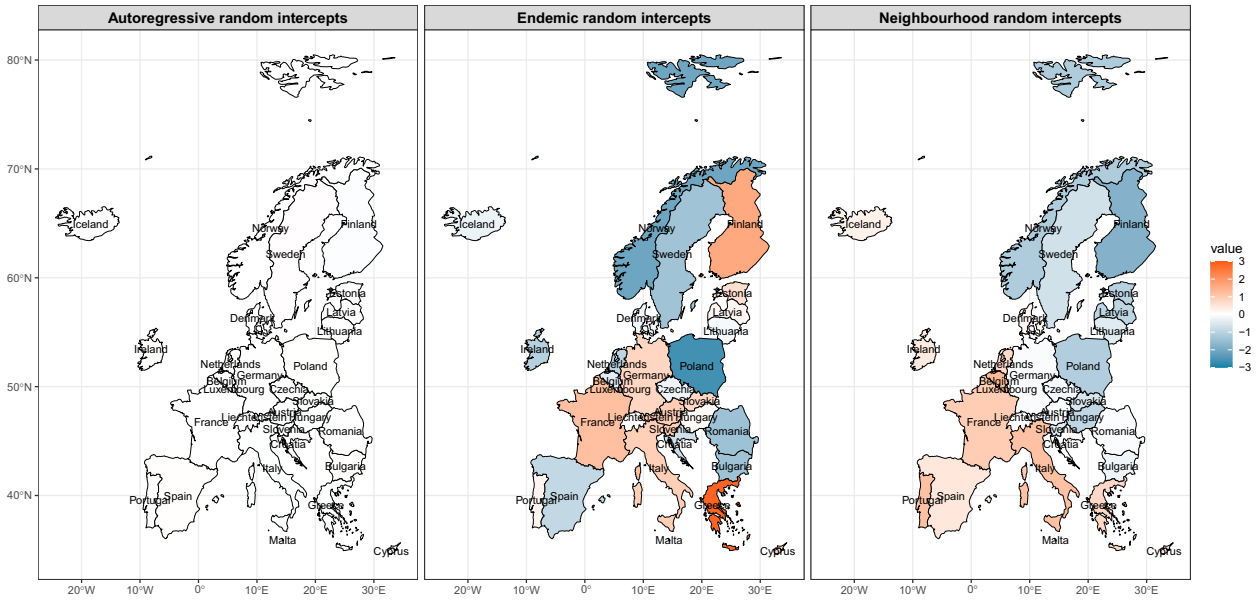

Fig. 7 The estimated country-specific random intercepts in the random effects model. The random intercepts $b_i^{(\lambda)}$, $b_i^{(\phi)}$, and $b_i^{(\nu)}$ of the autoregressive, spatiotemporal, and endemic components, respectively, are visualized in maps. In the left panel, the color cannot be visually displayed because the estimated random intercepts in the autoregressive component are very small compared to the estimated random intercepts in the other two components

autoregressive random effect exhibited minimal variation ($\sigma_\lambda^2 = 0.001$). In contrast, the estimated variance of the neighbourhood component was larger with $\sigma_\phi^2 = 0.807$, suggesting that there was considerable variability in between-country transmission that could not be explained by covariates. In Fig. 7, the neighbourhood random intercepts in the Western and Southern countries were relatively higher, implying a greater likelihood of cross-country transmission compared to the overall transmission rate. The endemic random effects exhibited the greatest variance with $\sigma_\nu^2 = 1.514$. It may present a substantial amount of residual variability, which was not absorbed by covariates in the endemic component due to, e.g., introduction events from outside the study region. Nevertheless, the fixed intercepts in the spatiotemporal and endemic components were found to be remarkably small, which explains why the endemic and neighbourhood components contributed minimally to the overall fit.

Simulation results

All the models fitted on the 800 simulated time series converged. On average, the median values of all estimated parameter values were found to be in close proximity to their true values. The coverage frequency of how often the true value falls within the 95% confidence interval ranged from 77% to 100%. To compare the contribution of each component in explaining the simulated observed data, we plugged the true values into each simulated dataset in each scenario. Subsequently, the fitted `hhh4` models were capable of separately identifying the three components, even though the impact of the spatiotemporal component may have been relatively small. For the summary of simulation results, see Fig. 9, the Appendix Tables A1 to A4, and Appendix Figures A17 to A19.

Discussion

We utilized a multivariate endemic-epidemic model to gain insights into the spatiotemporal spread of COVID-19 epidemics in 30 EU/EEA countries between 2020 - 2022. By facilitating an analysis of weekly aggregated COVID-19 case counts while extensively incorporating other data such as the population data, vaccination coverage taking into account waning immunity, the circulation of the main VOCs, the Stringency Index, and the assumption of a power-law decay of the spatial interaction between EU/EEA countries, we could characterize and quantify the transmission of COVID-19 across multiple countries in the EU/EEA region during the first three years of the pandemic. Insights from our study are beneficial for scientists and public health authorities in seeking a global overview with regard to the emergency phase of the COVID-19 pandemic. The employed methodological approach therein proves its advantages in investigating epidemiological issues through the use of a statistically-sound instrument with a minimal computational cost.

Before discussing the results, it is important to recall that, in reality, “endemic” refers to the constant or expected level of an infectious disease within a population, while an “epidemic” indicates a sudden increase in

Table 3 Different simulated scenarios of the COVID-19 epidemics after changing the parameter estimates from the original fitted model without random effects

Scenario			Scenario 0	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6	Scenario 7
Mean total cases over 100 simulated datasets vs. the original observed data			145.53%	99.78%	96.99%	100.50%	101.39%	103.36%	49.91%	202.76%
Number of models that converged			100	100	100	100	100	100	100	100
Component	**Parameter**	**Notation**	**Point estimates**							
Endemic	Intercept	$\alpha^{(\nu)}$	-19.107	**-16.1**	**-16.6**	**-6.4**	**-6.5**	**-7.1**	**-6.9**	**-6.1**
	Log susceptible	$\beta^{(\nu)}_{sus}$	-0.172	-0.172	-0.172	-0.172	-0.172	-0.172	-0.172	-0.172
	Variant Alpha	$\beta^{(\nu)}_{\mathrm{VOC}_{[Alpha]}}$	-2.307	-2.307	-2.307	**0**	**0**	**0**	**0**	**0**
	Variant Delta	$\beta^{(\nu)}_{\mathrm{VOC}_{[Delta]}}$	8.444	8.444	8.444	**0**	**0**	**0**	**0**	**0**
	Variant Omicron	$\beta^{(\nu)}_{\mathrm{VOC}_{[Omicron]}}$	10.882	10.882	10.882	**0**	**0**	**0**	**0**	**0**
Autoregressive	Intercept	$\alpha^{(\lambda)}$	-0.038	**-1.5**	**-0.8**	**-3**	**-2.9**	**-1.1**	**-2.6**	**-1**
	Log population	$\beta^{(\lambda)}_{pop}$	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	The Stringency Index	$\beta^{(\lambda)}_{\mathrm{SI}}$	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
	Log susceptible	$\beta^{(\lambda)}_{sus}$	-0.053	-0.053	-0.053	-0.053	-0.053	-0.053	-0.053	-0.053
	Variant Alpha	$\beta^{(\lambda)}_{\mathrm{VOC}_{[Alpha]}}$	-0.129	-0.129	-0.129	**0**	**0**	**0**	**0**	**0**
	Variant Delta	$\beta^{(\lambda)}_{\mathrm{VOC}_{[Delta]}}$	0.077	0.077	0.077	**0**	**0**	**0**	**0**	**0**
	Variant Omicron	$\beta^{(\lambda)}_{\mathrm{VOC}_{[Omicron]}}$	-0.180	-0.180	-0.180	**0**	**0**	**0**	**0**	**0**
Neighbourhood	Intercept	$\alpha^{(\phi)}$	-13.094	**-11.9**	**-10**	**-13.3**	**-13.2**	**-13.4**	**-13.9**	**-13.8**
	Log population	$\beta^{(\phi)}_{pop}$	0.701	0.701	0.701	0.701	0.701	0.701	0.701	0.701
	The Stringency Index	$\beta^{(\phi)}_{\mathrm{SI}}$	-0.087	-0.087	-0.087	-0.087	-0.087	-0.087	-0.087	-0.087
	Variant Alpha	$\beta^{(\phi)}_{\mathrm{VOC}_{[Alpha]}}$	-3.492	-3.492	-3.492	**0**	**0**	**0**	**0**	**0**
	Variant Delta	$\beta^{(\phi)}_{\mathrm{VOC}_{[Delta]}}$	-1.968	-1.968	-1.968	**0**	**0**	**0**	**0**	**0**
	Variant Omicron	$\beta^{(\phi)}_{\mathrm{VOC}_{[Omicron]}}$	-8.539	-8.539	-8.539	**0**	**0**	**0**	**0**	**0**
	Log decay parameter	d	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018
	Overdispersion	ψ	0.238	0.238	0.238	0.238	0.238	0.238	0.238	0.238

Numbers in bold indicate that the parameter estimates are modified differently than in Scenario 0

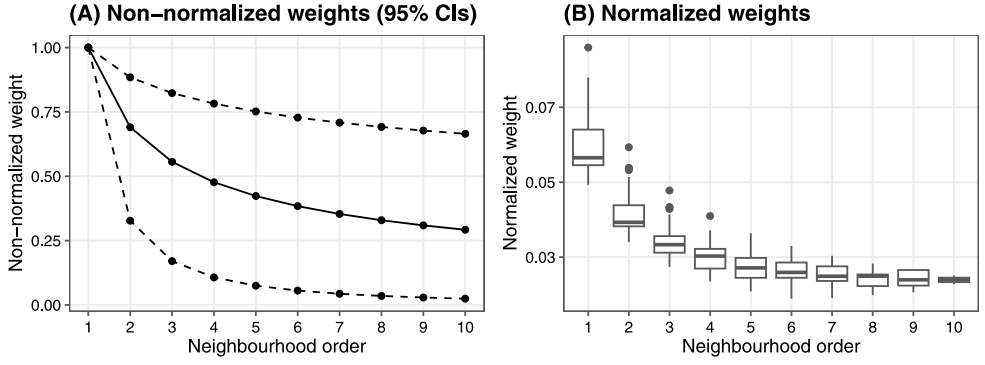

Fig. 8 The non-normalized spatial power law weights $w_{ji} = o_{ji}^{-d}$ and the normalized version of the maritime weight matrix [21]

cases above this expected level [39]. The EE model determines the risk of each susceptible being infected from external sources such as environmental reservoirs or cases imported from outside the study region (endemic component) and from infectives within the same or neighbouring areas (epidemic component) [40, 41]. Thus, this decomposition is a pragmatic simplification, providing an interpretable framework to distinguish baseline incidence from excess transmission in reality.

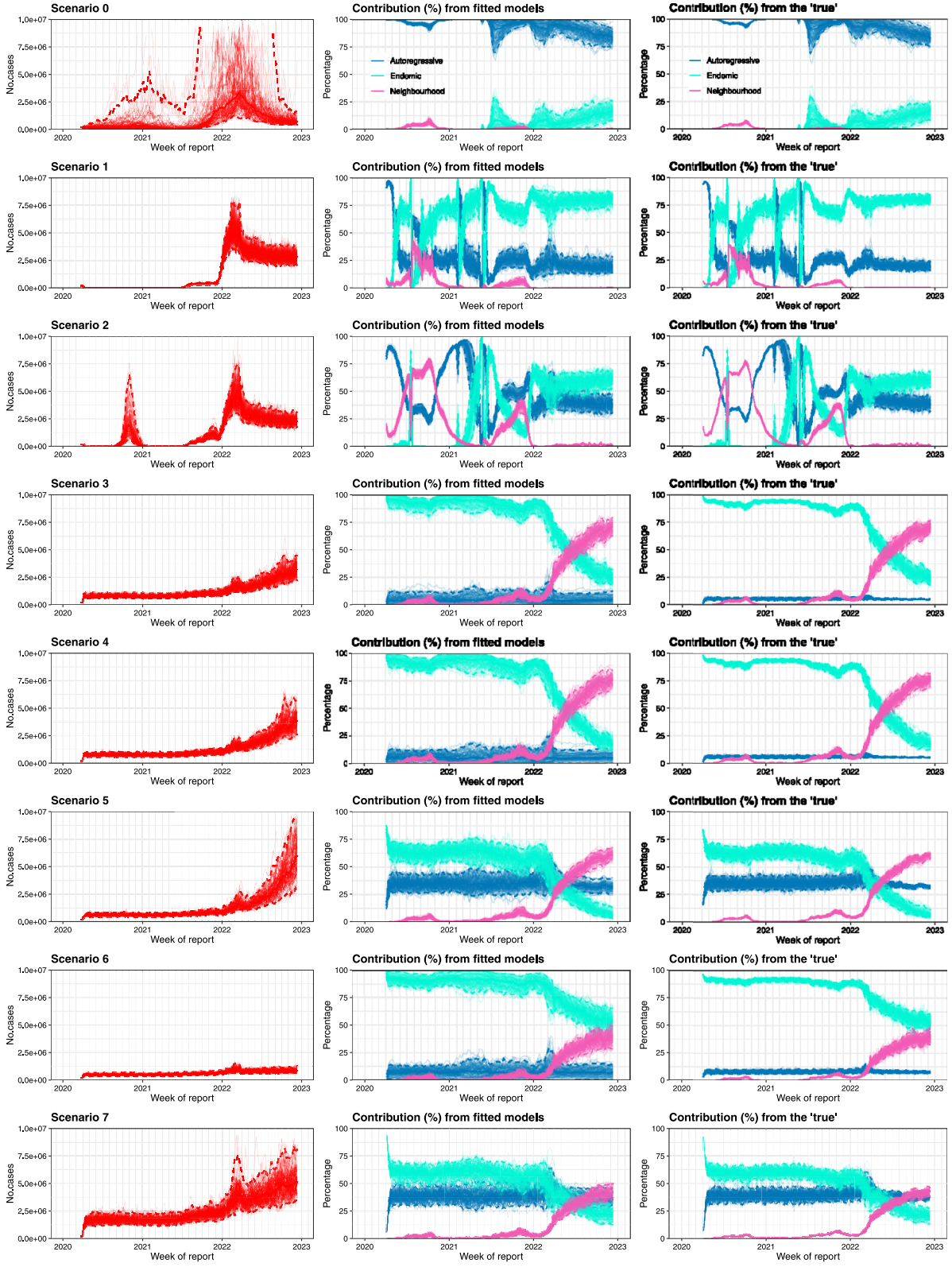

Fig. 9 The first column depicts various simulated scenarios of the epidemic, while the second illustrates the contribution of each of the three components in percentage over time, estimated from fitted models with simulated data. The third column presents the contribution of each component over time when the true parameter estimates from Table 3 are plugged into the simulated datasets. The autoregressive, the spatiotemporal, and the endemic components are illustrated by blue, pink, and cyan, respectively

From the study findings, we conclude that within-country transmission was the dominant transmission mode across EU/EEA countries during 2020 - 2022 as shown in the high fraction of cases attributable to the autoregressive component in each country. This suggests that domestic transmission was already established at the time this study commenced (i.e., March - April 2020) and probably even occurred before this time period [10, 42]. In fact, Davis et al. (2020) proved that the onset of local transmission in European countries was very likely in January and February 2020, a period during which testing capacity was limited, by using a stochastic, spatial, age-structured metapopulation epidemic model [10]. Besides, in many European countries, the elevated contribution of the neighbourhood component during June - November 2020 and the escalation of the autoregressive component contribution from November 2020 onward indicate that SARS-CoV-2 seeding events originating from European countries could have played a major role in successive local epidemic waves. We believe that this could be mainly due to the modest relaxation of NPIs from April to September 2020 [2] after the initial pandemic peak, as also shown in the slight reduction of the Stringency Index (Fig. 1B) in the corresponding period. This exemplifies the rationale behind one of the transmission mechanisms, which posits that introduction events at the outset of the outbreaks, for example via human travel, were instrumental in initiating localized spread. Nevertheless, those introduction events could also provide a means for the independent introduction of virus lineages despite the establishment of local spread [43]. Thus, the influence of spatiotemporal effects is nontrivial in our study, especially from late 2021 onward, when the contribution of the spatiotemporal component was remarkably small.

Importantly, our results showed that there was an early sign of the transition from epidemic to endemic states of the COVID-19 disease from late 2021 onward. The tendency to endemicity is contingent on two key factors: firstly, the evolution of immune-escape and/or more transmissible variants, and secondly, the presence of immunity in the population [44, 45]. In the present study, the former was favored by the strong effects of the Delta and Omicron variants, which are known to be associated with increased transmissibility and immune evasion properties [34, 46]. The timing of the dominance of Delta and Omicron variants was in accordance with the onset of an elevated endemic level of disease. Likely, the geographic variation in the prevalence of VOCs may explain the temporal heterogeneity in the endemic fractions across countries. Regarding the latter, the build-up of immunity in the population through vaccination or natural infection subsides in the number of infections. In circumstances where the persistence of SARS-CoV-2 within the population is attributable to infections arising from susceptible individuals, whether fueled by birth or waning immunity, endemicity is present [45, 47]. This means that the endemic level is marked by the stability in the number of infections in the population [45]. Nevertheless, although the transition can be anticipated, whether the endemicity observed in our study is temporary or becomes more stationary remains uncertain for two reasons. First, SARS-CoV-2, e.g., the Omicron variant and its descendants, continues to evolve and evade immunity [34, 44]. The emergence of a more immune-evasive variant poses a significant challenge to individuals who are partially protected against the virus, as it may fail to elicit variant-specific antibodies in their immune systems. Moreover, given that SARS-CoV-2 immunity is leaky, either through vaccination or previous infection, the likelihood of (re-)infection is increased, especially when more transmissible variants circulate in the community [44]. This results in an increase in the endemic level of the disease, thereby exposing the risk of new endemic waves. Second, the endemic level is also influenced by the potential of introducing new cases from outside the study settings. New infections from external sources not directly related to the observed epidemic history, for example, by immigration, could enter the endemic component. In our study, this immigration effect was captured by the random effects $b_i^{(\nu)}$ into the endemic component. However, it would be preferable to account for the immigration effect by explicitly incorporating immigration and tourism inflow data when it is available [22]. Genetic data could provide valuable insights into cross-border viral introductions and help disentangle local transmission from imported cases, highlighting a promising avenue for future research of inclusion in the EE framework. Country-specific clade patterns may reveal spatiotemporal clusters of importation and refine the assumptions of the endemic component. For example, phylogenomic studies in France and West Africa have traced geographical sources of introduction events and illuminated early transmission dynamics and variant circulation across regions [42, 48]. In sum, the endemic level of disease may increase over time, but this escalation can be countered by implementing protective measures, both pharmaceutical and non-pharmaceutical interventions [44].

In this work, we found that the NPIs measured by the Stringency Index were associated with lowering case counts in the neighbourhood component after accounting for its delayed effects, but this association was not observed in the autoregressive component. We attributed this mixed finding to factors such as population compliance, which is not captured by the SI, and its inability to assess the effectiveness or interaction of individual NPIs [11]. Despite these limitations, we chose the SI as a reasonable proxy for evaluating government responses because of its public availability, standardized structure,

and interpretability across time and countries. Interestingly, higher levels of susceptibility derived from un-vaccinated individuals seem to have a positive (but insignificant) impact on the endemic level of disease. Similarly, the vaccination coverage and the occurrence of epidemics within the country did not yield a significant association. When fitting the EE model with highly multivariate time series and multiple covariates, a high level of complexity is introduced, and thus, identifiability problems during the estimation procedure are likely [16, 19]. However, the results from the simulation study posit that this issue is not responsible for the insignificant estimates of the log of susceptible proportions in the endemic and autoregressive components. Instead, a more epidemiological explanation was proposed: the time-varying susceptibility level in each country was assumed to be proportionate to the proportion of unvaccinated individuals over time [22]; however, while the immunity level, which was solely based on the vaccination coverage with waning immunity, may be underestimated given that it failed to consider the immunity resulting from prior infection with SARS-CoV-2. Furthermore, the conjecture that immunity declines after six months for all vaccine products authorized for use during the study period seems to be a strong assumption [30, 31, 34], given the availability of more refined waning models [49]. The duration of effectiveness of vaccines against SARS-CoV-2 infection may exhibit variation between different vaccine products [30, 31] and the assumed 6-month protection of the vaccines may be less applicable in the presence of Omicron variants [34]. These factors would imply an overestimation of the vaccination coverage in the population, thus imposing biases in the estimation of the true level of vaccine-induced immunity in the population. A population's immunity profile can be collectively enhanced through vaccination and/or previous infection and this phenomenon likely occurs at the local level, rather than the global scale. This collective immunity state does not mean that herd immunity has been achieved; it may be transient and fragile. However, its contribution to the suppression of individual epidemic waves is non-negligible, but subsequent waves may emerge as a consequence of changes in human behaviour, such as human travel and contact behaviour [50].

One of the main drivers of spatial disease diffusion is human mobility. As evidenced in other EE studies [15, 20–22], our study once again verified the existence of an agglomeration phenomenon. When scaling the country susceptibility level by population size in the spatiotemporal component, countries with a higher population density resulted in a greater likelihood of attracting cases from their neighbours. However, it seems naive to assume that the epidemic can only come from their shared-bordered countries. Long-range transmission of cases was addressed by imposing a power law relationship between countries with respect to the neighbourhood order adjacency matrix. It was intuitive that individuals typically travel to locations in closer proximity to their residence, whereas travel to more distant destinations is less frequent. Accordingly, closer neighbouring countries were assigned for greater influence, i.e., more power law weight, while those situated at higher-order neighbours were assigned less weight. However, the estimated slow decay across neighbourhood orders in this work indicated that countries further apart were of importance in spatial disease dispersion. This forms a coarse implication of a pronounced interconnection between countries in the global spread of COVID-19 disease. From a macroscopic perspective, the spatiotemporal spread of the pandemic is not random but rather typically patterned by a human movement network that can be modeled as either constant or time-varying. Examples of such networks are airline transportation data [10], mobile network [22, 23], movement pattern approximation based on generated data sets from Meta-Facebook or Google [9, 17]. Given the interconnected nature of the modern world, a multi-dimensional network paradigm beyond the typical transportation network may provide a superior approach to elucidating the spatiotemporal patterns of the global dissemination of the novel coronavirus [51].

The strength of our study lies in the utilization of the endemic-epidemic modeling framework to project the transmission mechanism of COVID-19 disease at a global scale. Specifically, the `hhh4` model allows us to examine the spatiotemporal transmission process in a granular manner using multiple highly structured data that vary in both in time and space, while requiring a minimal computational burden as compared to other parameter-driven methodologies. Nevertheless, several limitations need to be recorded. The lack of information on a-/pre-symptomatic community incidence is our first limitation when monitoring the COVID-19 pandemic. There is compelling evidence suggesting the importance of a-/pre-symptomatic transmission of SARS-CoV-2. Serological survey data can be used to estimate the fractions of asymptomatic in the population [52]. However, this type of data is not often universally available, and the question remains as to whether these data can be extrapolated to represent the entire country. Secondly, the study did not consider information on testing strategies, which evolved during the emergency phase. This testing information may prove useful in assessing the reporting bias in the model [53]. Finally, the strength of inter-country connectivity, as modeled by the power law, was kept constant over time. This constancy is not likely realistic in the context of the pandemic, given there must be significant fluctuations in mobility patterns in our study setting [54, 55].

Conclusion

From a statistical and evolutionary perspective, we presented a comprehensive view of the spatiotemporal dynamics of SARS-CoV-2 in the EU/EEA region during 2020 – 2022. We found that within-country transmission was the main mode of transmission across all countries over the three years. This work also emphasized a basic transmission mechanism: infections introduced by between-country transmission could be of great importance for subsequent domestic outbreaks. Furthermore, early signs of the transition to endemicity since the beginning of 2022 were conceivable, particularly in light of the evolving VOCs. Our study highlighted the benefits of the endemic-epidemic framework, to cover many epidemiological aspects of COVID-19 diseases at a vast spatial and temporal scale. This is of great interest to those seeking to examine public health questions with multiple data sources available at the time while maintaining the flexibility to update the results when more information becomes available in the future.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-025-24795-6.

Supplementary Material 1: Appendix Tables and Figures.

Authors' contributions

THTN participated in data aggregation, statistical analysis, interpretation of the study results, and writing the initial draft of the manuscript. NH and CF participated in study conception, study design and analysis, interpretation of the study results, supervision and critical revision of the manuscript. All authors approved the final version of the manuscript.

Funding

This study was supported by the Special Research Fund of Hasselt University (BOF08M01, Methusalem grant).

Data availability

The data utilized in this study are publicly accessible on the websites indicated in the manuscript. All downloaded data are shared in Zenodo: https://zenodo.org/records/15533888. DOI:10.5281/zenodo.15533888. To make this study as reproducible as possible, the associated codes are available at the GitHub repository at https://github.com/trangnguyenpmd/COVID-19-EUEEA.

Declarations

Ethics approval and consent to participate

The present study analyzed publicly available data. Accordingly, no ethical approval is required.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 28 May 2025 / Accepted: 5 September 2025
Published online: 21 October 2025

References

1. World Health Organization. WHO COVID-19 dashboard. WHO website. 2024. https://data.who.int/dashboards/covid19/. Accessed 06 Oct 2024.
2. Lionello L, Stranges D, Karki T, Wiltshire E, Proietti C, Annunziato A, et al. Non-pharmaceutical interventions in response to the covid-19 pandemic in 30 European countries: the ECDC-JRC response measures database. Eurosurveillance. 2022;27(41):2101190.
3. European Centre for Disease Control and Prevention. ECDC Dashboard for COVID-19 Vaccine Tracker. ECDC website. 2023. https://vaccinetracker.ecdc.europa.eu/public/extensions/COVID-19/vaccine-tracker.html. Accessed 15 Mar 2023.
4. World Health Organization, et al. Ending the COVID-19 emergency and transitioning from emergency phase to longer-term disease management: guidance on calibrating the response, 4 September 2023. 2023. WHO-WHE-SPP-2023-2; 2023.
5. Jit M, Ainslie K, Althaus C, Caetano C, Colizza V, Paolotti D, et al. Reflections on epidemiological modeling to inform policy during the covid-19 pandemic in Western Europe, 2020–23: commentary reflects on epidemiological modeling during the covid-19 pandemic in Western Europe, 2020–23. Health Aff. 2023;42(12):1630–6.
6. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 13 March 2020. WHO website. 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-mission-briefing-on-covid-19---13-march-2020. Accessed 07 Oct 2024.
7. Sannigrahi S, Pilla F, Basu B, Basu AS, Molter A. Examining the association between socio-demographic composition and covid-19 fatalities in the European region using spatial regression approach. Sustain Cities Soc. 2020;62:102418.
8. Kianfar N, Mesgari MS. Gis-based spatio-temporal analysis and modeling of covid-19 incidence rates in Europe. Spat Spatio-Temporal Epidemiol. 2022;41:100498.
9. Nichita DR, Dima M, Boboc L, Hâncean MG. Data analysis evidence beyond correlation of a possible causal impact of weather on the covid-19 spread, mediated by human mobility. Sci Rep. 2024;14(1):17782.
10. Davis JT, Chinazzi M, Perra N, Mu K, Pastore y Piontti A, Ajelli M, et al. Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave. Nature. 2021;600(7887):127–32.
11. Fajgenblat M, Molenberghs G, Verbeeck J, Willem L, Crèvecoeur J, Faes C, et al. Evaluating the direct effect of vaccination and non-pharmaceutical interventions during the Covid-19 pandemic in Europe. Commun Med. 2024;4(1):178.
12. Nunes B, Caetano C, Antunes L, Dias C. Statistics in times of pandemics: The role of statistical and epidemiological methods during the COVID-19 emergency (invited paper with discussion). Revstat Stat J. 2020;18:553–64. https://doi.org/10.57805/revstat.v18i5.317.
13. Höhle M. Infectious Disease Modeling. In: Handbook of spatial epidemiology. CRC press; 2016. pp. 477–500.
14. Held L, Höhle M, Hofmann M. A statistical framework for the analysis of multivariate infectious disease surveillance counts. Stat Model. 2005;5(3):187–99.
15. Dunbar MBN, Held L. Endemic-epidemic framework used in covid-19 modelling:[discussion on the paper by nunes, caetano, antunes and dias]. REVSTAT-Stat J. 2020;18(5):565–74.
16. Herzog S, Paul M, Held L. Heterogeneity in vaccination coverage explains the size and occurrence of measles epidemics in German surveillance data. Epidemiol Infect. 2011;139(4):505–15.
17. Grimée M, Dunbar MBN, Hofmann F, Held L, et al. Modelling the effect of a border closure between Switzerland and Italy on the spatiotemporal spread of covid-19 in switzerland. Spat Stat. 2022;49:100552.
18. Meyer S, Held L. Incorporating social contact data in spatio-temporal models for infectious disease spread. Biostatistics. 2017;18(2):338–51.
19. Dunbar MBN, Held L. The COVID-19 vaccination campaign in Switzerland and its impact on disease spread. Epidemics. 2024;100745.
20. Nguyen THT, Faes C, Hens N. Measles epidemic in southern Vietnam: an age-stratified spatio-temporal model for infectious disease counts. Epidemiol Infect. 2022;150:e169.
21. Meyer S, Held L. Power-law models for infectious disease spread. Ann Appl Stat. 2014;8(3):1612–39. https://doi.org/10.1214/14-AOAS743.
22. Nguyen MH, Nguyen THT, Molenberghs G, Abrams S, Hens N, Faes C. The impact of national and international travel on spatio-temporal transmission of Sars-Cov-2 in Belgium in 2021. BMC Infect Dis. 2023;23(1):428.

23. Ensoy-Musoro C, Nguyen MH, Hens N, Molenberghs G, Faes C. Spatio-temporal model to investigate covid-19 spread accounting for the mobility amongst municipalities. Spat Spatio-Temporal Epidemiol. 2023;45:100568.
24. Meyer S, Held L, Höhle M. Spatio-temporal analysis of epidemic phenomena using the R Package surveillance. J Stat Softw. 2017;77(11):1–55. https://doi.org/10.18637/jss.v077.i11.
25. Bracher J, Held L. Endemic-epidemic models with discrete-time serial interval distributions for infectious disease prediction. Int J Forecast. 2022;38(3):1221–33.
26. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A global panel database of pandemic policies (oxford covid-19 government response tracker). Nat Hum Behav. 2021;5(4):529–38.
27. Mathieu E, Ritchie H, Rodés-Guirao L, Appel C, Giattino C, Hasell J, et al. Coronavirus Pandemic (COVID-19). Our World in Data. 2020. https://ourworldindata.org/coronavirus. Accessed 20 Feb 2023.
28. European Centre for Disease Control and Prevention. Data on SARS-CoV-2 variants in the EU/EEA. ECDC website. 2023. https://www.ecdc.europa.eu/en/publications-data/data-virus-variants-covid-19-eueea. Accessed 27 Feb 2023.
29. World Bank. Total population by country and year. World Bank website. 2024. https://data.worldbank.org/indicator/SP.POP.TOTL. Accessed 04 Jul 2024.
30. Feikin DR, Higdon MM, Abu-Raddad LJ, Andrews N, Araos R, Goldberg Y, et al. Duration of effectiveness of vaccines against Sars-Cov-2 infection and covid-19 disease: results of a systematic review and meta-regression. Lancet. 2022;399(10328):924–44.
31. Hall V, Foulkes S, Insalata F, Kirwan P, Saei A, Atti A, et al. Protection against Sars-Cov-2 after covid-19 vaccination and previous infection. N Engl J Med. 2022;386(13):1207–20.
32. Paul M, Held L. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. Stat Med. 2011;30(10):1118–36.
33. Dyson L, Hill EM, Moore S, Curran-Sebastian J, Tildesley MJ, Lythgoe KA, et al. Possible future waves of Sars-Cov-2 infection generated by variants of concern with a range of characteristics. Nat Commun. 2021;12(1):5730.
34. Cao Y, Jian F, Wang J, Yu Y, Song W, Yisimayi A, et al. Imprinted Sars-Cov-2 humoral immunity induces convergent omicron RBD evolution. Nature. 2023;614(7948):521–9.
35. Kremer C, Braeye T, Proesmans K, André E, Torneri A, Hens N. Serial intervals for Sars-Cov-2 omicron and delta variants, Belgium, November 19-December 31, 2021. Emerg Infect Dis. 2022;28(8):1699.
36. Torneri A, Libin P, Scalia Tomba G, Faes C, Wood JG, Hens N. On realized serial and generation intervals given control measures: the Covid-19 pandemic case. PLoS Comput Biol. 2021;17(3):e1008892.
37. Czado C, Gneiting T, Held L. Predictive model assessment for count data. Biometrics. 2009;65(4):1254–61.
38. Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration and sharpness. J R Stat Soc Ser B Stat Methodol. 2007;69(2):243–68.
39. Center for Disease Control and Prevention. Section 11: Epidemic Disease Occurrence. USCDC website. 2025. https://archive.cdc.gov/www_cdc_gov/csels/dsepd/ss1978/lesson1/section11.html. Accessed 30 Jul 2025.
40. Wakefield J, Dong TQ, Minin VN. Spatio-temporal analysis of surveillance data. In: Handbook of infectious disease data analysis. Chapman and Hall/CRC; 2019. pp. 455–75.
41. Höhle M. Infectious Disease Modeling. In: Handbook of spatial epidemiology. CRC press; 2016. pp. 477–500.
42. Gámbaro F, Behillil S, Baidaliuk A, Donati F, Albert M, Alexandru A, et al. Introductions and early spread of Sars-Cov-2 in France, 24 January to 23 March 2020. Eurosurveillance. 2020;25(26):2001200.
43. Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, et al. Establishment and lineage dynamics of the Sars-Cov-2 epidemic in the UK. Science. 2021;371(6530):708–12.
44. Otto SP, MacPherson A, Colijn C. Endemic does not mean constant as Sars-Cov-2 continues to evolve. Evolution. 2024;6:1092–108.
45. Telenti A, Arvin A, Corey L, Corti D, Diamond MS, García-Sastre A, et al. After the pandemic: perspectives on the future trajectory of Covid-19. Nature. 2021;596(7873):495–504.
46. Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, et al. The evolution of Sars-Cov-2. Nat Rev Microbiol. 2023;21(6):361–79.
47. Antia R, Halloran ME. Transition to endemicity: understanding Covid-19. Immunity. 2021;54(10):2172–6.
48. Wruck W, Adjaye J. Detailed phylogenetic analysis tracks transmission of distinct sars-cov-2 variants from china and europe to west africa. Sci Rep. 2021;11(1):21108.
49. Conway E, Walker C, Lydeamore M, Golding N, Ryan G, Mavec D, et al. Optimal timing of booster doses in a highly vaccinated population with minimal natural exposure to COVID-19. medRxiv. 2024;2024–05.
50. Tkachenko AV, Maslov S, Elbanna A, Wong GN, Weiner ZJ, Goldenfeld N. Time-dependent heterogeneity leads to transient suppression of the covid-19 epidemic, not herd immunity. Proc Natl Acad Sci. 2021;118(17):e2015972118.
51. Tsiotas D, Tselios V. Understanding the uneven spread of covid-19 in the context of the global interconnected economy. Sci Rep. 2022;12(1):666.
52. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-Olmeda M, et al. Prevalence of Sars-Cov-2 in Spain (Ene-Covid): a nationwide, population-based seroepidemiological study. Lancet. 2020;396(10250):535–44.
53. Geilhufe M, Held L, Skrøvseth SO, Simonsen GS, Godtliebsen F. Power law approximations of movement network data for modeling infectious disease spread. Biom J. 2014;56(3):363–82.
54. Docquier F, Golenvaux N, Nijssen S, Schaus P, Stips F. Cross-border mobility responses to covid-19 in Europe: new evidence from facebook data. Glob Health. 2022;18(1):41.
55. Shepherd HE, Atherden FS, Chan HMT, Loveridge A, Tatem AJ. Domestic and international mobility trends in the United Kingdom during the covid-19 pandemic: an analysis of facebook data. Int J Health Geogr. 2021;20:1–13.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.