Journal Pre-proof

Real-time estimates of the emergence and dynamics of SARS-CoV-2 variants of concern: A modeling approach

Nicolò Gozzi, Matteo Chinazzi, Jessica T. Davis, Kunpeng Mu, Ana Pastore y Piontti, Marco Ajelli, Alessandro Vespignani, Nicola Perra

PII: S1755-4365(24)00066-5
DOI: https://doi.org/10.1016/j.epidem.2024.100805
Reference: EPIDEM 100805

To appear in: *Epidemics*

Received date : 3 April 2024
Revised date : 19 November 2024
Accepted date : 20 November 2024

Please cite this article as: N. Gozzi, M. Chinazzi, J.T. Davis et al., Real-time estimates of the emergence and dynamics of SARS-CoV-2 variants of concern: A modeling approach. *Epidemics* (2024), doi: https://doi.org/10.1016/j.epidem.2024.100805.

Real-time Estimates of the Emergence and Dynamics of SARS-CoV-2 Variants of Concern: A Modeling Approach

Nicolò Gozzi[1,*], Matteo Chinazzi[2], Jessica T. Davis[2], Kunpeng Mu[2], Ana Pastore y Piontti[2], Marco Ajelli[3], Alessandro Vespignani[2,1], Nicola Perra[4,2,5]

[1] ISI Foundation, Turin, Italy

[2] Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA USA

[3] Laboratory for Computational Epidemiology and Public Health, Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN, USA

[4] School of Mathematical Sciences, Queen Mary University of London, UK

[5] The Alan Turing Institute, London, UK

[*] nicolo.gozzi@isi.it

Abstract

The emergence of SARS-CoV-2 variants of concern (VOCs) punctuated the dynamics of the COVID-19 pandemic in multiple occasions. The stages subsequent to their identification have been particularly challenging due to the hurdles associated with a prompt assessment of transmissibility and immune evasion characteristics of the newly emerged VOC. Here, we retrospectively analyze the performance of a modeling strategy developed to evaluate, in real-time, the risks posed by the Alpha and Omicron VOC soon after their emergence. Our approach utilized multi-strain, stochastic, compartmental models enriched with demographic information, age-specific contact patterns, the influence of non-pharmaceutical interventions, and the trajectory of vaccine distribution. The models' preliminary assessment about Omicron's transmissibility and immune evasion closely match later findings. Additionally, analyses based on data collected since our initial assessments demonstrate the retrospective accuracy of our real-time projections in capturing the emergence and subsequent dominance of the Alpha VOC in seven European countries and the Omicron VOC in South Africa. This study shows the value of relatively simple epidemic models in assessing the impact of emerging VOCs in real time, the importance of timely and accurate data, and the need for regular evaluation

of these methodologies as we prepare for future global health crises.

1 Introduction

The evolution of the COVID-19 pandemic was significantly marked by the emergence of SARS-CoV-2 Variants of Concern (VOCs). While the virus' mutations have been monitored since the pandemic's onset [1], certain lineages exhibited distinct characteristics in terms of transmissibility, mortality, or immune escape, altering the global pandemic landscape. Many variants eventually subsided or remained geographically contained. Others, like Alpha and Omicron, dramatically changed the course of the pandemic.

To assess the initial risk of VOCs importation from one region to another, several studies have proposed models that integrate international and domestic traffic patterns [2–4]. Others have focused on estimating the relative transmissibility of emerging VOCs compared to pre-existing strains [5–9]. These efforts have also assessed potential increases in the severity of the disease caused by VOCs and their impact on cases, hospitalizations, and fatalities [7]. To this end, researchers have harnessed genomic data and provided statistical analyses of VOCs' detection [5, 6, 8]. By integrating data of social contacts, mobility, and demographic indicators, other studies have adapted epidemiological models to account for the attributes of these novel VOCs [7, 9]. Furthermore, a significant body of research has been devoted to characterizing the initial introduction and establishment of VOCs in different countries [4, 6, 10–12]. These studies have predominantly leveraged phylodynamic analyses of large scale genomic datasets reconstructing the emergence of VOCs and quantifying their initial introduction thereby shedding light on their spatial dissemination [10–12]. This is particularly relevant given that many VOCs typically undergo a cryptic initial phase where they spread silently without being detected by traditional surveillance systems [13].

The development of parsimonious models able to synthesize intelligence from diverse data sources is essential to assess the risks and shape responses. However, during the initial phases of a VOC's emergence and spread, the reliability of any analytical method is challenged by time constraints and limited information. In this context, we undertake a critical and retrospective analysis of the modeling strategy we developed in real-time during the emergence and spread of Alpha and Omicron. This study focuses on detailing and assessing a mechanistic modeling approach developed and calibrated during the early stages of VOCs onset, employing minimal data. We employed this framework in two real-world scenarios providing: i) an early assessment of a VOC's increased transmissibility and potential for immune escape; ii) an early estimation of its relative prevalence and the timeline for its dominance. In particular, we provide a retrospective evaluation of our initial, real-time estimates concerning the Alpha variant's timeline for dominance in seven European countries (made public on 23/02/2021 [14])

and of our preliminary analyses of the Omicron variant's transmissibility and immune escape potential in South Africa (made public on 05/01/2022 [15]). The results reveal that for Alpha, the actual dates of dominance fell within our model's confidence intervals in most countries studied. In the case of Omicron, our model effectively mirrored its rapid rise in South Africa, with initial estimates of transmissibility and immune escape closely matching later studies. However, data limitations at the time led to a broad range of credible values for these parameters, underscoring the challenges of modeling in real-time scenarios.

Overall, our findings highlight the significance of relatively streamlined, yet robust, epidemic models in promptly responding to the challenges posed by emerging VOCs. The study not only validates our initial results but also underscores the necessity for reliable data in the early stages of a VOC's emergence. This research contributes to the ongoing discussion on effective methodologies for gathering essential intelligence on new VOCs, affirming the value of mechanistic modeling approaches in understanding and managing pandemic emergencies.

2 Results

In this section, we show the performance of our modeling approach, designed to address typical challenges faced during the emergence of novel VOCs. Our methodology focuses on two primary objectives: i) conducting an early assessment of a VOC's increased transmissibility and potential for immune escape; ii) estimating its relative prevalence and the timeline for its dominance in its apparent source or in other countries.

Our modeling framework relies on an age-structured SEIR (Susceptible, Exposed, Infectious, Recovered) compartmental structure, augmented with additional compartments to account for COVID-19-related fatalities, individuals infected with the newly emerged VOC, and vaccinated individuals (only considered for the Omicron variant).

A detailed discussion of our model, including its structure and the parameters used, is reported in Section 4.1 and in the Supplementary Information.

2.1 Estimate ranges of increased transmissibility and immune escape potential

As a prototypical example for the early assessment of the properties of a VOC, we consider the case of the emergence of Omicron in South Africa. Omicron, classified as lineage B.1.1.529, was first identified in South Africa in November 2021 and rapidly spread to over a hundred countries [16, 17]. Its swift global dissemination, even in regions with high vaccination rates, raised concerns about increased transmissibility and immune evasion. Omicron's numerous mutations, some linked to neutralizing antibody escape, were confirmed in preliminary studies to reduce the effectiveness of two vaccine doses [18].

Additionally, data on breakthrough infections in South Africa suggested Omicron's significant ability to evade naturally acquired immunity [19]. Furthermore, the variant emerged while many high and upper-middle-income countries were administering booster doses, several lower-middle and lower-income countries were still advancing their primary vaccination campaigns [20]. Fortunately, existing vaccines, natural immunity, and booster doses continued to provide substantial protection against severe disease outcomes and deaths [17, 18].

In order to estimate Omicron's properties we adopted a multi-stage Approximate Bayesian Computation (ABC) calibration process. As first step we fitted the epidemic trajectory prior to the emergence of Omicron considering confirmed deaths in South Africa in the period May 1^{st}, 2021 – November 23^{rd}, 2021. In doing so, we calibrated the model on the conditions where the VOC initially spread. Indeed, starting in May 2021, the country experienced a third pandemic wave fueled by the Delta VOC [19]. This variant was able to replace Beta, which was responsible for the second wave. Hence, before the appearance of Omicron, Delta was responsible for the large majority of cases. From this standpoint we studied the characteristics of Omicron by performing a second calibration to constrain the model's outcomes to the number of cases reported in the country up to December 14^{th}, 2021. We focused on cases rather than deaths because the latter being a lagging indicator were not yet reflecting the emergence of Omicron at that time. The second calibration step allowed us to estimate posterior distributions of the Omicron's transmissibility advantage with respect to the previously circulating strain (η) and its immune evasion potential (ν) due to the reduction of natural and vaccine induced protection against this VOC. Full details on model calibration is provided in Sec. 4.6

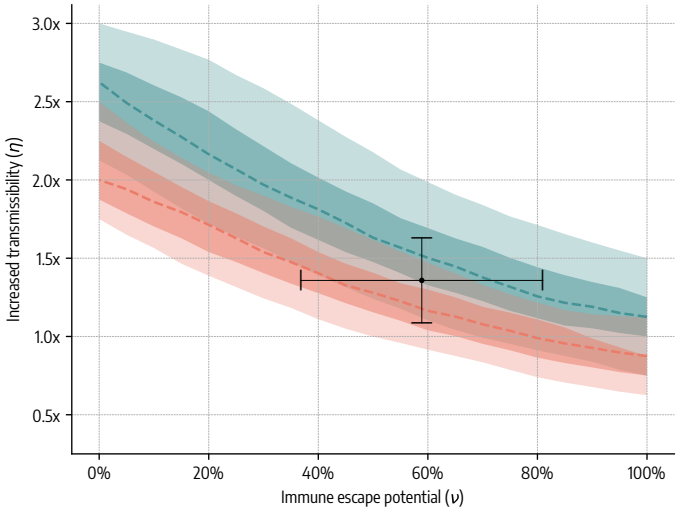

Figure 1: Posterior distribution (median, 50%, and 95% confidence intervals) of Omicron increased transmissibility η and immune escape potential ν. We report results for a $T_G = 5.5$ days (green) and for a $T_G = 3.5$ days (orange). We also show the estimated intervals for η and ν taken from subsequent independent studies.

Figure 1 shows the joint posterior distribution of η and ν, considering two different generation times

(T_G): 5.5 days (represented by the green line and shaded areas) and 3.5 days (represented by the orange line and shaded areas). Indeed, some works hinted to a possible shortening of the generation time of Omicron compared to previous lineages [21]. While large regions of the parameter space are rejected by the calibration, a plausible region that corresponds to the non-identifiable nature of both parameters emerges. Generally speaking, the higher the immune escape ν, the smaller η. Intuitively, a large value of immune escape can be compensated by changes in the relative transmissibility with respect to other strains and vice versa.

In the plot, we also display the ranges of η and ν obtained from subsequent independent studies grounded in more evidence. Specifically, for the range of η, we refer to Ref. [22], which presents a modeling study on the spread of COVID-19 in South Africa, including a comparison of transmissibility among different variants of concern. As for ν, we consider the findings from a systematic review on the waning of SARS-CoV-2 immunity [23]. Our model's credible region aligns with the ranges obtained from independent studies, indicating the validity of our preliminary insights. Notably, the central estimate of these ranges falls within the 50% confidence interval of our model projections when using a T_G of 5.5 days, and within the 95% confidence interval when considering a T_G of 3.5 days.

2.2 Estimating VOCs' potential and path to dominance

During the emergence of new VOCs, understanding their dynamics and pathways to dominance presents additional challenges. In this study, we analyze the dynamics of both the Omicron and Alpha VOCs to demonstrate the effectiveness of our approach.

In the case of Omicron, we modeled its growth within South Africa, which is the country where it was first reported. Hence, we introduced a number of initial Omicron seeds based on the evidence available at the time as initial conditions of the model [24].

Fig. 2A shows the evolution of the percentage of infections due to Omicron in South Africa as simulated by our models (median, 50%, and 95% confidence intervals) and as reported by official genomic surveillance [25] in late 2021. It is important to notice how the genomic data plotted serves as independent evaluation of the model's output as it was not used in any stage of calibration. Fig. 2B, instead, shows the date when Omicron was responsible for at least half of new cases as estimated by our model (50% and 95% confidence intervals) and as estimated on genomic surveillance data via a logistic fit (see Section 4). This threshold is generally regarded as the *date of dominance* and it is used to indicate when the VOC has overcome the previously circulating strain(s) in terms of share of cases. We can see that our model can accurately reproduce the extremely rapid growth of the VOC in the country ($wMAPE = 13\%$), with the estimated date of dominance from genomic data that falls within 95% model's confidence intervals, as shown also in Table 1. Here, we assumed a T_G of 5.5 days for Omicron, in the Supplementary Information we also show results assuming a shorter generation time (3.5 days) obtaining minor differences.

With respect to Omicron, our analysis of Alpha, started later relative to its first detection. At that moment several studies, based on different approaches, already provided a picture of the characteristics of the variant and its initial spread in the UK [26, 27]. Therefore, we focused on projecting its dynamic and path to dominance in other countries outside of its apparent source. The Alpha VOC was first detected in the United Kingdom in September 2020 [28, 29] and identified as lineage B.1.1.7. Its emergence, associated with a surge in cases [30, 31], led to stringent non-pharmaceutical interventions such as lockdowns and travel bans [32]. Despite these measures, Alpha spread globally, becoming dominant in several countries and challenging the initial stages of COVID-19 vaccination efforts, primarily impacting the unvaccinated population. At the moment of our analysis, the Alpha variant was already well established in several countries and the challenge was how to account for this prior dispersion in the modeling approach. For this task, we used GLEAM, a global stochastic metapopulation model that simulates the mobility of people across more than $3,300$ sub-populations in about 190 countries and territories [33–36] (more details on GLEAM in Sec. 4.3). To account for the stochastic nature of Alpha introductions and the beginning of local transmission, we considered a total of $\sim 300,000$ stochastic simulations generated by GLEAM. Specifically, we only focused on the arrivals, in the target countries, of exposed individuals within each age group. Indeed, at the time of Alpha's emergence, travellers were required to exhibit a negative test and other checks to prevent symptomatic individuals from travelling were conducted in airports. The first two specimens of the Alpha variant were collected on September 20 and 21, 2020 in London and Kent areas, respectively. Since UK sequenced about 5% of positive cases [37], we assumed an initial cluster of Alpha variant infections in week 38 of 2020 drawn from a Poisson distribution with a mean value of 40. As a second step, we modeled the local evolution of the newly emerged VOC once community transmission had started. We set the relative transmissibility $\eta = 1.5$ based on the available evidence at that time [26, 27], while maintaining the same generation time as the previously circulating strain(s). In the Supplementary Information, we include a sensitivity analysis exploring different η values. The analysis reveals that lower values of η result in a slower replacement of the wild type (WT) by the Alpha variant, whereas higher values of η lead to a faster replacement. Here with term *wild type* we indicate the previously circulating strain with respect to the VOC under examination. However, it is noteworthy that $\eta = 1.5$ generally yields more accurate results in terms of capturing the actual progression of the Alpha VOC in the target countries. As discussed below however, for two countries higher values of η lead to better results.

We used this parametrization and initialization to model the growth of Alpha in seven European countries (Denmark, Germany, Greece, Italy, Poland, Portugal, and Spain). The choice of countries was dictated by the risk of importing the VOC. Indeed, Germany, Greece, Italy, Poland, Portugal, and Spain are the six top countries with highest number of trips originating from London airports in October 2020. Additionally, Denmark was the first country to report an Alpha infection after United Kingdom [38].

Fig. 2C shows the evolution of the percentage of infections due to Alpha in the target countries as simulated by our models (median, 50%, and 95% confidence intervals) and as reported by official genomic surveillance [25]. As for Omicron, we note how not only we did not use the genomic data in the calibration, but many of the data points shown in the plots have been released only after our initial work. Indeed, our calibration stopped on 14/02/2021. Here, we assume that after this date, contacts remained at the level observed in the last calibration week. In the Supplementary Information, we explore alternatives where NPIs are relaxed, finding that the different NPIs scenarios considered have very limited impact on the estimated date of dominance. Fig. 2C also includes the weighted mean absolute percentage error ($wMAPE$) between model's median and reported data. Overall, the model reproduced the rise of Alpha in the countries considered, with $wMAPE$ values ranging from 7% in the case of Denmark to 35% in the case of Poland and Portugal (average $wMAPE$ is 23%). Despite some clear deviations (see Poland, Italy and Portugal), most of points fall within the confidence intervals of our simulations. Fig. 2D, instead, shows the dominance dates as estimated by our model (50% and 95% confidence intervals) and as estimated from genomic surveillance. For Denmark, Germany, Greece, Portugal, and Spain the date of dominance estimated from genomic data falls in the 50% confidence intervals of our model. The dates of dominance for Italy, and Poland precede the model's median but nonetheless fall within the 95% confidence intervals. Lower accuracy of the model for this two countries is also confirmed by the higher $wMAPE$ between simulated and reported Alpha percentage incidence discussed above. The accuracy of the model in some instances might be less reliable due to several factors. For instance, the model imposed a higher transmissibility rate for the Alpha variant, characterized by $\eta = 1.5$. As mentioned, this value was derived from preliminary findings in the UK. However, it is crucial to note that the establishment of the Alpha variant occurred in considerably diverse epidemiological contexts and under different non-pharmaceutical interventions (NPIs). Consequently, relative transmissibility η may well vary across countries. A sensitivity analysis on plausible values of η is presented in the Supplementary Information. The results indicate that higher values of η yield more accurate estimates for both Italy and Poland. Secondly, sequencing data may contain biases resulting from alterations in testing policies, sample methods, and sample sizes. Specifically, enhanced testing efforts directed towards the detection of a VOC could potentially skew estimates towards higher values. Third, it is important to acknowledge the inherent stochasticity of events like the initial emergence of a new VOC within a country. Such events can be influenced by a multitude of factors, random occurrences, and shifts in population behavior. Finally, as shown in the Supplementary Information, it's important to highlight that the proportion of sequenced cases differs considerably among the countries under consideration. It is encouraging to see that our model performs particularly well in Denmark, a country that managed to sequence nearly half of all reported infections during the period in question. Conversely, our model's performance was lower in Italy and Poland, the countries with the lowest sequencing rates. Both countries sequenced less than 1%

of reported infections, which may have contributed to higher noise in reported sequencing and therefore lower performance of our model in these contexts.

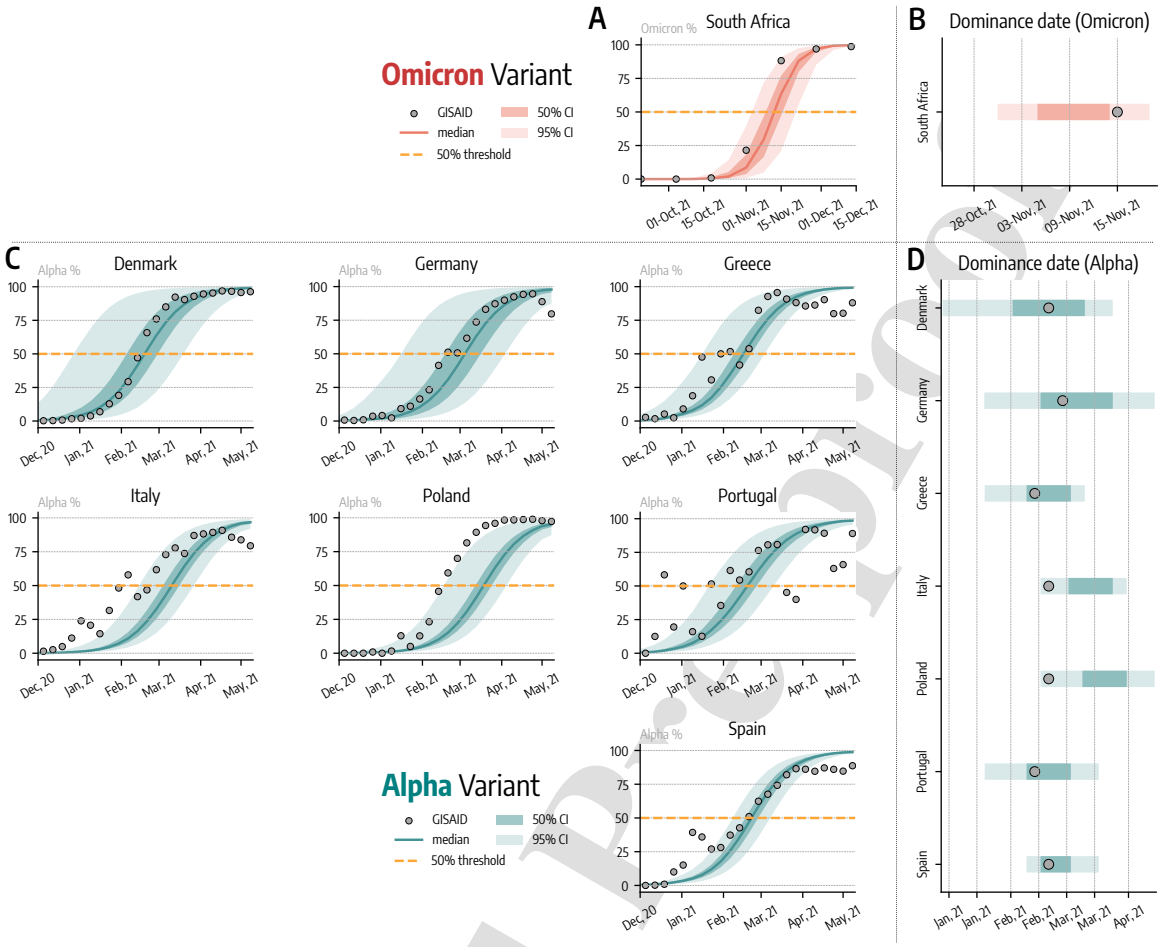

Figure 2: **Path to dominance of Alpha and Omicron VOCs in selected countries.** A) Percentage of new infections due to the Omicron VOC in South Africa as reported by genomic surveillance (grey dots) and as estimated by our model (median, 50%, and 95% confidence intervals). B) Date when Omicron was responsible for at least half of new cases as estimated from genomic surveillance (grey dot) and as projected by our model (50% and 95% confidence intervals). C) Percentage of new infections due to the Alpha VOC in seven European countries as reported by genomic surveillance (grey dots) and as estimated by our model (median, 50%, and 95% confidence intervals). D) Date when Alpha was responsible for at least half of new cases as estimated from genomic surveillance (grey dot) and as projected by our model (50% and 95% confidence intervals).

3 Discussion

The emergence of new SARS-CoV-2 VOCs has represented pivotal moments in the evolution of the COVID-19 pandemic. Characterizing the behavior of these variants is vital, yet the high degree of uncertainty at the early stage of their emergence has presented substantial obstacles to formulating effective responses. In these circumstances, computational epidemic models emerge as essential tools,

		Median	CI 50%	CI 95%	Estimated on Data
Alpha	*Poland*	2021-11	[2021-10; 2021-12]	[2021-07; 2021-14]	2021-07
	Greece	2021-07	[2021-06; 2021-08]	[2021-03; 2021-09]	2021-06
	Denmark	2021-07	[2021-05; 2021-09]	[2020-53; 2021-11]	2021-07
	Portugal	2021-07	[2021-06; 2021-08]	[2021-03; 2021-10]	2021-06
	Germany	2021-09	[2021-07; 2021-11]	[2021-03; 2021-14]	2021-08
	Italy	2021-10	[2021-09; 2021-11]	[2021-07; 2021-12]	2021-07
	Spain	2021-08	[2021-07; 2021-08]	[2021-06; 2021-10]	2021-07
Omicron	*South Africa*	2021-45	[2021-44; 2021-45]	[2021-44; 2021-46]	2021-46

Table 1: Date when VOC was responsible for at least half of new cases as estimated from genomic surveillance and as projected by our model (median, 50% and 95% confidence intervals). Dates are reported in the format year, ISO week.

offering critical insights that combined with genomic data and phylodynamic analysis can contribute to the development of better informed strategies for containment and mitigation [39].

The retrospective analysis of the early, and in real-time, results obtained with our modeling approach offers valuable insights and lessons. The results presented here show that our modeling approach provided the correct probable ranges for two key epidemiological aspects of Omicron: its heightened transmissibility and potential for immune evasion. Moreover, the models were effective in projecting the expansion of the Alpha variant across seven European nations. Notably, the actual dates when Alpha became dominant were within our model's 50% confidence intervals for five of these countries and within the 95% confidence intervals for the remaining two. Similarly, our models precisely captured the rapid surge of the Omicron variant in South Africa by matching both the observed growth and the date of dominance.

Computational epidemiological models are not without limitations. Indeed, they operate on certain assumptions and simplifications that may not always capture the complexity of real-world dynamics. One such challenge lies in the reliance on detailed data regarding epidemiological metrics and behavior, such as mobility or contact patterns, to accurately isolate the impact of a new variant of concern from other factors. In our study, we employed data from the COVID-19 Community Mobility Report published by Google LLC [40], to characterize shifts in behavior induced by NPIs and the spread of the disease. Nevertheless, this data may be biased due to variations in smartphone usage. Indeed, the availability and quality of such data have exhibited heterogeneity across countries, geographical resolutions, and stages of the pandemic [41]. In addition, we assumed that a given level of mobility translated into specific contact levels, thus overlooking changes in the adoption of other NPIs, such as the use of face masks, which could potentially influence the spread of the disease. Furthermore, due to lack of data across all the countries studied, our modeling of NPIs translated only in a modulation of the intensity of interactions. We did not consider possible changes in the overall structure of contact patterns possibly resulting from the adoption of social distancing measures. Our models operate at the national level, which limits their ability to account for heterogeneity at more localized levels in factors such as the epidemiological

situation, prevalence of variants, and the adoption of NPIs. Lastly, while our compartmentalization setup is commonly used [35, 42, 43], it is relatively simplistic and does not explicitly account for factors such as asymptomatic individuals. In addition to these limitations, modeling choices can also influence the results. As shown in the Supplementary Information, variations in epidemiological parameters (e.g., increased transmissibility of Alpha, generation time of Omicron) and assumptions about the impact of NPIs on contact patterns (in the case of Alpha) can affect the outcomes of the model.

It is also worth acknowledging the fundamental role of accurate sequencing data, which delineates variant prevalence. Our model's assumptions were largely based on the analysis of such data, which also underpinned several key studies informing public health responses [39]. This was particularly evident with the Alpha variant, where our model performed better in countries with more robust sequencing capabilities. Finally, let's stress that this type of data is also essential for conducting retrospective analyses like the one presented here.

Despite potential limitations, it is crucial to highlight that the models we proposed yielded satisfactory results with minimal information and assumptions about the VOCs under study. For the Omicron variant, we determined a credible range for its key epidemiological characteristics by using a two-step Approximate Bayesian Computation calibration technique, relying solely on reported deaths and cases data. Similarly, for the Alpha variant, our model relied only on VOC importation estimates derived from GLEAM and assumed a reasonable range for increased transmissibility based on available evidence at the time. Despite the lack of precise data allowing a more detailed model structure and initialization, our results exhibited considerable accuracy in both cases, effectively capturing the growth dynamics of the VOCs and identifying valid ranges for essential epidemiological parameters. When interpreting these results, it is worth remarking that the models have not been re-calibrated to incorporate information that became available later, such as revisions and data backfills. This approach ensures a truly retrospective evaluation of real-time analyses. The success of our approach can be attributed to its mechanistic nature, which enabled us to dissect various factors influencing the VOCs' growth, including transmissibility, immune escape, generation time, and behavioral changes. As we transition away from the emergency phase of the COVID-19 pandemic, retrospective evaluations like ours acquire even more importance, demonstrating how mechanistic epidemiological models can offer crucial insights into the initial growth of future VOCs, and better equip the response to upcoming health threats.

4 Materials and methods

4.1 Epidemic model

We consider a SEIR compartmentalization setup. Healthy and susceptible individuals are placed in compartment S and can get infected by interacting with infectious individuals. When this happens,

they transition to the compartment of the Exposed (E). These are infected but not yet infectious. Only after the latent period (ϵ^{-1}) they become infectious themselves and transition to compartment I. Infected individuals remain infected for the duration of the infectious period (μ^{-1}), after which they transition to the compartment of the Recovered (R). Additionally, we divide individuals into 10 age groups ($[0-9, 10-19, 20-24, 25-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+]$), therefore we use the notation X_k to indicate individuals in compartment X and age group k. We also introduce the contact matrix $\mathbf{C}$, whose elements C_{ij} indicate the number of contact that, on average, an individual in age group i has with individuals in age group j in a given time. We consider country specific synthetic contact matrices from Ref. [44]. While the structure of $\mathbf{C}$ is not modified throughout our simulation period, the overall intensity of contacts is modulated using proxy data on the impact of mitigation policies, as explained in Sec. 4.2. Nonetheless, we acknowledge that non-pharmaceutical interventions implemented during the COVID-19 pandemic may have induced structural changes in contact matrix as well.

In this context, the per capita rate at which susceptible individuals in age group k get infected (i.e., the force of infection) is given by $\lambda_k = \beta \sum_{k'} C_{kk'} I_{k'}/N_{k'}$, where β is the transmission rate of a single contact. Finally, we also calculate the number of deaths by applying the age-stratified Infection Fatality Rate (IFR) from Ref. [45] on the number of daily recovered (i.e., individuals transitioning from I to R). To account for delays between the transition $I \to R$ and actual death due to hospitalization and reporting lags, we record deaths computed on recovered of day t on day $t + \Delta$. The model has also a seasonality term $s(t)$ which is multiplied by λ_k to account for variations in humidity, temperature, and other factors which might affect transmissibility and contact patterns [46]. The seasonality term is equal to $s_i(t) = \frac{1}{2}\left[\left(1 - \frac{\alpha_{min}}{\alpha_{max}}\right)\sin\left(\frac{2\pi}{365}(t - t_{max,i}) + \frac{\pi}{2}\right) + 1 + \frac{\alpha_{min}}{\alpha_{max}}\right]$here i refers to the hemisphere considered [34]. The value $t_{max,i}$ is the time corresponding to the maximum of $s(t)$. It is fixed to January 15^{th} in the northern hemisphere and six months later in the southern one, while we assume $s(t) = 1$ in the tropical region (i.e., no seasonal modulation). We set $\alpha_{max} = 1$ and consider α_{min} as free parameter.

We extend this framework with specific compartments to consider the emergence of a VOC and vaccinations. First, we add specific compartments for individuals infected by the VOC, namely L_{VOC}, I_{VOC}, R_{VOC} and the related compartments for vaccinated individuals. The VOC can have a different transmission rate β_{VOC}, latent period ($1/\epsilon_{VOC}$), infectious period ($1/\mu_{VOC}$), and infection fatality rate (IFR_{VOC}). More in detail, we use the parameter η to describe the increased transmissibility of the VOC with respect to the wild type circulating ($\eta = R_0^{VOC}/R_0^{WT}$).

Second, we model vaccinations as follows. Individuals who have received the first vaccine dose are moved into a novel compartment designated as S^{V1}. The susceptibility of S^{V1} individuals to infection is attenuated by a factor of $1 - VES_1$, where VES_1 is vaccine efficacy against infection. Upon infection,

their IFR is also reduced by a factor $1 - VEM_1$. Thus, the overall efficacy of the first vaccine dose against death is $VE_1 = 1 - (1 - VES_1)(1 - VEM_1)$. Subsequent to receiving the second dose, S^{V1} individuals transition to the compartment S^{V2}. The second administration has an efficacy against infection and mortality, denoted as VES_2 and VEM_2 respectively, and its overall efficacy against fatality is $VE_2 = 1 - (1 - VES_2)(1 - VEM_2)$. Additionally, all vaccinated individuals present a decreased infectiousness by a factor of $(1 - VEI)$ [47]. We assume that S, L, and R individuals are eligible for vaccination, and a delay of Δ_V days is considered between the administration of the vaccine (both first and second doses) and the emergence of the vaccine's actual effect. We consider the number of doses administered daily from Ref. [48] and we assume that the vaccine rollout proceeded prioritizing the elderly. More in detail, in the simulations we assign doses in decreasing order of age (starting from 80+). Once all individuals aged 50 and above have received vaccination, we proceed by uniformly distributing doses among those under 50, while excluding those under 18.

It is important to underline that vaccinations are considered only in the case of Omicron and not Alpha. Indeed, Alpha emerged prior to the initiation of vaccination campaigns and reached dominance in Europe at the outset of COVID-19 vaccination efforts. Furthermore, when we introduce the first Omicron infection we also move R individuals (i.e., individuals previously infected by the wild type) to a new compartment $\hat{R}$. We use a parameter ν to describe the ability of Omicron to escape natural and vaccine acquired immunity. We assume that individuals in S^{V1}, S^{V2}, and $\hat{R}$ compartments see their protection against Omicron provided by immunity reduced by a factor $(1 - \nu)$. We assume also that initial protection of $\hat{R}$ is equal to that of V_2 individuals.

A schematic representation of the full model (with deaths, VOC, vaccinations) is provided in Fig. 3.

4.2 Modeling of mitigation policies

We measure the temporal variations in contacts influenced by non-pharmaceutical interventions utilizing the COVID-19 Community Mobility Report published by Google LLC [40]. The dataset provides, for various regions, the percentage change, denoted as $r_l(t)$, in total visits to specific locations l on day t, compared to a pre-pandemic baseline. We convert this percentage into a contact rescaling factor, denoted as $\omega_l(t) = (1+r_l(t)/100)^2$, under the assumption that the potential number of contacts per location scales proportionally to the square of visitor numbers. When considering the Omicron variant, we calculate a single $\omega(t)$ by averaging the fields for workplaces, retail and recreation, and transit stations, which we then apply to the overall contact matrix $\mathbf{C}$. In the case of the Alpha variant, instead, we consider the different layers of the contact matrix—home, workplaces, school, and the general community setting—separately. The workplace contact layer is multiplied by $\omega_l(t)$ computed from the workplaces field of the report, while the general community settings are adjusted by $\omega_l(t)$ derived from the average of the retail and recreation and transit stations fields. For school contacts, we use the ordinal index `C1 School closing`

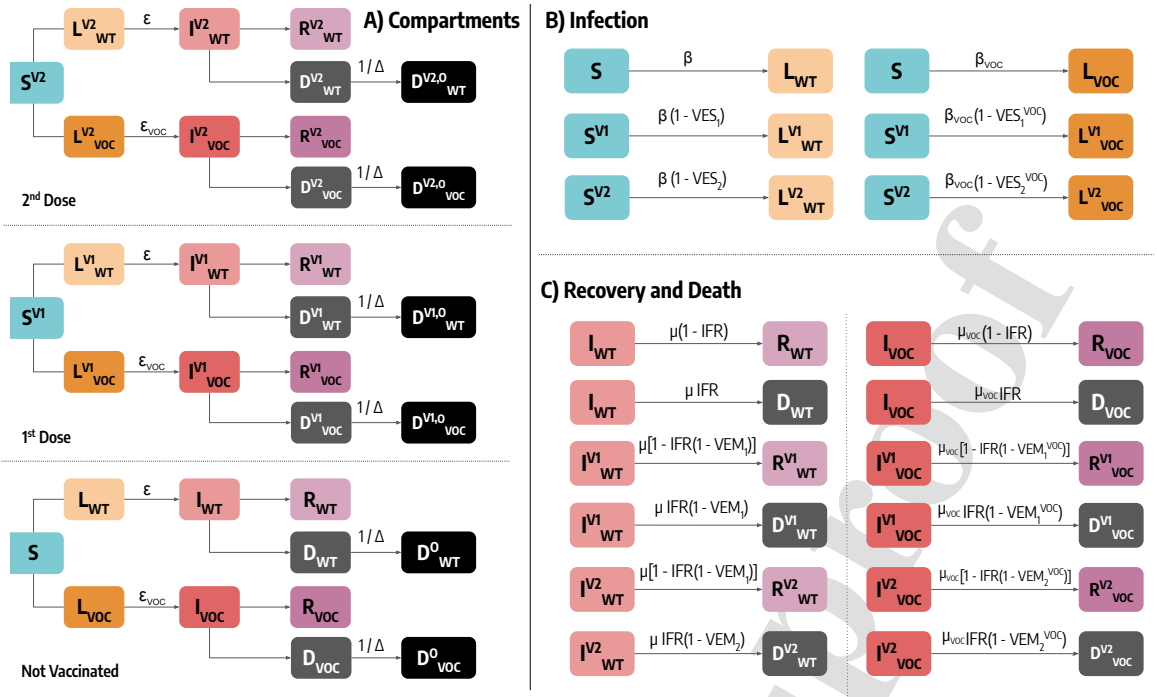

Figure 3: Model compartmental structure.

from the Oxford Coronavirus Government Response Tracker [49]. This index spans from a minimum of 0 (no measures) to a maximum of 3 (all levels required to close). We translate this into a contact reduction coefficient, represented as $\omega_{school}(t) = 1 - \texttt{C1 School closing}/3$, which we then apply to the school layer.

4.3 Modeling the introduction of VOCs

We model the introduction of the Alpha VOC in the countries considered using GLEAM, a global stochastic metapopulation model that simulates the mobility of people across more than 3,300 sub-populations in about 190 countries and territories [33–36]. Sub-populations correspond to the catchment area of major transportation hubs. Mobility among them accounts for both long-range movements related to global air travel and short-range commuting between adjacent sub-populations. For international airline travel, we used 2020 origin-destination data provided by the International Air Transport Association and Official Airline Guide (OAG) [50]. GLEAM was calibrated using early international importations of SARS-CoV-2 from China as well as the unfolding of COVID-19-related deaths in each country. The model also accounted for travel restrictions, mobility reductions, and government interventions adopted during the pandemic [36]. We use GLEAM to estimate the arrivals of Alpha exposed individuals, in the target countries, within each age group considering $\sim 300,000$ stochastic simulations. Since the first two specimens of the Alpha variant were collected on September 20 and 21, 2020 in London and Kent areas

and since UK sequenced about 5% of positive cases [37], we assumed an initial cluster of Alpha variant infections in week 38 of 2020 drawn from a Poisson distribution with a mean value of 40.

In the case of Omicron, instead, we assumed an initial seeding in the first week of October 2021, as suggested by preliminary evidences from phylogenetic analysis [51]. We sampled from a flat prior distribution ($\sigma = [10 - 1000]$) the number of initial seeds and distributed them in different age groups proportionally to their size (more details in Sec. 4.6).

4.4 Model initialization

In the case of the Alpha VOC we start simulations on 01/09/2020. The initial distribution of individuals in S, L_{WT} (wild type latent), I_{WT} (wild type infected), and R compartments, in each country, is obtained from GLEAM.

In the case of the Omicron VOC, our approach is as follows. First, we consider I_{rep} as the number of infections reported in the week prior to 01/05/2021 (our simulation starting date). The initial number of infected individuals in the model (distributed between I and E proportionally to the time individuals spend in each compartment), say I_{ini}, is calibrated within the uniform range $[I_{rep}, 10I_{rep}]$ to account for under-reporting. Similarly, for the R compartment, we place a total number of individuals between 1 and 20 times the cumulative infections reported up to 01/05/2021, which is also calibrated. We do not initialize individuals in vaccinated compartments, as South Africa had administered approximately 0.5 doses per 100 as of 01/05/2021.

4.5 Model implementation

Model implementation is stochastic and transitions among compartments are modelled through chain binomial processes. In practice, the number of individuals transitioning from compartment X_k to Y_k at time t is sampled from a binomial distribution $Pr^{Bin}(X_k(t), p_{X_k \to Y_k}(t))$, where $p_{X_k \to Y_k}(t)$ is the transition probability. In case of multiple possible transitions from compartment X (e.g., $X \to Y$ and $X \to Z$) we consider a multinomial distribution $Pr^{Mult}(X_k(t), p_{X_k \to Y_k}(t), p_{X_k \to Z_k}(t))$, where $p_{X_k \to Y_k}(t)$ and $p_{X_k \to Z_k}(t)$ are the two transition probabilities. We consider an integration step δt of 1 day.

4.6 Model calibration

The proposed epidemic models are calibrated using an Approximate Bayesian Computation (ABC) technique [52, 53]. In particular, we use a rejection algorithm that works as follows. First, we define a prior distribution $P(\boldsymbol{\theta})$ on model's free parameters $\boldsymbol{\theta}$. Then, at each iteration, a parameter set $\boldsymbol{\theta}^*$ is sampled from $P(\boldsymbol{\theta})$ and an instance of the model is generated from it. An output quantity ϕ' of the model is compared to the corresponding real quantity ϕ via a distance metric $d(\phi, \phi')$. If $d(\phi, \phi')$ is smaller than a predefined tolerance δ, then $\boldsymbol{\theta}^*$ is accepted, otherwise is rejected. This process is repeated

until n parameter sets are accepted. The distribution of accepted $\boldsymbol{\theta}^*$ will be an approximation of the true posterior distribution. Here, we consider the weighted mean absolute percentage error ($wMAPE$) as distance metric.

With regards to the Alpha variant, we conducted model calibration for the seven countries under consideration during the time frame $01/09/2020 - 14/02/2021$ considering weekly deaths as output quantity. This allowed us to estimate posterior distributions of key free parameters including the reproductive number, seasonality, and initial conditions (see Tab. 2) for a fixed value of Alpha increased transmissibility η. Following this calibration, we generated out-of-sample projections on the growth of Alpha up to 08/05/2021 sampling from the obtained posterior distributions. In the out-of-sample period, we assumed three different NPIs scenarios. The one presented in the main text, assumes a *status quo* situation in which contact remains unchanged with respect to the last calibration week. In the Supplementary Information, we propose two additional scenarios describing a *conservative* and a *moderate* relaxation of social distancing and other NPIs.

With regards to the Omicron variant, we conducted a two-stage calibration process. We first fitted the model to weekly reported deaths in South Africa allowing just for one strain in the period May 1^{st}, 2021 – November 23^{rd}, 2021. As mentioned, starting in May 2021, the country experienced a third pandemic wave fueled by the Delta VOC [19]. This variant was able to replace the Beta VOC, which was responsible for the second wave. Through this first calibration step we obtained the posterior distributions for transmissibility, delay between deaths and their reporting, initial conditions, seasonality, infection fatality rates multiplier respect to the estimates from Ref. [45], and under-reporting in deaths. We then studied the impact of Omicron by assuming an initial seeding in the first week of October 2021. Preliminary evidences from phylogenetic analysis suggested the median date of the common ancestor, of all available Omicron samples, in early October (90% CI: [30 September - 20 October 2021]) [51]. As we did not have information about the number of initial seeds, we sampled a flat prior distribution of initial Omicron seeds. We distributed them in different age groups proportionally to their size. We defined Omicron's transmissibility advantage by setting $\eta = R_0^{Omicron}/R_0^{Delta}$ as the ratio of the basic reproductive numbers for Omicron and Delta. The immune evasion of Omicron was introduced with a single factor ν which describes the reduction of vaccine efficacy and protection from reinfection, as described in Sec. 4.1. We explored the $\boldsymbol{\eta} \times \boldsymbol{\nu} \times \boldsymbol{\sigma}$ parameter space and applied a second ABC calibration to select values compatible with the number of confirmed cases up to December 14^{th}, 2021. This choice was made because deaths, as a lagging indicator, had not yet fully reflected the impact of Omicron at that time. For η and ν we explore, uniformly, the range $[0.5, 3.0]$ and $[0\%, 100\%]$ respectively. For σ, we explore uniformly the values 10, 50, 500, and 1000.

In Table 2, we provide a list of parameters considered in our models for both the Omicron and Alpha VOC, including values from existing literature as well as those calibrated for the study along with

	Alpha	**Omicron**
R_t^{start}	$U(0.8, 2.0)$	$U(1.0, 2.5)$
α_{min}	$U(0.5, 1.0)$	$U(0.5, 1.0)$
Δ	$U(12, 25)$	$U(10, 30)$
initial infected	GLEAM	1 to 10 times infections reported in week prior to start simulation
initial recovered	GLEAM	1 to 20 times infections reported up to start simulation
% deaths reported	N/A	$U(25\%, 100\%)$
IFR multiplier	N/A	$U(0.5, 2.0)$
Increased transmissibility (ν)	1.5 (main text) 1.3, 1.7 (SI)	$U(0.5, 3.0)$
Immune escape potential (η)	N/A	$U(0\%, 100\%)$
Initial VOC seeds (σ)	GLEAM	$U(10, 1000)$
Latent period	WT, Alpha: 4 days	WT: 3 days, Omicron: 3 and 2 days
Infectious period	WT, Alpha: 2.5 days	WT: 2.5 days, Omicron: 2.5 and 1.5 days
Infection Fatality Rate (age-stratified)	Ref. [45]	Ref. [45]
VE_1	N/A	80%
VES_1	N/A	70%
VE_2	N/A	90%
VES_2	N/A	80%
VEI	N/A	40%
Δ_V	N/A	14 days

Table 2: **Model parameters.** WT stands for *wild type*, that we use to indicate the previously circulating strain with respect to the VOC under examination. N/A stands for not applicable reflecting some additional parameters used for Omicron but not for Alpha.

their prior distributions. For a more in-depth discussion of the model's parameters, we direct readers to the Supplementary Information, where we also report the estimated posterior distributions of key free parameters for both Alpha and Omicron. In the Supplementary Information, we also show fitted deaths and cases resulting from the calibrations. We note how, in the case of Alpha, we consider a latent period of $\epsilon^{-1} = 4$ days and an infectious period $\mu^{-1} = 2.5$ days for both Alpha VOC and the previously circulating strain [54, 55]. When studying the emergence of Omicron instead, we consider $\epsilon^{-1} = 3$ days and $\mu^{-1} = 2.5$ days for the previously circulating strain (Delta) [56], while we test two possible parameters combination for the Omicron VOC (i.e., the second strain): $\epsilon^{-1} = 3$ days and $\mu^{-1} = 2.5$ days (i.e., same as Delta) and $\epsilon^{-1} = 2$ days and $\mu^{-1} = 1.5$ days (shorter generation time as suggested by some works [21]).

4.7 Estimating the date of dominance from sequencing data

We considered sequencing data obtained from GISAID [25], accessed through the ECDC website [57] for the Alpha VOC, and from CoVariants [58] for the Omicron VOC. This data provides information on the proportion of processed samples attributed to each specific virus variant in each country. Let $s_{VOC}(t)$ represent the fraction of samples associated with the VOC of interest at time t. To obtain more accurate estimates of the date of dominance and reduce the impact of noise, for each country we fitted a logistic

curve of the form $\hat{s}_{VOC}(t) = 1/(1+e^{-\gamma(t-t_{1/2})})$ to the actual fractions of VOC samples $s_{VOC}(t)$. Similar approaches have been widely applied in the context of SARS-CoV-2 VOCs growth modeling [59, 60]. The fitting process was performed via least square using the python library *scipy* [61]. By employing this approach, we determined the estimated value of $t_{1/2}$ (the time at which the VOC achieves 50% dominance) as the actual date of dominance.

Acknowledgments

M.C., M.A., and A.V. acknowledge support from COVID Supplement CDC-HHS-6U01IP001137-01. M.C. and A.V. acknowledge support from Google Cloud and Google Cloud Research Credits program to fund this project. The findings and conclusions in this study are those of the authors and do not necessarily represent the official position of the funding agencies, the National Institutes of Health, or the U.S. Department of Health and Human Services.

References

[1] James Hadfield, Colin Megill, Sidney M Bell, John Huddleston, Barney Potter, Charlton Callender, Pavel Sagulenko, Trevor Bedford, and Richard A Neher. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics, 34(23):4121–4123, 2018.

[2] Zhanwei Du, Lin Wang, Bingyi Yang, Sheikh Taslim Ali, Tim K Tsang, Songwei Shan, Peng Wu, Eric HY Lau, Benjamin J Cowling, and Lauren Ancel Meyers. Risk for international importations of variant SARS-CoV-2 originating in the United Kingdom. Emerging Infectious Diseases, 27(5):1527, 2021.

[3] Matteo Chinazzi, Jessica T Davis, Kunpeng Mu, A Pastore y Piontti, Nicola Perra, Samuel V Scarpino, and Alessandro Vespignani. Preliminary estimates of the international spreading risk associated with the SARS-CoV-2 VUI 202012/01, 2020.

[4] Tara Alpert, Anderson F Brito, Erica Lasek-Nesselquist, Jessica Rothman, Andrew L Valesano, Matthew J MacKay, Mary E Petrone, Mallery I Breban, Anne E Watkins, Chantal BF Vogels, et al. Early introductions and transmission of SARS-CoV-2 variant B. 1.1. 7 in the United States. Cell, 184(10):2595–2604, 2021.

[5] Erik Volz, Swapnil Mishra, Meera Chand, Jeffrey C Barrett, Robert Johnson, Lily Geidelberg, Wes R Hinsley, Daniel J Laydon, Gavin Dabrera, Áine O'Toole, et al. Assessing transmissibility of SARS-CoV-2 lineage B. 1.1. 7 in England. Nature, 593(7858):266–269, 2021.

[6] Thomas Y Michaelsen, Marc Bennedbæk, Lasse E Christiansen, Mia SF Jørgensen, Camilla H Møller, Emil A Sørensen, Simon Knutsson, Jakob Brandt, Thomas BN Jensen, Clarisse Chiche-Lapierre, et al. Introduction and transmission of SARS-CoV-2 lineage B. 1.1. 7, Alpha variant, in Denmark. Genome Medicine, 14(1):47, 2022.

[7] Nicholas G Davies, Sam Abbott, Rosanna C Barnard, Christopher I Jarvis, Adam J Kucharski, James D Munday, Carl AB Pearson, Timothy W Russell, Damien C Tully, Alex D Washburne, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B. 1.1. 7 in England. Science, 372(6538):eabg3055, 2021.

[8] Kathy Leung, Marcus HH Shum, Gabriel M Leung, Tommy TY Lam, and Joseph T Wu. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Eurosurveillance, 26(1):2002106, 2021.

[9] Paola Stefanelli, Filippo Trentini, Giorgio Guzzetta, Valentina Marziano, Alessia Mammone, Monica Sane Schepisi, Piero Poletti, Carla Molina Grané, Mattia Manica, Martina Del Manso, et al. Co-circulation of SARS-CoV-2 alpha and gamma variants in Italy, February and march 2021. Eurosurveillance, 27(5):2100429, 2022.

[10] John T McCrone, Verity Hill, Sumali Bajaj, Rosario Evans Pena, Ben C Lambert, Rhys Inward, Samir Bhatt, Erik Volz, Christopher Ruis, Simon Dellicour, et al. Context-specific emergence and growth of the SARS-CoV-2 Delta variant. Nature, 610(7930):154–160, 2022.

[11] Joseph L-H Tsui, John T McCrone, Ben Lambert, Sumali Bajaj, Rhys PD Inward, Paolo Bosetti, Rosario Evans Pena, Houriiyah Tegally, Verity Hill, Alexander E Zarebski, et al. Genomic assessment of invasion dynamics of SARS-CoV-2 Omicron BA. 1. Science, 381(6655):336–343, 2023.

[12] Moritz UG Kraemer, Verity Hill, Christopher Ruis, Simon Dellicour, Sumali Bajaj, John T McCrone, Guy Baele, Kris V Parag, Anya Lindström Battle, Bernardo Gutierrez, et al. Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B. 1.1. 7 emergence. Science, 373(6557):889–895, 2021.

[13] Benjamin Faucher, Chiara E Sabbatini, Peter Czuppon, Moritz UG Kraemer, Philippe Lemey, Vittoria Colizza, François Blanquart, Pierre-Yves Boëlle, and Chiara Poletto. Drivers and impact of the early silent invasion of SARS-CoV-2 Alpha. Nature Communications, 15(1):2152, 2024.

[14] Nicolò Gozzi, Matteo Chinazzi, Jessica T Davis, Kunpeng Mu, Ana Pastore y Piontti, Marco Ajelli, Nicola Perra, and Alessandro Vespignani. Estimating the spreading and dominance of SARS-CoV-2 VOC 202012/01 (lineage B. 1.1. 7) across Europe. medRxiv, pages 2021–02, 2021.

[15] Nicolò Gozzi, Matteo Chinazzi, Jessica T Davis, Kunpeng Mu, Ana Pastore y Piontti, Alessandro Vespignani, and Nicola Perra. Preliminary modeling estimates of the relative transmissibility and

immune escape of the Omicron SARS-CoV-2 variant of concern in South Africa. medRxiv, pages 2022–01, 2022.

[16] Ewen Callaway. Heavily mutated coronavirus variant puts scientists on alert. Nature.

[17] Implications of the further emergence and spread of the SARS-CoV-2 B.1.1.529 variant of concern (Omicron) for the EU/EEA – first update. https://www.ecdc.europa.eu/sites/default/files/documents/threat-assessment-covid-19-emergence-sars-cov-2-variant-omicron-december-2021.pdf, 2021. Accessed: 2021-12-02.

[18] Ewen Callaway. Tracking COVID-19 variant Omicron. Nature, 2021.

[19] Juliet RC Pulliam, Cari van Schalkwyk, Nevashan Govender, Anne von Gottberg, Cheryl Cohen, Michelle J Groome, Jonathan Dushoff, Koleka Mlisana, and Harry Moultrie. Increased risk of sars-cov-2 reinfection associated with emergence of omicron in south africa. Science, 376(6593):eabn4947, 2022.

[20] Nicolò Gozzi, Matteo Chinazzi, Natalie E Dean, Ira M Longini Jr, M Elizabeth Halloran, Nicola Perra, and Alessandro Vespignani. Estimating the impact of COVID-19 vaccine inequities: a modeling study. Nature Communications, 14(1):3272, 2023.

[21] Sam Abbott, Katharine Sherratt, Moritz Gerstung, and Sebastian Funk. Estimation of the test to test distribution as a proxy for generation interval distribution for the Omicron variant in England. MedRxiv, pages 2022–01, 2022.

[22] Wan Yang and Jeffrey L Shaman. COVID-19 pandemic dynamics in South Africa and epidemiological characteristics of three variants of concern (Beta, Delta, and Omicron). Elife, 11, 2022.

[23] Francesco Menegale, Mattia Manica, Agnese Zardini, Giorgio Guzzetta, Valentina Marziano, Valeria d'Andrea, Filippo Trentini, Marco Ajelli, Piero Poletti, and Stefano Merler. Evaluation of Waning of SARS-CoV-2 Vaccine–Induced Immunity: A Systematic Review and Meta-analysis. JAMA Network Open, 6(5):e2310650–e2310650, 2023.

[24] Raquel Viana, Sikhulile Moyo, Daniel G Amoako, Houriiyah Tegally, Cathrine Scheepers, Christian L Althaus, Ugochukwu J Anyaneji, Phillip A Bester, Maciej F Boni, Mohammed Chand, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. Nature, 603(7902):679–686, 2022.

[25] Stefan Elbe and Gemma Buckland-Merrett. Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Challenges, 1(1):33–46, 2017.

[26] PHE. Investigation of novel SARS-CoV-2 variant. Variant of Concern 202012/01. Technical briefing 3. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950823/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3_-_England.pdf, 2021. [Online; accessed 13-January-2021].

[27] N. Davies, R.C. Barnard, C.I. Jarvis, A.J. Kucharski, J.D. Munday, C.A.B. Pearson, T.W. Russell, D.C. Tully, S. Abbott, A. Gimma, W. Waites, K.L.M. Wong, K. van Zandvoort, CMMID COVID-19 working group, R.M. Eggo, S. Funk, M. Jit, Atkins K.E., and W.J. Edmunds. Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. https://cmmid.github.io/topics/covid19/uk-novel-variant.html, 2020. [Online; accessed 23-December-2020].

[28] Science. Mutant coronavirus in the United Kingdom sets off alarms, but its importance remains unclear. https://www.sciencemag.org/news/2020/12/mutant-coronavirus-united-kingdom-sets-alarms-its-importance-remains-unclear, 2020. [Online; accessed 21-December-2020].

[29] Andrew Rambaut, Nick Loman, Oliver Pybus, Wendy Barclay, Jeff Barrett, Alesandro Carabelli, Tom Connor, Tom Peacock, David L Robertson, Erik Volz, on behalf of COVID-19 Genomics Consortium UK (CoG-UK)9. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563, 2020. [Online; accessed 21-December-2020].

[30] NERVTAG. NERVTAG meeting on SARS-CoV-2 variant under investigation VUI-202012/01. https://khub.net/documents/135939561/338928724/SARS-CoV-2+variant+under+investigation%2C+meeting+minutes.pdf/962e866b-161f-2fd5-1030-32b6ab467896?t=1608470511452, 2020. [Online; accessed 21-December-2020].

[31] ECDC. Rapid increase of a SARS-CoV-2 variant with multiple spike protein mutations observed in the United Kingdom. https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf, 2020. [Online; accessed 21-December-2020].

[32] Covid: New lockdowns for England and Scotland ahead of 'hardest weeks'. https://www.bbc.com/news/uk-55538937, 2021. [Online; accessed 5-January-2021].

[33] M Chinazzi, J.T. Davis, K. Mu, A. Pastore y Piontti, N. Perra, S.V. Scarpino, and A. Vespignani. Preliminary estimates of the international spreading risk associated with

the SARS-CoV-2 VUI 202012/01. https://www.mobs-lab.org/uploads/6/7/8/7/6787877/covid_19_uk_new_strain__3_.pdf, 2020. [Online; accessed 19-January-2021].

[34] Balcan, Duygu and Gonçalves, Bruno and Hu, Hao and Ramasco, José J. and Colizza, Vittoria and Vespignani, Alessandro. Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. Journal of Computational Science, 1(3):132–145, aug 2010.

[35] Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. Science, 368(6489):395–400, 2020.

[36] Jessica T Davis, Matteo Chinazzi, Nicola Perra, Kunpeng Mu, Ana Pastore y Piontti, Marco Ajelli, Natalie E Dean, Corrado Gioannini, Maria Litvinova, Stefano Merler, Luca Rossi, Kaiyuan Sun, Xinyue Xiong, Ira M Longini, M Elizabeth Halloran, Cécile Viboud, and Alessandro Vespignani. Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave. Nature, 600(7887):127–132, 2021.

[37] WHO. SARS-CoV-2 Variant – United Kingdom of Great Britain and Northern Ireland. https://www.who.int/csr/don/21-december-2020-sars-cov2-variant-united-kingdom/en/, 2020. [Online; accessed 21-December-2020].

[38] Áine O'Toole, Verity Hill, Oliver G Pybus, Alexander Watts, Issac I Bogoch, Kamran Khan, Jane P Messina, The COVID, Brazil-UK CADDE Genomic Network, Houriiyah Tegally, et al. Tracking the international spread of SARS-CoV-2 lineages B. 1.1. 7 and B. 1.351/501Y-V2 with grinch. Wellcome Open Research, 6, 2021.

[39] Erik Volz. Fitness, growth and transmissibility of SARS-CoV-2 genetic variants. Nature Reviews Genetics, pages 1–11, 2023.

[40] Google LLC "Google COVID-19 Community Mobility Reports". https://www.google.com/covid19/mobility/, 2020. Accessed: 2021-08-01.

[41] Jack Wardle, Sangeeta Bhatia, Moritz UG Kraemer, Pierre Nouvellet, and Anne Cori. Gaps in mobility data and implications for modelling epidemic spread: a scoping review and simulation study. Epidemics, page 100666, 2023.

[42] Shengjie Lai, Nick W. Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R. Floyd, Amy Wesolowski, Mauricio Santillana, Chi Zhang, Xiangjun Du, Hongjie Yu, and Andrew J. Tatem. Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature, 2020.

[43] Jianhong Wu, Kathy Leung, and Gabriel Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in wuhan, china: a modelling study. The Lancet, 395, 01 2020.

[44] Dina Mistry, Maria Litvinova, Matteo Chinazzi, Laura Fumanelli, Marcelo FC Gomes, Syed A Haque, Quan-Hui Liu, Kunpeng Mu, Xinyue Xiong, M Elizabeth Halloran, et al. Inferring high-resolution human mixing patterns for disease modeling. Nature Communications, 2021.

[45] Robert Verity, Lucy Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick Walker, Han Fu, Amy Dighe, Jamie Griffin, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Anne Cori, Zulma M. Cucunubá, Rich FitzJohn, Katy Gaythorpe, and Neil Ferguson. Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases, 20, 03 2020.

[46] Ben S Cooper, Richard J Pitman, W John Edmunds, and Nigel J Gay. Delaying the international spread of pandemic influenza. PLoS Med, 3(6):e212, 2006.

[47] Julia Shapiro, Natalie E. Dean, Zachary J. Madewell, Yang Yang, M.Elizabeth Halloran, and Ira Longini. Efficacy Estimates for Various COVID-19 Vaccines: What we Know from the Literature and Reports. medRxiv, 2021.

[48] Edouard Mathieu, Hannah Ritchie, Esteban Ortiz-Ospina, Max Roser, Joe Hasell, Cameron Appel, Charlie Giattino, and Lucas Rodés-Guirao. A global database of COVID-19 vaccinations. Nature Human Behaviour, 5(7):947–953, 2021.

[49] Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nature human behaviour, 5(4):529–538, 2021.

[50] OAG, Aviation Worlwide Limited, 2020. `http://www.oag.com/`.

[51] Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. `https://ceri.africa/publication/?token=369`, 2021. Accessed: 2021-12-14.

[52] Amanda Minter and Renata Retkute. Approximate bayesian computation for infectious disease modelling. Epidemics, 29:100368, 2019.

[53] Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. Approximate Bayesian Computation. PLOS Computational Biology, 9(1):1–10, 01 2013.

[54] Jantien A Backer, Don Klinkenberg, and Jacco Wallinga. Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from wuhan, china, 20–28 january 2020. Eurosurveillance, 25(5), 2020.

[55] Stephen M. Kissler, Christine Tedijanto, Edward Goldstein, Yonatan H. Grad, and Marc Lipsitch. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science, 368(6493):860–868, 2020.

[56] Baisheng Li, Aiping Deng, Kuibiao Li, Yao Hu, Zhencui Li, Yaling Shi, Qianling Xiong, Zhe Liu, Qianfang Guo, Lirong Zou, et al. Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. Nature communications, 13(1):460, 2022.

[57] ECDC. Data on SARS-CoV-2 variants in the EU/EEA . `https://www.ecdc.europa.eu/en/publications-data/data-virus-variants-covid-19-eueea`.

[58] Emma B. Hodcroft. CoVariants: SARS-CoV-2 Mutations and Variants of Interest. `https://covariants.org`, 2021.

[59] Erik Volz, Verity Hill, John T McCrone, Anna Price, David Jorgensen, Áine O'Toole, Joel Southgate, Robert Johnson, Ben Jackson, Fabricia F Nascimento, et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. Cell, 184(1):64–75, 2021.

[60] Nicole L Washington, Karthik Gangavarapu, Mark Zeller, Alexandre Bolze, Elizabeth T Cirulli, Kelly M Schiabor Barrett, Brendan B Larsen, Catelyn Anderson, Simon White, Tyler Cassens, et al. Emergence and rapid transmission of SARS-CoV-2 B. 1.1. 7 in the United States. Cell, 184(10):2587–2594, 2021.

[61] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17:261–272, 2020.

Nicolò Gozzi: Conceptualization, Data curation, Investigation, Writing – original draft. **Matteo Chinazzi:** Data curation, Investigation, Writing – review and editing. **Jessica T. Davis:** Data curation, Investigation, Writing – review and editing. **Kunpeng Mu:** Data curation, Investigation, Writing – review and editing. **Ana Pastore y Piontti:** Data curation, Investigation, Writing – review and editing. **Marco Ajelli:** Conceptualization, Data curation, Investigation, Writing – review and editing. **Alessandro Vespignani:** Conceptualization, Data curation, Investigation, Writing – original draft. **Nicola Perra:** Conceptualization, Data curation, Investigation, Writing – original draft.

Declaration of Interest

M.C., M.A., and A.V. acknowledge support from COVID Supplement CDC-HHS-6U01IP001137-01. M.C. and A.V. acknowledge support from Google Cloud and Google Cloud Research Credits program to fund this project. The findings and conclusions in this study are those of the authors and do not necessarily represent the official position of the funding agencies, the National Institutes of Health, or the U.S. Department of Health and Human Services.