A two-variant model of SARS-COV-2 transmission: estimating the characteristics of a newly emerging strain

Teresa K. Yamana[1*], Swetha Rajagopal[2], Donald C. Hall Jr.[3], Ahmed M. Moustafa[2,4], Andries Feder[2], Azad Ahmed[3], Colleen Bianco[2], Rebecca Harris[2], Susan Coffin[2], Amy E. Campbell[2], Sen Pei[1], Joshua Chang Mell[3], Paul J. Planet[2,4*] and Jeffrey Shaman[1]

Abstract

Background The Covid-19 pandemic has been characterized by the emergence of novel SARS-CoV-2 variants, each with distinct properties influencing transmission dynamics, immune escape, and virulence, which, in turn, influence their impact on local populations. Swift analysis of the properties of newly emerged variants is essential in the initial days and weeks to enhance readiness and facilitate the scaling of clinical and public health system responses.

Methods This paper introduces a two-variant metapopulation compartmental model of disease transmission to simulate the dynamics of disease transmission during a period of transition to a newly dominant strain. Leveraging novel S-gene dropout analysis data and genomic sequencing data, combined with confirmed Covid-19 case data, we estimate the epidemiological characteristics of the Omicron variant, which replaced the Delta variant in late 2021 in Philadelphia, PA. We utilized a grid-search method to identify plausible combinations of model parameters, followed by an ensemble adjustment Kalman filter for parameter inference.

Results The model successfully estimated key epidemiological parameters; we estimated the ascertainment rate of 0.22 (95% credible interval 0.15–0.29) and transmission rate of 5.0 (95% CI 2.4–6.6) for the Omicron variant.

Conclusions The study demonstrates the potential for this model-inference framework to provide real-time insights during the emergence of novel variants, aiding in timely public health responses.

Keywords SARS-COV-2, Infectious disease modelling, Parameter estimation

*Correspondence:
Teresa K. Yamana
tky2104@cumc.columbia.edu
Paul J. Planet
planetp@chop.edu
[1]Columbia University, New York, NY, USA
[2]Children's Hospital of Philadelphia, Philadelphia, PA, USA
[3]Drexel University, Philadelphia, PA, USA
[4]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

Background

The Covid-19 pandemic has been marked by numerous waves, many of which were driven by new variants of SARS-CoV-2. The specific characteristics of these new variants determine levels of immune escape, transmissibility, and virulence, which, in turn, influence their impact on local populations [1–3]. The B.1.617.2 (Delta) variant, for example, was characterized by increased transmissibility and virulence, and decreased vaccine efficacy against symptomatic infection compared to previously circulating strains. This led to resurgences in cases, hospitalizations, and deaths worldwide [4–7].

The largest wave of Covid-19 cases to date was driven by the emergence of the B.1.1.529 (Omicron) variant, which was first identified in the United States on December 1, 2021 [8]. This variant featured multiple mutations on the spike protein, leading to a high rate of immune escape [9–11]. The Omicron variant was found to have a higher secondary attack rate than Delta even among unvaccinated cases and contacts, suggesting that the Omicron variant could have additional intrinsic properties leading to higher transmissibility compared to the Delta variant [12], which supported its rapid spread. The impact of Omicron was mitigated by its decreased severity compared to Delta [13, 14].

Rapid characterization of new variant properties is needed in the days and weeks immediately following emergence in order to better prepare and scale clinical and public health system response [15]. Infectious disease models, if properly integrated to real-time data and inference methods, provide one such means for estimating epidemiological characteristics such as transmissibility, immune escape, and incubation period in real time. Such systems have been developed to infer the characteristics of novel variants using case and mortality data [16].

Here, we present newly collected variant type data derived from sequencing and S-gene dropout analysis, used in conjunction with classical epidemiological surveillance data to produce an estimated time-series of Covid-19 cases by variant type. We develop and demonstrate the use of a two-strain model of disease transmission driven by these data to estimate the characteristics of Omicron as it replaced the Delta variant in late 2021 to early 2022 in the city of Philadelphia, PA. In the future, this model-inference framework could be deployed in real-time to provide insight during the emergence of other novel variants.

Methods.

Data.

Weekly confirmed cases of Covid-19 by ZIP code from October 4, 2020 through April 3, 2022 were obtained from the Philadelphia Department of Public Health. We grouped Philadelphia ZIP codes into 5 geographic regions, first by grouping according to the 18 planning analysis sections, then by merging adjacent sections into larger geographical units. The resulting regions are shown in Fig. 1. The timing and population-adjusted magnitude of case rates were very similar between locations. Approximately 8% of case data were missing ZIP code information; we distributed these to the 5 geographic regions proportionally to the data for which geographic information was available under the assumption that data with missing information were evenly distributed across the city.

The relative incidence of the two principal circulating variants during Dec 2021 through Jan 2022, Delta and Omicron, was determined using weekly data from S-gene target drop out and genomic sequencing analyses of 1157 SARS-CoV-2 samples collected from residents of Philadelphia by the Children's Hospital of Philadelphia and the Philadelphia Department of Public Health as part of a variant-surveillance project. S-gene drop out analyses was successfully applied to 1007 samples, and of these, 174 samples underwent whole-genome sequencing as well. An additional 122 samples were whole-genome sequenced, without S-gene drop out analysis. These data have been deposited at the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA).

Viral RNA from specimens were extracted using the ThermoFisher MagMAX™ Viral/Pathogen Nucleic Acid Isolation Kit on the ThermoFisher KingFisher extraction system following the assay standard protocol. S-gene dropout monitoring was performed using the ThermoFisher TaqPath™ COVID-19 Combo Kit (Cat No: A47814) on a QuantStudio 6 Plus. Amplification of the S-gene was monitored with the reporter dye ABY. A confirmed S-drop was determined from the noted amplification of ORF1ab (FAM) and N-gene (VIC) and absence of S-gene (ABY) amplification. Data were analyzed using ThermoFisher Design & Analysis software (version 2.6.0) with automated thresholding for Ct value determination.

Whole-genome sequencing was conducted at the Genomics Core Facility at Drexel University using the Paragon Genomics CleanPlex SARS-CoV-2 Research and Surveillance NGS Panel. Library preparation consisted of measuring concentration with the Qubit dsDNA High-Sensitivity Assay Kit, quality assessment using the Agilent High Sensitivity DNA Kit and 2100 Bioanalyzer instrument, and standardization to 8 nM. The libraries were again quantified, then diluted to a final concentration of 4 nM, and loaded onto the MiSeq system at 10 pM. Paired-end and dual-indexed 2×150-bp sequencing was performed using MiSeq Reagent Kits. Sequences were demultiplexed, and base calls were transformed into FASTQ format using bcl2fastq2, version 2.20. The FASTQ reads were subsequently processed to generate a consensus sequence, and variants were identified using
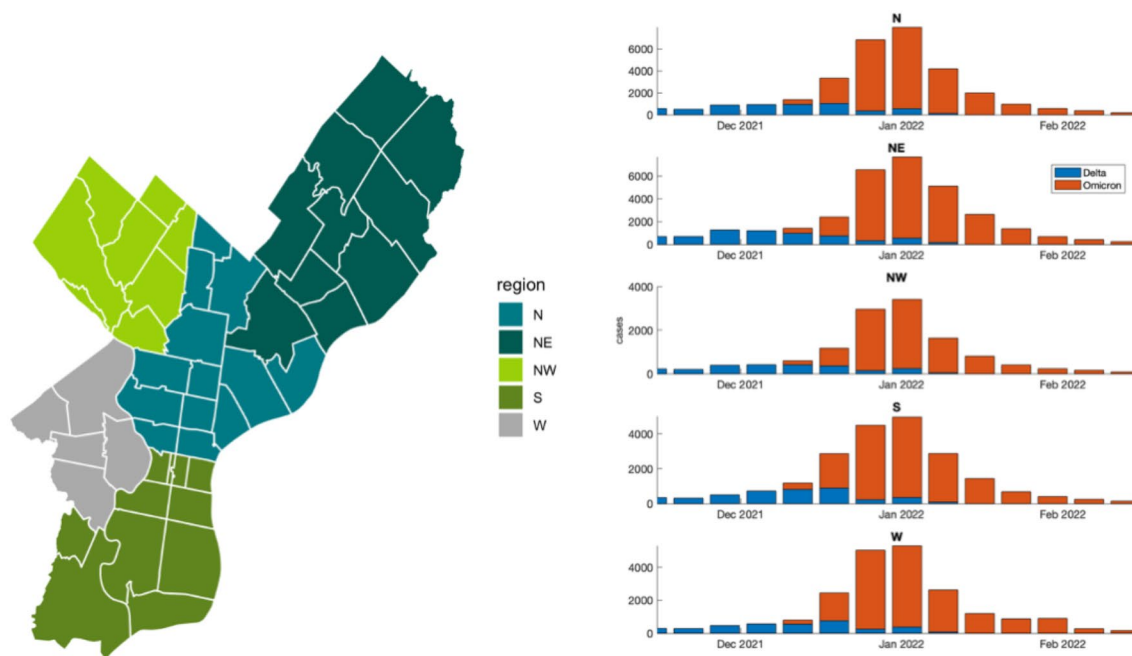
Fig. 1 The left image shows the ZIP codes of Philadelphia, outlined in white, divided into the 5 regions used in this analysis. The bar graphs on the right show the number of confirmed Covid-19 cases in each of the 5 regions. The colors indicate the estimated number of cases of each of the two variants

the ncov2019-artic-nf pipeline (https://github.com/connor-lab/ncov2019-artic-nf) [17].

The number of Covid-19 cases for each of the two variant types was estimated by multiplying the number of confirmed cases by the proportion of each variant in the genotyped samples for each corresponding week (Fig. 1; Supplemental Table 1). These time-series were interpolated to daily values using shape-preserving piecewise cubic interpolation, which then served as the observational input to the two-strain disease transmission model.

The average rate of daytime movement of individuals from their home geographic region to each of the other geographic regions was estimated using data on visits to geo-referenced points of interests from SafeGraph [18]. We aggregated these visits to the 5 geographic regions and then normalized the number of visits to estimate the number of individuals travelling to other parts of the city during the day.

Model description.

The time period for this analysis begins on November 12, 2021, several weeks prior to the first reported cases of the Omicron variant in December 2021. We modeled disease transmission using a two-strain metapopulation compartmental structure in order to explicitly represent transmission dynamics during this critical period of a newly emerging variant Each strain was modeled separately using the adapted form of a single-strain

metapopulation model previously employed to model Covid-19 and influenza transmission for a range of locations and spatial scales [19–22]. The population of Philadelphia was divided into subpopulations based on region of daytime location (e.g. work) and region of residence. Movement between the subpopulations (e.g. people living in region *i* and working in region *j)* was imposed based on the estimates derived from SafeGraph data.

We modeled the transmission dynamics of COVID-19 separately for day and night periods, reflecting the movement of individuals to daytime destination regions, the return home for the night, and mixing with the wider populations at both locations. Transmission was represented as a discrete Markov process with Susceptible (S), Exposed (E), Reported Infected (Ir), Unreported Infected (Iu), and Recovered (R) compartments. Here, we added a second Covid-19 variant to the model (Fig. 2). We assumed that all individuals susceptible to the original variant were also susceptible to the new variant, and that some additional individuals immune to the original variant are susceptible to the new variant due to immune escape. We assumed that infection by one strain over the course of the simulation confers full immunity to both strains. This immunity wanes over time; however, we assumed a long duration of immunity such that over the timescale of the simulation, the effect of reinfection is negligible. Model parameters are listed in Table 1. Full
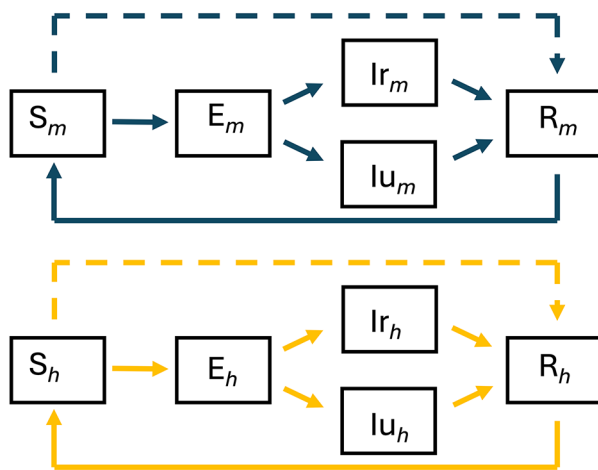
Fig. 2 Model schematic. The solid lines indicate individuals moving between through the Susceptible (S), Exposed (E), Reported Infected (Ir), Unreported Infected (Iu) and Recovered (R) compartments due to an infection of strains *m* (top) and *h* (bottom), respectively. The dashed lines indicate cross-immunity conferred by infection by the other strain (i.e. movement from S_m to R_m due to infection by strain *h* and vice-versa)

model equations and additional details are provided in the Supplementary Material.

At the start date of the analysis November 12, 2021, the predominant variant in circulation across the United States was SARS-CoV-2 B.1.617.2 (Delta), which accounted for an estimated 99.9% of transmission [8]. The Delta variant in our model was initialized using estimates of population susceptibility, infection rate, and parameter values for Philadelphia County as of November 12, 2021 from an operational county-resolved model of Covid-19 transmission throughout the US; these initial values are listed in Table 1 [23]. Initialization of the Omicron variant is described below.

Parameterization using grid search method.

We employed two methods to infer unknown model parameters α^h and β^h; (1) a grid search approach, and (2) a data assimilation approach. The grid search approach was a relatively simple analysis intended to narrow model uncertainty by identifying plausible combinations of parameters and state variables that produce outbreaks with comparable magnitude and speed of progression as observed. While this approach could be employed as a stand-alone method, in this study, we used it to provide initial conditions to our more sophisticated data assimilation method.

We limited parameter estimation to the ascertainment rate α^h and transmission rate β^h parameters of the emerging Omicron variant, as well as the model state variables for each strain and location. We focused on these two parameters as they are difficult to measure directly using traditional epidemiological approaches such as contact tracing. The remaining parameters were assigned and remained fixed over time; values are shown in Table 1.

We first used a grid search approach to explore the parameter space and identify optimal combinations of parameters and initial susceptibility by assessing the fit between model output and observations. We ran a set of 100 stochastic simulations for each of 1008 combinations of transmission rate (β), ascertainment rate (α), and initial susceptibility (S_0). Initial susceptibility to the Omicron variant was assigned values between 20 and 100% of the population in increments of 10%. The case reporting rate for the Omicron variant was assigned values from 5 to 40%, in increments of 5%. The transmission rate for Omicron was assigned values from 0.5 to 7 in increments

Table 1 List and description of model parameters and initial conditions, including values used for Grid Search analysis and initial conditions for inference system. Superscript *m* indicates parameters relating to the Delta strain and superscript *h* indicates parameters relating to the emerging omicron strain

Model parameters	Description	Value	Reference
α^m	Ascertainment rate	0.24	[23]
β^m	Transmission rate	2.2	[23]
S_0^m	Initial susceptibility to original strain	0.20–0.25	[23]
Z^m	Incubation period	3.6 days	[22]
D^m	Duration of infectiousness	3.5 days	[22]
L^m	Duration of immunity	5 years	n/a
α^h	Ascertainment rate	0.05–0.4	estimated
β^h	Transmission rate	0.5–7	estimated
S_0^h	Initial susceptibility to emerging strain	0.20–1.00	estimated
Z^h	Incubation period	3 days	[24–27]
D^h	Duration of infectiousness	3.5 days	n/a
L^h	Duration of immunity	5 years	n/a
μ	Relative infectiousness of asymptomatic infection	0.61	[22]
θ	Mobility factor	0.15	[22]

of 0.5. The corresponding basic reproductive number R_0 ranged from 1.1 to 18.8.

For each combination of parameters and initial conditions, we assessed the fit of the simulated reported cases to observations using the continuous rank probability score (CRPS). This measure considers the full cumulative distribution function (CDF) of the modeled probabilities and calculates the average discrepancy between this modeled CDF and the CDF of the observed variable.

Model fitting with Ensemble Adjustment Kalman Filter (EAKF).

We used the results of the grid search analysis to inform a Bayesian inference approach, the ensemble adjustment Kalman filter (EAKF), coupled with the transmission model to infer epidemiological parameters and state variables. This more sophisticated method provides more precise probabilistic estimates compared to the grid search analysis. The EAKF is a data assimilation method originally developed for weather forecasting [24], which has been successfully applied for parameter inference to infectious disease systems, including COVID-19 [19], influenza [25], and RSV [26]. We use an ensemble of 250 simulations in which parameters and state variables are initialized and optimized using the EAKF in a prediction-update cycle following each data input of reported Covid-19 cases. In the prediction step, the transmission model advances the state variables forward in time. This is followed by an update step in which the EAKF algorithm adjusts ensemble members to better match case observations. These adjustments are also applied to unobserved variables and parameters based on the prior ensemble covariance between the observed variable and the unobserved variables and parameters. The update step ensures that the posterior ensemble mean and variance match the predicted mean and variance according to Bayes theorem, assuming a Gaussian distribution. As with the grid search method, we limited our estimation to the ascertainment rate α and transmission rate β parameters of the emerging Omicron variant, and the state variables for each strain and location.

Synthetic testing of model-inference system.

We validated the ability of the EAKF model-inference system to infer epidemiological parameters and state variables by first conducting synthetic testing. We ran the model in free simulation with randomly drawn parameters and initial conditions to simulate daily observations in the 5 neighborhoods for each of the two circulating Covid-19 variants. We then added random noise to the model-generated daily Covid-19 case counts, representing observational noise. We repeated this process 20 times to generate a range of plausible outbreak trajectories. Subsequently, we employed the model-inference system using these synthetic observations as inputs to estimate the α and β parameters for the emerging

Omicron variant, as well as state variables for both variants over time. For this synthetic testing step, we did not use results from the grid-search analysis; rather, initial conditions were drawn from the wider parameter space listed in Table 1. We repeated this process 5 times for each outbreak trajectory, then pooled the results from the 5 realizations to obtain overall estimates.

We also repeated this synthetic validation procedure using the model to estimate α, β, and the latent period Z for the Omicron variant to test whether the system could simultaneously estimate all three parameters. However, the model was not able to reliably estimate Z. We therefore fixed a value of 3 days for Z, consistent with findings from investigations of outbreak clusters and transmission pairs [27–30].

Inference using observed data.

Following validation of the EAKF model-inference system, we used the same system with the location- and strain-specific observations, described above, to simulate Covid-19 transmission dynamics in Philadelphia during the period when the Omicron variant emerged and replaced Delta as the dominant strain. Our primary interest was to estimate the ascertainment rate (α) and transmission rate (β) parameters. We first selected initial conditions for the Omicron variant that were drawn from the plausible parameter space identified in the grid search analysis. We used an ensemble of 250 members and repeated the inference 25 times. We pooled the results from the 25 realizations to obtain overall estimates.

While the grid search analysis was useful for informing initial conditions of the model-inference system, it had the disadvantage of relying on observational data that would not have been available in the initial weeks of variant emergence. We therefore tested the model-inference system's ability to produce estimates without the narrowed initial conditions provided by the grid-search analysis. We repeated the parameter estimation while relaxing the constraint on the initial conditions for α. Initial conditions for β, and the initial susceptibility (S_0) for the Omicron variant were drawn from combinations of values identified in the grid search analysis while initial conditions for α were drawn randomly from a larger parameter space.

Finally, we repeated the inference using initial conditions for α, β, and S_0 for the Omicron variant, randomly drawn from a larger parameter space that would have been considered plausible in the early days of its arrival. These initial conditions are listed in Table 1.

Results
Parameterization using grid search method.

The continuous rank probability score (CRPS) for each combination of transmission rate, case reporting rate, and initial population susceptibility to Omicron is shown

in Fig. 3; low CRPS corresponds with the closest match to the observed outbreak trajectory. We found a clear relationship between combinations of the three parameters and variables and the fit between the resulting trajectories and observed Covid-19 cases. A tradeoff between S_0 and β exists; outbreaks resembling observations could be produced by simulations with high β and low initial susceptibility, or low β and high S_0. This result is not surprising, as disease transmission in a compartmental model is largely driven by the effective reproductive number, R_{eff}, which is proportional to the product of β times S_0.

We found that the lowest CRPS scores were achieved for values of β x S_0 between 1.2 and 1.8. (Fig. 4). Higher values of α generally resulted in higher CRPS (Figs. 3 and 4).

We arbitrarily set an upper limit of CRPS=100 to narrow the parameter space to the best fitting combinations. Of the 1008 combinations of α, β, and S_0 tested, 166 combinations had CRPS below this threshold.

Synthetic testing of model-inference system.

Our second estimation method used the dynamical disease transmission model coupled with the EAKF.

Before applying this estimation method to observed data, we first validated the model-inference system's ability to ascertain parameters and state variables from a set of 20 plausible synthetic truths. These results are shown in Fig. 5. The 'true' value of α and β used to generate the outbreaks are shown in yellow, and the green areas show the density of the estimates. The model-inference system was generally able to infer parameters.

Inference using observed data.

Initial conditions for α, β, and S_0 for the Omicron variant were drawn from the combinations of values identified in the grid search analysis. The median value for S_0 was 40%, corresponding to an additional 15–20% susceptibility with respect to immunity derived from infection from previous variants and vaccines. The model-inference system was able to produce a good fit of observed case data for all 5 neighborhoods (Fig. 6, Supplemental Figs. 1 and 2). We estimated the average values of the α and β parameters over the duration of the Omicron wave, from November 24, 2021 through February 11, 2022. The model estimated value for α during this period was 0.22 (95% CI 0.15–0.29); the estimate
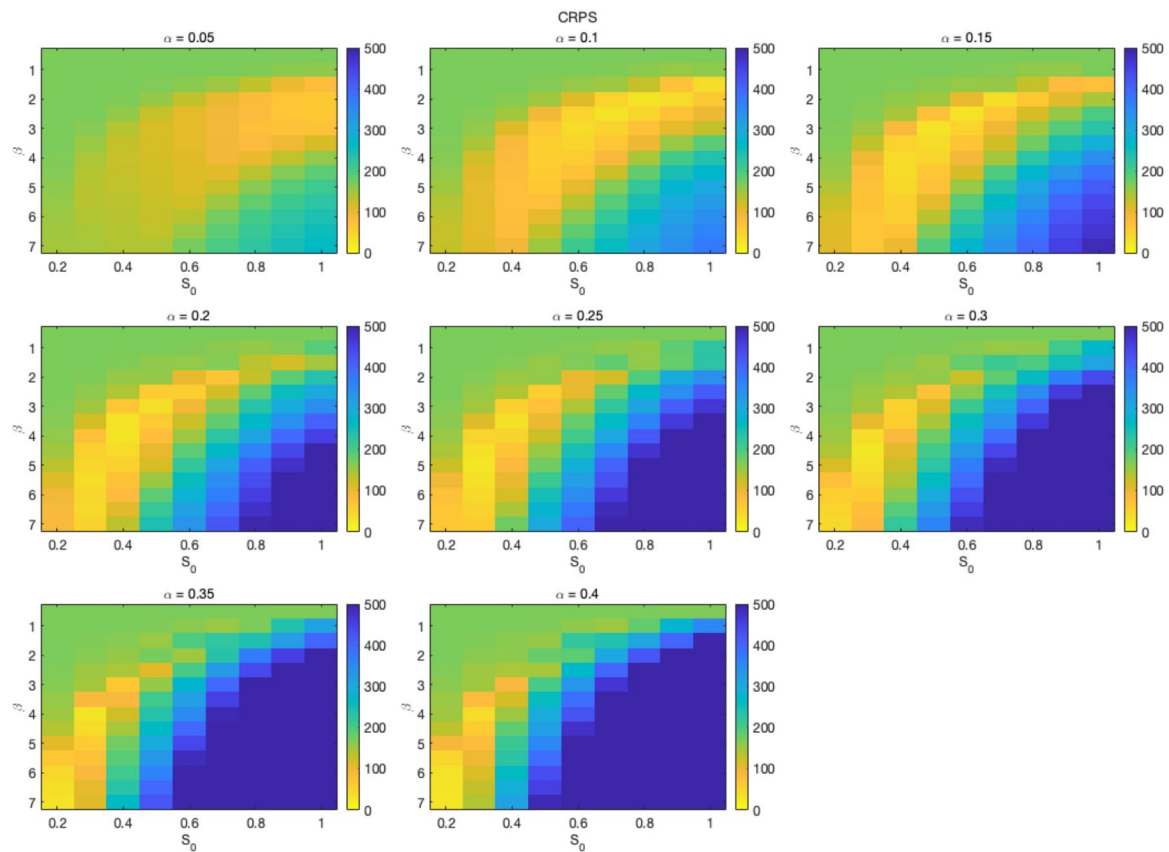

Fig. 3 Continuous rank probability score (CRPS) of modeled output with each combination of α, β and S_0. The color scale shows CRPS
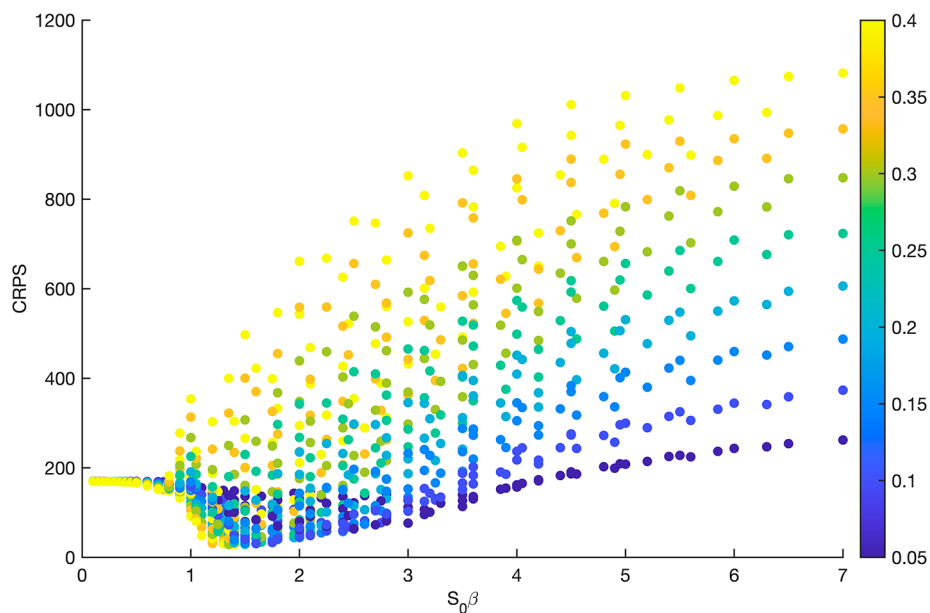
Fig. 4 CRPS for each combination of parameters. The product of $S_o \times \beta$ is on the x-axis and CRPS is on the y-axis. The colors indicate the value of α

for β was 5.0 (95% CI 2.4–6.6). The corresponding value of the time varying reproductive number Rt, a standard measure of disease transmission computed here as $R_t = \beta D[\alpha + \mu(1 - \alpha)]$, was 12.3 (95% CI 5.9–16.7) (Table 2).

We repeated the inference with the initial conditions for β, and S_0 for the Omicron variant drawn from combinations of values identified in the grid search analysis while initial conditions for α were drawn randomly from a larger parameter space. This relaxation of initial conditions led to similar outcomes and produced a good fit to observed case data (Supplemental Fig. 3, Table 2).

Finally, we repeated the inference with initial conditions for α, β, and S_0 for the Omicron variant were all drawn randomly from a wide range of plausible values. We found that this less constrained model-inference system was also able to produce a good fit to observed case data (Supplemental Fig. 4, Table 2). The estimate for β was 3.9 (95% CI 2.6–5.4). The corresponding value of the time varying reproductive number Rt was 9.5 (95% CI 6.3–13.3) These values are generally consistent with those estimated under with constrained initial conditions.

Discussion and Conclusion.

This study presents a framework for simultaneously modeling two variants of an infectious disease and estimating the parameters of the emerging variant using a combination of weekly confirmed positive cases, and newly collected S-drop PCR tests that distinguish Delta and Omicron, and genome sequencing analyses. We developed a two-strain model of disease transmission, and using a grid-search approach, identified combinations of parameters that would produce outbreaks similar

to what was observed in the weeks and months following the emergence of the Omicron variant in the city of Philadelphia. However, the disadvantage of the grid search method as applied here is that it required the benefit of hindsight and a relatively long time series of observations, and therefore could not be used in real-time to estimate the properties of a newly emerging variant.

Results from the grid search analysis were used to initialize a model-inference system that could better infer the underlying epidemiological parameters of the emerging Omicron variant as it replaced Delta. We found that both immune escape, as quantified by the additional susceptibility to Omicron compared Delta at the start of the Omicron wave, and enhanced transmissibility, as estimated by the difference in β between the two variants, contributed to the transmission advantage of Omicron. The values of the basic reproductive number for the Omicron variant estimated here are consistent with estimates compiled in a recent review of published estimates [31].

Finally, we showed that the model-inference system could be used to estimate parameters even without the constraints derived from the grid search analysis, indicating it could be applied before a larger record of the emerging outbreak is documented. While this study was conducted retrospectively, given the timely availability of case observations and serotype data, the methods illustrated here could be applied in real-time to estimate important epidemiological characteristics of an emerging variant. However, we note that these data are not always readily available in real-time. In practice, the case data collected by the Philadelphia Department of Public Health were subject to revisions over time as results
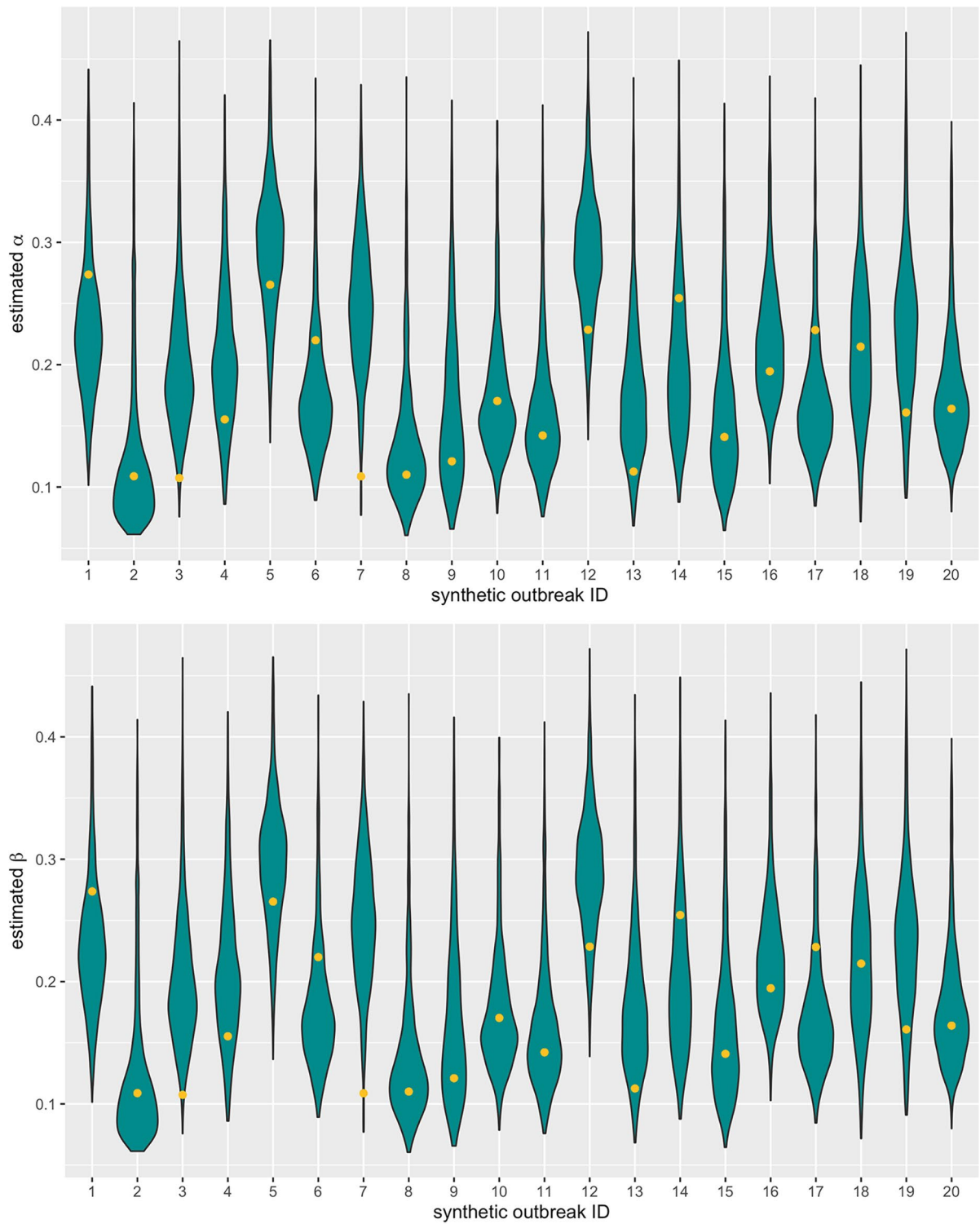
Fig. 5 Violin plots of estimated α (top panel) and β (bottom panel) parameters values for the Omicron variant for each of the 20 synthetic outbreaks. The 'true' value is shown in yellow, and the green areas show the density of the estimates. The x-axis enumerates 20 independent synthetic runs
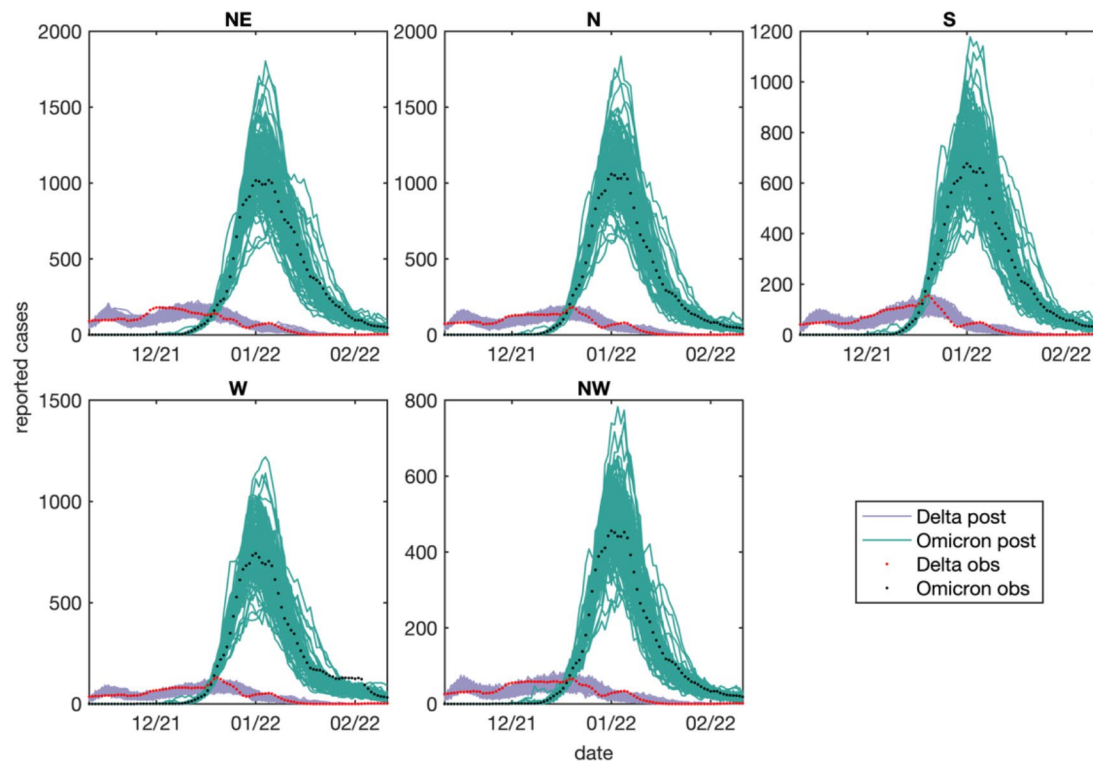
Fig. 6 Model posterior fit for Delta (purple) and Omicron (green) variants for each neighborhood, compared to reconstructed observed data (red and black for Delta and Omicron, respectively). These model posteriors show one of 25 model iterations; the process was repeated 25 times and the posteriors were pooled to compute overall parameter estimates. The model posteriors for the two variants individually are shown in Supplemental Figs. 1 and 2

Table 2 Estimated parameter values for the Omicron variant

Initialization	α	β	Rt
α, β, and S_0 from grid search	0.22 (0.15–0.29)	5.0 (2.4–6.6)	12.3 (5.9–16.7)
β, and S_0 from grid search, α random	0.21 (0.14–0.27)	4.6 (1.9–6.4)	11.2 (4.4–16.0)
α, β, and S_0 random (no grid search)	0.20 (0.14–0.27)	3.9 (2.6–5.4)	9.5 (6.3–13.3)

were processed and recorded; this is regularly observed in health systems throughout the country and worldwide, and is a known challenge to real-time applications [32, 33]. While correcting for data availability is beyond the scope of this analysis, there are several nowcasting approaches that have been developed to address missing data that could be applied in real-time (e.g. [32–36]).

Another limitation for the real-time application of the methods presented here is the model's reliance on independently published estimates of the incubation period *Z* for the emerging variant. The published estimates we used here were from investigations of individual outbreaks and were available from mid-December 2021 through January 2022 [27, 28, 30, 37]. In the absence of such information, we would have repeated the analysis for a range of plausible values of *Z*, and obtained estimates of the remaining parameters conditional on *Z*.

In this application, the Omicron variant spread rapidly throughout the city of Philadelphia, resulting in minimal spatial variability in cases. We expect that the utility of such a spatially resolved model would be more apparent for a less transmissible variant, for which the model-inference system could capture and shed light on spatial dynamics. Additionally, spatial variability would likely improve the performance of the model-inference system as the case observations from different locations would provide more independent data streams with which to fit the model.

Here, we used a combination of S-gene dropout and whole genome sequencing data. The S-gene dropout condition was a major advantage in identifying the Omicron variant as it provided a convenient and low-cost method to distinguish between variant types. If this type of rapid testing was not available, we would rely more on

sequence data, which would incur greater cost and longer lag times between sample collection and reported results.

Historically over the course of the SARS-CoV-2 pandemic, we have seen repeated instances in which a novel variant replaces the dominant circulating variant in circumstances similar to those observed in this study. The two-variant model framework presented here could be adapted to retrospectively represent the takeover dynamics of Alpha over the ancestral variant, and Alpha over Delta. In the case that the replacement dynamics are less straightforward, as has been observed in the years following the emergence of Omicron, we could expand and fit the model to allow for interactions between variants and thus co-circulation and competition of multiple strains. We believe this framework would be of particular interest for a novel variant with immune escape and high virulence.

Abbreviations

CRPS	Continuous rank probability score
CDF	Cumulative distribution function
S	Susceptible
E	Exposed
Ir	Reported Infected
Iu	Unreported Infected
R	Recovered
EAKF	Ensemble adjustment Kalman filter
R_{eff}	Effective reproductive number
R_t	Time-varying basic reproductive number

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-024-09823-x.

Supplementary Material 1

Author contributions
TY, JCM, PP and JS: conception; design of the work; the acquisition, analysis; interpretation of data; the creation of new software used in the work; drafted the work or substantively revised itSR: conception; acquisition, analysis; interpretation of data; DH: conception; design of the work; the acquisition, analysis; interpretation of data; drafted the work or substantively revised itAM and SP: the creation of new software used in the workAF: the acquisition, analysis; interpretation of data; the creation of new software used in the work; CB, RH, AC and AA: the acquisition, analysis; interpretation of data; SC: conception; design of the work; drafted the work or substantively revised itAll authors reviewed the manuscript.

Data availability
Sequence data presented is being uploaded to the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) with BioProject ID: PRJNA1072733Philadelphia COVID-19 case data are available on the OpenDataPhilly portal https://opendataphilly.org/datasets/covid-tests-and-cases/.

Declarations

Ethics approval and consent to participate
All experimental protocols were approved by the Children's Hospital of Pennsylvania IRB (IRB 21-018726). The study did not require direct interaction with human subjects and only involved review of medical records, with appropriate protections for storage and destruction of data. The IRB concluded that the study was minimal risk and was performed on specimens solely collected for non-research purposes that would have been discarded. Therefore, the CHOP IRB granted a waiver of consent. All data transmitted to outside institutions was completely deidentified and approved by the CHOP IRB.

Consent for publication
n/a.

Competing interests
JLS and Columbia University disclose partial ownership of SK Analytics. JLS discloses consulting for BNI. The other authors declare no competing interests.

References
1. Thakur V, Bhola S, Thakur P, Patel SKS, Kulshrestha S, Ratho RK, Kumar P. Waves and variants of SARS-CoV-2: understanding the causes and effect of the COVID-19 catastrophe. Infection 2021:1–16.
2. Marques AD, Sherrill-Mix S, Everett JK, Reddy S, Hokama P, Roche AM, Hwang Y, Glascock A, Whiteside SA, Graham-Wooten J. SARS-CoV-2 variants associated with vaccine breakthrough in the Delaware Valley through summer 2021. MBio. 2022;13(1):e03788–03721.
3. SeyedAlinaghi S, Mirzapour P, Dadras O, Pashaei Z, Karimi A, MohsseniPour M, Soleymanzadeh M, Barzegary A, Afsahi AM, Vahedi F. Characterization of SARS-CoV-2 different variants and related morbidity and mortality: a systematic review. Eur J Med Res. 2021;26(1):1–20.
4. Eyre DW, Taylor D, Purver M, Chapman D, Fowler T, Pouwels KB, Walker AS, Peto TE. Effect of Covid-19 vaccination on transmission of alpha and delta variants. N Engl J Med. 2022;386(8):744–56.
5. Shiehzadegan S, Alaghemand N, Fox M, Venketaraman V. Analysis of the delta variant B. 1.617. 2 COVID-19. Clin Pract. 2021;11(4):778–84.
6. Mlcochova P, Kemp SA, Dhar MS, Papa G, Meng B, Ferreira IA, Datir R, Collier DA, Albecka A, Singh S. SARS-CoV-2 B. 1.617. 2 Delta variant replication and immune evasion. Nature. 2021;599(7883):114–9.
7. Hoffmann M, Hofmann-Winkler H, Krüger N, Kempf A, Nehlmeier I, Graichen L, Arora P, Sidarovich A, Moldenhauer A-S, Winkler MS. SARS-CoV-2 variant B. 1.617 is resistant to bamlanivimab and evades antibodies induced by infection and vaccination. Cell Rep 2021, 36(3).
8. Covid C, Team R. Sars-cov-2 b. 1.1. 529 (omicron) variant—United States, December 1–8, 2021. Morb Mortal Wkly Rep. 2021;70(50):1731.
9. Saxena SK, Kumar S, Ansari S, Paweska JT, Maurya VK, Tripathi AK, Abdel-Moneim AS. Characterization of the novel SARS-CoV-2 omicron (B. 1.1. 529) variant of concern and its global perspective. J Med Virol. 2022;94(4):1738–44.
10. Dejnirattisai W, Huo J, Zhou D, Zahradník J, Supasa P, Liu C, Duyvesteyn HM, Ginn HM, Mentzer AJ, Tuekprakhon A. SARS-CoV-2 Omicron-B. 1.1. 529 leads to widespread escape from neutralizing antibody responses. Cell. 2022;185(3):467–84. e415.
11. Chen J, Wang R, Gilby NB, Wei G-W. Omicron variant (B. 1.1. 529): infectivity, vaccine breakthrough, and antibody resistance. J Chem Inf Model. 2022;62(2):412–22.
12. Allen H, Tessier E, Turner C, Anderson C, Blomquist P, Simons D, Løchen A, Jarvis CI, Groves N, Capelastegui F. Comparative transmission of SARS-CoV-2 Omicron (B. 1.1. 529) and Delta (B. 1.617. 2) variants and the impact of vaccination: national cohort study, England. Epidemiol Infect. 2023;151:e58.
13. Wolter N, Jassat W, Walaza S, Welch R, Moultrie H, Groome M, Amoako DG, Everatt J, Bhiman JN, Scheepers C. Early assessment of the clinical severity of the SARS-CoV-2 Omicron variant in South Africa. MedRxiv. 2021;2021(2012):2021–21268116.
14. Sigal A, Milo R, Jassat W. Estimating disease severity of Omicron and Delta SARS-CoV-2 infections. Nat Rev Immunol. 2022;22(5):267–9.

15. Gao SJ, Guo H, Luo G. Omicron variant (B. 1.1. 529) of SARS-CoV-2, a global urgent public health alert! J Med Virol. 2022;94(4):1255.
16. Yang W, Shaman JL. COVID-19 pandemic dynamics in South Africa and epidemiological characteristics of three variants of concern (Beta, Delta, and Omicron). eLife. 2022;11:e78933.
17. Moustafa AM, Bianco C, Denu L, Ahmed A, Coffin SE, Neide B, Everett J, Reddy S, Rabut E, Deseignora J. Comparative analysis of emerging B. 1.1. 7 + E484K SARS-CoV-2 isolates. Open Forum Infectious diseases: 2021. Oxford University Press US; 2021. p. ofab300.
18. [https://safegraph.com].
19. Pei S, Yamana TK, Kandula S, Galanti M, Shaman J. Burden and characteristics of COVID-19 in the United States during 2020. Nature. 2021;598(7880):338–41.
20. Pei S, Kandula S, Yang W, Shaman J. Forecasting the spatial transmission of influenza in the United States. *Proceedings of the National Academy of Sciences* 2018, 115(11):2752–2757.
21. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science. 2020;368(6490):489–93.
22. Pei S, Shaman J. Initial simulation of SARS-CoV2 spread and intervention effects in the continental US. *MedRxiv* 2020:2020-03.
23. COVID-19Projection. [https://github.com/shaman-lab/COVID-19Projection]
24. Anderson JL. An ensemble adjustment Kalman filter for data assimilation. Mon Weather Rev. 2001;129(12):2884–903.
25. Yang W, Lipsitch M, Shaman J. Inference of seasonal and pandemic influenza transmission dynamics. Proc Natl Acad Sci. 2015;112(9):2723–8.
26. Reis J, Shaman J. Retrospective parameter estimation and forecast of respiratory syncytial virus in the United States. PLoS Comput Biol. 2016;12(10):e1005133.
27. Jansen L, Tegomoh B, Lange K, Showalter K, Figliomeni J, Abdalhamid B, Iwen PC, Fauver J, Buss B, Donahue M. Investigation of a SARS-CoV-2 B. 1.1. 529 (omicron) variant cluster—Nebraska, November–December 2021. Morb Mortal Wkly Rep. 2021;70(51–52):1782.
28. Tanaka H, Ogata T, Shibata T, Nagai H, Takahashi Y, Kinoshita M, Matsubayashi K, Hattori S, Taniguchi C. Shorter incubation period among COVID-19 cases with the BA. 1 Omicron variant. Int J Environ Res Public Health. 2022;19(10):6330.
29. Backer JA, Eggink D, Andeweg SP, Veldhuijzen IK, van Maarseveen N, Vermaas K, Vlaemynck B, Schepers R, van den Hof S, Reusken CB. Shorter serial intervals in SARS-CoV-2 cases with Omicron BA. 1 variant compared with Delta variant, the Netherlands, 13 to 26 December 2021. Eurosurveillance. 2022;27(6):2200042.
30. Brandal LT, MacDonald E, Veneti L, Ravlo T, Lange H, Naseer U, Feruglio S, Bragstad K, Hungnes O, Ødeskaug LE. Outbreak caused by the SARS-CoV-2 Omicron variant in Norway, November to December 2021. Eurosurveillance. 2021;26(50):2101147.
31. Liu Y, Rocklöv J. The effective reproductive number of the Omicron variant of SARS-CoV-2 is several times relative to Delta. J Travel Med. 2022;29(3):taac037.
32. De Nicola G, Schneble M, Kauermann G, Berger U. Regional now-and forecasting for data reported with delay: toward surveillance of COVID-19 infections. AStA Adv Stat Anal. 2022;106(3):407–26.
33. Harris JE. Timely epidemic monitoring in the presence of reporting delays: anticipating the COVID-19 surge in New York City, September 2020. BMC Public Health. 2022;22(1):871.
34. Kandula S, Hsu D, Shaman J. Subregional nowcasts of seasonal influenza using search trends. J Med Internet Res. 2017;19(11):e370.
35. Beesley LJ, Osthus D, Del Valle SY. Addressing delayed case reporting in infectious disease forecast modeling. PLoS Comput Biol. 2022;18(6):e1010115.
36. Greene SK, McGough SF, Culp GM, Graf LE, Lipsitch M, Menzies NA, Kahn R. Nowcasting for real-time COVID-19 tracking in New York City: an evaluation using reportable disease data from early in the pandemic. JMIR Public Health Surveillance. 2021;7(1):e25538.
37. Backer JA, Eggink D, Andeweg SP, Veldhuijzen IK, van Maarseveen N, Vermaas K, Vlaemynck B, Schepers R, van den Hof S, Reusken CBEM et al. Shorter serial intervals in SARS-CoV-2 cases with Omicron BA.1 variant compared to Delta variant in the Netherlands, 13–26 December 2021. *medRxiv* 2022:2022.2001.2018.22269217.

Publisher's note