Integrating genomic data into test-negative designs for estimating lineage-specific COVID-19 vaccine effectiveness

Kevin C. Ma PhD[1], Diya Surie MD[1], Natalie Dean PhD[2], Clinton R. Paden PhD[1], Natalie J. Thornburg PhD[1], Fatimah S. Dawood MD[1]

[1] National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention (CDC), Atlanta, Georgia, USA

[2] Rollins School of Public Health, Emory University, Atlanta, Georgia, USA

Word count: 2987

Correspondence: Kevin C. Ma (tra3@cdc.gov), Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Atlanta, Georgia, USA

Disclaimer: The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention (CDC).

Abstract

Background: SARS-CoV-2 lineage-specific vaccine effectiveness (VE) studies can inform decision-making on whether COVID-19 vaccine composition updates are needed to maintain effectiveness against severe disease as SARS-CoV-2 continues to evolve. Lineage assignment methods in VE test-negative design (TND) studies include sequence-based (whole-genome sequencing), proxy-based (e.g., S-gene target failure during polymerase chain reaction), and period-based (variant predominance time periods) approaches.

Methods: We first summarize benefits, challenges (including cost and timeliness), and methodologic considerations related to estimating lineage-specific COVID-19 VE against emerging variants using these different assignment approaches. We then use a deterministic model to illustrate the potential impact of lineage misclassification error on VE estimates in a TND using period-based lineage assignment across different variant emergence scenarios.

Results: Our model suggests period-based analyses may underestimate differences in VE between two lineages due to lineage misclassification error in certain variant circulation scenarios. This effect is most evident during prolonged variant co-circulation or in early time periods following new variant takeover. Using higher predominance thresholds can reduce VE estimate bias in period-based analyses but at the expense of sample size, reducing precision or outright precluding estimation under some scenarios.

Conclusions: TND studies using sequence-, proxy-, and period-based lineage assignment have contributed substantially towards understanding SARS-CoV-2 variant-mediated vaccine escape, but biases that can affect each study design vary. Our results identify analytic considerations for robust estimation and suggest principles that may translate to other pathogens that undergo continuous antigenic drift.

Introduction

The continued emergence of SARS-CoV-2 lineages with increased immune evasion from COVID-19 vaccines and therapeutics presents a public health challenge. The earliest signals of immune escape are often obtained through laboratory characterization of neutralizing antibody titers against emerging variants using sera from vaccinated individuals [1–6]. Timely epidemiologic estimates of COVID-19 vaccine effectiveness (VE) against SARS-CoV-2 variants provide important real-world validation of serologic signals to guide public health risk assessments and decisions about updating COVID-19 vaccine composition [7].

The test-negative design (TND) has been increasingly applied for estimating overall and lineage-specific COVID-19 VE, but methods vary for integrating viral genetic data and addressing confounding and other biases [8–13]. A key consideration is the method used to ascertain and classify SARS-CoV-2 lineages. Common methods for lineage assignment include 1) sequence-based methods using whole-genome sequencing (WGS) of SARS-CoV-2 specimens; 2) proxy-based methods relying on a genetic marker associated with a lineage of interest, such as the spike protein residue 69–70 deletion resulting in S-gene polymerase chain reaction (PCR) amplification failure [14,15]; and 3) time period-based assignment using variant predominance periods. Importantly, these platforms have varied strengths, costs, timeliness, and susceptibility to biases that may affect the accuracy and precision of lineage-specific VE estimates.

Here, we highlight benefits, challenges, and methodologic considerations for different lineage assignment approaches when estimating COVID-19 VE against emerging and co-circulating SARS-CoV-2 variants. We also use a simple deterministic model to illustrate the potential impact of lineage misclassification error on VE estimates based on period-based assignment across different variant emergence scenarios.

Lineage assignment approaches in TND VE studies

Sequence-based

Sequence-based approaches rely on whole-genome sequencing (WGS) of SARS-CoV-2 cases for assigning lineage, which remains the gold standard for variant identification. Sequence-based methods were used throughout the pandemic to assess VE as new variants emerged [16–22]. For example, Kirsebom et al. assessed VE of XBB.1.5 and BA.4/5 boosters given as

part of the fall 2023 program in the UK against XBB, EG.5.1, and JN.1 lineages using WGS. The authors found evidence for lower VE against JN.1 lineages, which emerged in late 2023 and became predominant in January 2024 [19], mirroring findings from a multi-site cohort in the US which also used WGS to assess VE against XBB and JN lineages [13]. Unlike proxy- or period-based methods, sequence-based approaches have the advantage of full genomic resolution and could potentially also be used to assess VE by spike protein substitutions or by genetic distance from the COVID-19 vaccine strain [23]. Methods to characterize how vaccine efficacy varies by infecting pathogen strain have been referred to as sieve analyses and have been conducted for HIV, malaria, and SARS-CoV-2 vaccines in randomized clinical trials [25–28], but are not yet routinely applied in observational vaccine studies.

Sequence-based lineage assignment using WGS requires active sample collection or salvaging of stored clinical samples (potentially including rapid antigen tests [29]) and efficient specimen shipping, processing, and sequencing to produce timely VE estimates. Completeness of sequencing results also depends on adequate respiratory specimen quality; specimens with high PCR cycle threshold values often contain insufficient viral material for sequencing. Sequence-based lineage assignment is not free from bias, as specimens that cannot be sequenced due to low viral load are excluded and viral load may correlate with variant or vaccination status [30–32]. Sensitivity analyses can be conducted comparing overall VE with VE among patients with sequenced specimens to assess the degree of bias [25]. Additionally, mixed infection with multiple SARS-CoV-2 strains can rarely occur, and lineage identification approaches need to distinguish between sequencing artifacts, true mixed infections, or novel variants [33]. Lastly, attaining adequate sample sizes for sufficient statistical power in VE studies is perhaps the greatest challenge for studies based on WGS compared to other assignment approaches due to cost, equipment, and personnel requirements. This may preclude precise estimates of VE stratifying simultaneously by lineage and other covariates including age, immune status, and time since vaccination.

Proxy-based

In the absence of WGS capacity, real-time reverse transcription PCR against select amplification targets can sometimes be used as a proxy for genetic lineage (i.e. proxy-based lineage assignment). For SARS-CoV-2, a commonly used proxy is the TaqPath COVID-19 multiplexed diagnostic test, which amplifies the *S*, *ORF1ab*, and *N* genes of the viral genome

[14]. Failure of *S* gene amplification can occur due to a deletion at spike protein residues 69 and 70. This mutation was first observed with the Alpha variant [34,35], and has since alternated in occurrence with the emergence of some subsequent variants, including Delta, Omicron, and various Omicron descendants [15,36]. Targeted RT-PCRs for specific spike mutations have also been previously used to detect known variants of concern [37,38], and further work is needed to develop new scalable and accurate proxies [39,40].

The use of S-gene target failure (SGTF) or presence (SGTP) as a lineage marker can enable rapid estimation of lineage-specific VE [9,41–43]. Tseng et al. used SGTP/F in a large integrated health network to estimate effectiveness of mRNA-1273 boosters against Delta and Omicron variant infection, respectively, in December 2021 [44]. Similarly, Link-Gelles et al. used SGTP/SGTF in a nationwide pharmacy testing network to conduct early estimates of VE against symptomatic XBB lineage infection in 2023 [8,9] and JN lineage infection in 2024 [43]. Use of SGTP/F to define lineages can be more cost-effective and higher-throughput than WGS because SGTP/F relies on existing diagnostic RT-PCR laboratory infrastructure, and results are typically available more quickly [15,36,45]. However, SGTP/F cannot discriminate between specific mutations or other sub-lineage variation that may affect immune escape.

One of the main challenges for proxy-based approaches is that sensitivity and positive predictive value for emerging variants of interest must be continuously assessed and validated using WGS. The sensitivity of SGTP/F for differentiating lineages has varied, and can be increased by including “partial” SGTF samples where amplification of the S-gene occurs but is reduced compared to other genomic targets [36,46]. The positive predictive value of SGTP/F can be influenced by prevalence, especially as a new variant is beginning to emerge. In December 2022 as the XBB.1.5 variant began to spread in the US, positive predictive value of SGTP was relatively low, with XBB or XBB.1.5 constituting approximately 60% of SGTP specimens and the remainder being other BA.2 sublineages [9,15]. Andeweg et al. restricted to time periods when SGTP/F had >85% positive predictive value for variants of interest to address this issue [41], but this could delay VE estimation for emerging variants. Further assessments are needed to understand how this type of misclassification could influence VE estimates and to develop approaches to adjust for misclassification, such as regression calibration [47].

Time period-based

Period-based VE analyses approximate lineage assignment by assuming that all case-patients enrolled during a specified time period are infected by the predominant circulating SARS-CoV-2 variant. The period is defined using a prevalence threshold for the variants of interest, and thresholds used have varied considerably across studies: 50% is common [10,17,48–51], but higher thresholds such as 60% [11,52–54], 70% [55], 75% [56], 80% [12,57–59], and 90% [60] have also been used. A key advantage of the period-based lineage assignment approach is that it enables estimation of lineage-specific VE using data from studies without individual WGS or PCR laboratory data, which may allow for larger sample sizes. For instance, Schrag et al. used a period-based approach in a large multisite electronic health record network to characterize mRNA VE against medically attended illness during Delta versus Omicron periods, with simultaneous stratifications by pregnancy status, number of doses, and days since dose receipt [51]. Period-based approaches cannot resolve variants during prolonged periods of co-circulation, such as during mid-2024 onwards, when multiple JN.1 descendants (e.g. KP.2, KP.3, LB.1, KP.3.1.1, XEC, etc.) emerged and co-circulated without establishing sustained predominance [24]. Period-based approaches may also face difficulty disentangling the effect of lineage on VE in the presence of time-varying confounders, such as waning VE from antibody level decreases following increasing time since dose receipt. Extensions of period-based analyses to incorporate variant proportion as a continuous covariate may expand utility to a broader range of co-circulation scenarios [61].

Variant prevalence data for defining predominance periods can be derived using sequencing data from clinical specimens collected from the study population [48,49] or from external data sources. External sources include global sequence repositories [62] and surveillance programs such as the US CDC's national genomic surveillance platform, which aggregated genomic data from CDC-contracted laboratories, the National Strain Surveillance program, and public sequence data repositories [63]. Sequencing data availability and capacity varies substantially globally [64,65], but if external data sources are available, period-based analyses likely offer cost-savings over sequence- or proxy-based approaches as it does not require collection, storage, and processing of SARS-CoV-2 clinical specimens. Robust period-based VE assessments therefore rely on sustained support for genomic surveillance programs. Decreasing counts of published sequences in some national surveillance platforms [24,66] present a challenge for timely estimation of VE. Additionally, due to geographic heterogeneity in variant circulation patterns, local variant prevalence at TND enrollment sites may not match

external data sources. For instance, in the US, the XBB.1.5 lineage first appeared in the Northeast in late 2022 and attained predominance several weeks later on the West Coast [63]. Defining site-specific predominance periods in multi-cohort analyses may be considered to better account for regional differences in variant circulation [48,51], provided that sufficient subregional sequencing data are available.

Other study designs

Some investigators have used a hierarchy of data sources to classify variants [18,22,49,57,67]. For instance, Adams et al. and Lauring et al. conducted analyses in a multi-site TND network using periods defined by WGS of case-patients and a 50% prevalence threshold. Cases with a sequenced discordant lineage in a given period were excluded, reducing misclassification error where possible [18,49]. Tartof et al. used a hierarchy of approaches beginning with WGS when available, SGTF, and finally period-based classification (80% or higher threshold), maximizing the number of included case-patients [57,67].

Additionally, some studies lack test-negative controls but can still provide evidence to evaluate serologic signals of immune escape from updated COVID-19 vaccines. In these case-only study designs, adjusted odds of receipt of updated COVID-19 vaccination are compared between two lineage groups of interest [68–70], which can indicate a relative degree of vaccine immune escape. Moustsen-Helms et al. used this approach for JN.1 versus XBB lineages to provide early evidence for increased immune escape by JN.1 lineages [69] that later TND assessments corroborated. Case-only study designs and traditional TNDs can also support early assessments of evolving disease severity as new variants emerge and circulate [13,16,69–71].

Potential impact of lineage misclassification error in period-based approaches

Methods

We conducted a modeling study to assess the potential impact of lineage misclassification errors using a period-based approach for estimating VE. We used empirical US COVID-19 incidence and genomic surveillance data to inform model structure and parameters by estimating COVID-19 cases attributable to variants circulating in the US during July 2021 to

September 2024 (Supplementary Methods). We found that most variant epidemic curves were symmetric with variable overlap between successive variants and a standard deviation of approximately 6 weeks (Supplementary Figure 1). To mirror these dynamics, we used a simple deterministic model for two co-circulating variants assuming normally-distributed epidemic curves, with the peak of variant 2 circulation occurring a range of 2 to 30 weeks after the peak of variant 1, allowing full control over the extent of co-circulation (Supplementary Methods). VE against COVID-19 hospitalization was assumed to be 50% against variant 1 and 20% against variant 2, representing approximate differences in VE against hospitalization observed empirically against recent vaccine strain-matched and divergent variants, respectively [13,19,53,59]. Vaccine coverage was assumed to increase linearly from the start of the season before plateauing at 20% [72]. Incidence of non-SARS-CoV-2 acute respiratory infection resulting in hospitalization was assumed to be constant. Using a TND framework, we estimated lineage-specific VE using logistic regression adjusting for calendar time (categorical week), comparing VE estimate bias and precision when assigning lineages with sequence-based and period-based approaches using different predominance thresholds.

Results

By varying the peak of variant 2 circulation relative to variant 1, we generated a range of variant co-circulation scenarios. Setting variant 2 circulation to peak two weeks after variant 1 yielded prolonged co-circulation (37/39 weeks of circulation where both variant 1 and 2 were >10% prevalence) (Figure 1A). With increasing delay in the start of variant 2 circulation, variant epidemic curves had reduced overlap and variant 1 predominance transitioned more rapidly to variant 2 predominance (Figure 1C–D, Supplementary Figure 2). These scenarios correspond to a range of variant dynamics observed since 2021 in the US, such as prolonged co-circulation of multiple XBB descendant lineages in summer and fall 2023 followed by rapid takeover of XBB by JN.1 by January 2024 [24].

We assessed how period-based lineage-specific VE analyses using a 50% predominance threshold would perform under different variant dynamics scenarios. With prolonged co-circulation, a period-based analysis using a 50% threshold exhibited notable VE estimate bias (>10 percentage points for both variants) and underestimated the difference in VE between lineages (Figure 2A, Supplementary Table 1). As the duration of co-circulation decreases (i.e.,

sharper transition between variant periods), cases with misclassified lineage represented an increasingly smaller share of data and the bias in VE estimates decreased for both variants.

Because lineage misclassification during variant transition periods appeared to drive estimate bias, one solution is to increase the predominance threshold and effectively exclude weeks in which a substantial fraction of cases is likely to be misclassified. During periods in which estimation was possible, increasing the threshold to 70% (Figure 2B) or 90% (Figure 2C) reduced point estimate bias. However, by excluding samples, standard errors increased and confidence intervals widened (Supplementary Table 1). Additionally, in some scenarios with high co-circulation, sample sizes were insufficient (number of COVID-19 cases for either variant <100) to estimate VE, highlighting a tradeoff between VE estimate precision and bias for period-based analyses (Figure 2B and 2C).

These findings suggested lineage misclassification error is mitigated in scenarios in which a new variant quickly overtakes an old one (variant 2 peak delayed relative to variant 1 peak in this simulation). Because such scenarios are often of epidemiologic and scientific interest as they suggest a fitness advantage for the new variant, rapid estimation of VE as variant 2 is emerging is critical for early assessments of risk and public health impact [73]. Therefore, we analyzed a scenario where variant takeover was rapid (Figure 1D) and VE was estimated weekly following variant 2 reaching >50% prevalence. Using a sequence-based approach, estimates were unbiased but imprecise, and increasingly gained precision as data accumulated (Figure 3A, Supplementary Table 2). In contrast, a 50% period-based analysis underestimated the difference in VE between lineages, though both precision and bias improved with accumulation of data (Figure 3B). Use of more stringent predominance thresholds reduced bias in the point estimate, but precision decreased and estimation was not possible using higher thresholds until prevalence of variant 2 reached >70% and >90%, respectively (Figure 3C, Figure 3D, Supplementary Table 2).

Discussion

Through a deterministic simulation, we highlight how period-based analyses may underestimate differences in VE between two lineages due to lineage misclassification error in certain scenarios. This issue was most pronounced during prolonged variant co-circulation or in early time periods following rapid variant takeover. Using higher predominance thresholds might reduce VE estimate bias but at the expense of sample size, reducing precision or outright

precluding estimation under some scenarios. Sequence-based analyses do not face this tradeoff but can introduce selection bias, which was not explicitly modeled here, as not all case-patients will have successful specimen sequencing results. Findings from this analysis are meant to illustrate important considerations for estimating lineage-specific VE rather than provide definitive estimates of the magnitude of bias. TND studies using sequence-, proxy-, and period-based lineage assignment have all contributed substantially towards our understanding of SARS-CoV-2 variant-mediated vaccine escape, but biases that can affect these different study designs still need to be elucidated.

This model is subject to several limitations. First, we did not model biases that affect sequence-based approaches, including delays in generating sequencing data and selection bias for specimens with successful sequencing. Second, we assumed *a priori* a sizeable difference in VE between the two lineages; if COVID-19 VE is similar against the two lineages, the observed impact of misclassification error will be lower [17]. Third, to focus on the effect of lineage misclassification error, we did not model other general sources of bias in VE TNDs like confounding or misclassification of vaccination status. Fourth, we used a simplified, deterministic model of variant evolution and did not model cross-reactive immunity between variants or account for stochastic effects that may cause variant extinction. Fifth, we assumed rapid uptake of COVID-19 vaccine at the start of the circulation period with minimal variability in vaccine uptake beyond the early circulation period. Finally, our model did not consider “natural” waning of immunologic protection due to declining antibody levels. Under real-world settings, observed decreases in VE may be attributable to multiple factors beyond lineage replacement, and disentangling the impact of specific factors, such as natural waning and variant replacement, may not be possible [8].

Lineage-specific VE studies can inform decision-making about whether updates to COVID-19 vaccine composition are needed to maintain protection against severe disease as SARS-CoV-2 continues to evolve [7]. Sustained SARS-CoV-2 WGS surveillance is critical to support lineage-specific VE studies, whether directly integrated into a TND through a sequence-based analysis or through epidemiologic surveillance sequencing to guide determination of predominance periods for period-based studies. Our results add to the growing number of analytic considerations for robust estimation of VE using TNDs [74], which also include accounting for prior infection [75,76], correlated vaccination behaviors [77], calendar time [78,79], and other potential confounders [74]. Findings from this analysis can support high-quality national and global COVID-19 VE assessment strategies and suggest principles that may translate to other

pathogens that undergo continuous antigenic drift, such as influenza viruses and rotavirus [80–82].

Funding. This research received no external funding.

Conflicts of interest. The authors have declared no competing interests.

References

1. Cromer D, Steain M, Reynaldi A, Schlub TE, Khan SR, Sasson SC, et al. Predicting vaccine effectiveness against severe COVID-19 over time and against variants: a meta-analysis. Nat Commun. 2023;14: 1–9.

2. Khoury DS, Cromer D, Reynaldi A, Schlub TE, Wheatley AK, Juno JA, et al. Neutralizing antibody levels are highly predictive of immune protection from symptomatic SARS-CoV-2 infection. Nat Med. 2021;27: 1205–1211.

3. Wang Q, Mellis IA, Wu M, Bowen A, Gherasim C, Valdez R, et al. KP.2-based monovalent mRNA vaccines robustly boost antibody responses to SARS-CoV-2. Lancet Infect Dis. 2025;25: e133–e134.

4. Jian F, Wang J, Yisimayi A, Song W, Xu Y, Chen X, et al. Evolving antibody response to SARS-CoV-2 antigenic shift from XBB to JN.1. Nature. 2024;637: 921–929.

5. Yang S, Yu Y, Xu Y, Jian F, Song W, Yisimayi A, et al. Fast evolution of SARS-CoV-2 BA.2.86 to JN.1 under heavy immune pressure. Lancet Infect Dis. 2024;24: e70–e72.

6. Cromer D, Reynaldi A, Mitchell A, Schlub TE, Juno JA, Wheatley AK, et al. Predicting COVID-19 booster immunogenicity against future SARS-CoV-2 variants and the benefits of vaccine updates. Nat Commun. 2024;15: 1–9.

7. Grant R, Sacks JA, Abraham P, Chunsuttiwat S, Cohen C, Figueroa JP, et al. When to update COVID-19 vaccine composition. Nat Med. 2023;29: 776–780.

8. Link-Gelles R, Britton A, Fleming-Dutra KE, CDC COVID-19 Vaccine Effectiveness Team. Building the U.S. COVID-19 vaccine effectiveness program: Past successes and future directions. Vaccine. 2023; 125492.

9. Link-Gelles R. Early Estimates of Bivalent mRNA Booster Dose Vaccine Effectiveness in Preventing Symptomatic SARS-CoV-2 Infection Attributable to Omicron BA.5– and XBB/XBB.1.5–Related Sublineages Among Immunocompetent Adults — Increasing Community Access to Testing Program, United States, December 2022–January 2023. MMWR Morb Mortal Wkly Rep. 2023;72. doi:10.15585/mmwr.mm7205e1

10. Britton A, Embi PJ, Levy ME, Gaglani M, DeSilva MB, Dixon BE, et al. Effectiveness of COVID-19 mRNA vaccines against COVID-19-associated hospitalizations among immunocompromised adults during SARS-CoV-2 Omicron predominance - VISION Network, 10 states, December 2021-August 2022. MMWR Morb Mortal Wkly Rep. 2022;71: 1335–1342.

11. Merdrignac L, Lanièce Delaunay C, Verdasca N, Vega-Pirisc L, O'Donnell J, Sève N, et al. Effectiveness of XBB.1.5 Vaccines against Symptomatic SARS-CoV-2 Infection in Older Adults during the JN.1 Lineage-Predominant Period, European VEBIS Primary Care Multicentre Study, 20 November 2023–1 March 2024. 2024. doi:10.2139/ssrn.4883286

12. Rose AM, Nicolay N, Sandonis Martín V, Mazagatos C, Petrović G, Niessen FA, et al. Vaccine effectiveness against COVID-19 hospitalisation in adults (≥ 20 years) during Alpha- and Delta-dominant circulation: I-MOVE-COVID-19 and VEBIS SARI VE networks, Europe, 2021. Euro Surveill. 2023;28. doi:10.2807/1560-7917.ES.2023.28.47.2300186

13. Ma KC, Surie D, Lauring AS, Martin ET, Leis AM, Papalambros L, et al. Effectiveness of updated 2023-2024 (Monovalent XBB.1.5) COVID-19 vaccination against SARS-CoV-2 omicron XBB and BA.2.86/JN.1 lineage hospitalization and a comparison of clinical severity-IVY network, 26 hospitals, October 18, 2023-March 9, 2024. Clin Infect Dis. 2024; ciae405.

14. McMillen T, Jani K, Robilotti EV, Kamboj M, Babady NE. The spike gene target failure (SGTF) genomic signature is highly accurate for the identification of Alpha and Omicron SARS-CoV-2 variants. Sci Rep. 2022;12: 18968.

15. Scobie HM. Spike Gene Target Amplification in a Diagnostic Assay as a Marker for Public Health Monitoring of Emerging SARS-CoV-2 Variants — United States, November 2021–January 2023. MMWR Morb Mortal Wkly Rep. 2023;72. doi:10.15585/mmwr.mm7205e2

16. Kirsebom FCM, Harman K, Lunt RJ, Andrews N, Groves N, Abdul Aziz N, et al. Vaccine effectiveness against hospitalisation estimated using a test-negative case-control study design, and comparative odds of hospital admission and severe outcomes with COVID-19 sub-lineages BQ.1, CH.1.1. and XBB.1.5 in England. Lancet Reg Health Eur. 2023;35: 100755.

17. Lind ML, Copin R, McCarthy S, Coppi A, Warner F, Ferguson D, et al. Use of whole-genome sequencing to estimate the contribution of immune evasion and waning immunity on decreasing COVID-19 vaccine effectiveness. J Infect Dis. 2023;227: 663–674.

18. Lauring AS, Tenforde MW, Chappell JD, Gaglani M, Ginde AA, McNeal T, et al. Clinical severity of, and effectiveness of mRNA vaccines against, covid-19 from omicron, delta, and alpha SARS-CoV-2 variants in the United States: prospective observational study. BMJ. 2022;376: e069761.

19. Kirsebom FCM, Stowe J, Lopez Bernal J, Allen A, Andrews N. Effectiveness of autumn 2023 COVID-19 vaccination and residual protection of prior doses against hospitalisation in England, estimated using a test-negative case-control study. J Infect. 2024;89: 106177.

20. Bruxvoort KJ, Sy LS, Qian L, Ackerson BK, Luo Y, Lee GS, et al. Effectiveness of mRNA-1273 against delta, mu, and other emerging variants of SARS-CoV-2: test negative case-control study. BMJ. 2021;375: e068848.

21. Chung H, He S, Nasreen S, Sundaram ME, Buchan SA, Wilson SE, et al. Effectiveness of BNT162b2 and mRNA-1273 covid-19 vaccines against symptomatic SARS-CoV-2 infection and severe covid-19 outcomes in Ontario, Canada: test negative design study. BMJ. 2021;374: n1943.

22. Skowronski DM, Febriani Y, Ouakki M, Setayeshgar S, El Adam S, Zou M, et al. Two-dose severe acute respiratory syndrome Coronavirus 2 vaccine effectiveness with mixed schedules and extended dosing intervals: Test-negative design studies from British Columbia and Quebec, Canada. Clin Infect Dis. 2022;75: 1980–1992.

23. Cao L, Lou J, Chan SY, Zheng H, Liu C, Zhao S, et al. Rapid evaluation of COVID-19 vaccine effectiveness against symptomatic infection with SARS-CoV-2 variants by analysis of genetic distance. Nat Med. 2022;28: 1715–1722.

24. Guidance on conducting vaccine effectiveness evaluations in the setting of new SARS-CoV-2 variants: Interim guidance, 22 July 2021. Addendum to Evaluation of COVID-19 vaccine effectiveness. World Health Organization; 23 Jul 2021 [cited 31 May 2024]. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-vaccine_effectiveness-variants-2021.1

25. Rolland M, Gilbert PB. Sieve analysis to understand how SARS-CoV-2 diversity can impact vaccine protection. PLoS Pathog. 2021;17: e1009406.

26. Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, et al. Genetic diversity and protective efficacy of the RTS,S/AS01 malaria vaccine. N Engl J Med. 2015;373: 2025–2037.

27. Magaret CA, Li L, deCamp AC, Rolland M, Juraska M, Williamson BD, et al. Quantifying how single dose Ad26.COV2.S vaccine efficacy depends on Spike sequence features. Nat Commun. 2024;15: 2175.

28. Martin GE, Taiaroa G, Taouk ML, Savic I, O'Keefe J, Quach R, et al. Maintaining genomic surveillance using whole-genome sequencing of SARS-CoV-2 from rapid antigen test devices. Lancet Infect Dis. 2022;22: 1417–1418.

29. Puhach O, Meyer B, Eckerle I. SARS-CoV-2 viral load and shedding kinetics. Nat Rev Microbiol. 2023;21: 147–161.

30. Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. SARS-CoV-2, SARS-CoV, and MERS-CoV viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. Lancet Microbe. 2021;2: e13–e22.

31. Janes H, Fisher LH, Kee JJ, Parameswaran L, Goepfert PA, Falsey AR, et al. Association between SARS-CoV-2 viral load and COVID-19 vaccination in 4 phase 3 trials. J Infect Dis. 2024;230: 1384–1389.

32. Pipek OA, Medgyes-Horváth A, Stéger J, Papp K, Visontai D, Koopmans M, et al. Systematic detection of co-infection and intra-host recombination in more than 2 million global SARS-CoV-2 samples. Nat Commun. 2024;15: 1–13.

33. Walker AS, Vihta K-D, Gethings O, Pritchard E, Jones J, House T, et al. Tracking the emergence of SARS-CoV-2 alpha variant in the United Kingdom. N Engl J Med. 2021;385: 2582–2585.

34. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science. 2021;372: eabg3055.

35. Clark C, Schrecker J, Hardison M, Taitel MS. Validation of reduced S-gene target performance and failure for rapid surveillance of SARS-CoV-2 variants. PLoS One. 2022;17: e0275150.

36. Nasir A, Aamir UB, Kanji A, Bukhari AR, Ansar Z, Ghanchi NK, et al. Tracking SARS-CoV-2 variants through pandemic waves using RT-PCR testing in low-resource settings. PLOS Glob Public Health. 2023;3: e0001896.

37. Chung H-Y, Jian M-J, Chang C-K, Lin J-C, Yeh K-M, Chen C-W, et al. Emergency SARS-CoV-2 variants of concern: Novel multiplex real-time RT-PCR assay for rapid detection and surveillance. Microbiol Spectr. 2022;10: e0251321.

38. Clark AE, Wang Z, Ostman E, Zheng H, Yao H, Cantarel B, et al. Multiplex fragment analysis for flexible detection of all SARS-CoV-2 variants of concern. Clin Chem. 2022;68: 1042–1052.

39. Jiang W, Ji W, Zhang Y, Xie Y, Chen S, Jin Y, et al. An update on detection technologies for SARS-CoV-2 variants of concern. Viruses. 2022;14: 2324.

40. Andeweg SP, de Gier B, Eggink D, van den Ende C, van Maarseveen N, Ali L, et al. Protection of COVID-19 vaccination and previous infection against Omicron BA.1, BA.2 and Delta SARS-CoV-2 infections. Nat Commun. 2022;13: 4738.

41. Accorsi EK, Britton A, Fleming-Dutra KE, Smith ZR, Shang N, Derado G, et al. Association Between 3 Doses of mRNA COVID-19 Vaccine and Symptomatic Infection Caused by the SARS-CoV-2 Omicron and Delta Variants. JAMA. 2022;327: 639–651.

42. Link-Gelles R, Ciesla AA, Mak J, Miller JD, Silk BJ, Lambrou AS, et al. Early estimates of updated 2023-2024 (monovalent XBB.1.5) COVID-19 vaccine effectiveness against symptomatic SARS-CoV-2 infection attributable to co-circulating omicron variants among immunocompetent adults - Increasing Community Access to testing program, United States, September 2023-January 2024. MMWR Morb Mortal Wkly Rep. 2024;73: 77–83.

43. Tseng HF, Ackerson BK, Luo Y, Sy LS, Talarico CA, Tian Y, et al. Effectiveness of mRNA-1273 against SARS-CoV-2 Omicron and Delta variants. Nat Med. 2022;28: 1063–1071.

44. Data quality report: national flu and COVID-19 surveillance report. In: Gov.uk [Internet]. [cited 16 Apr 2025]. Available: https://www.gov.uk/government/publications/sources-of-surveillance-data-for-influenza-covid-19-and-other-respiratory-viruses/data-quality-report-national-flu-and-covid-19-surveillance-report

45. Hilti D, Wehrli F, Berchtold S, Bigler S, Bodmer T, Seth-Smith HMB, et al. S-gene target failure as an effective tool for tracking the emergence of dominant SARS-CoV-2 variants in Switzerland and Liechtenstein, including Alpha, Delta, and Omicron BA.1, BA.2, and BA.4/BA.5. Microorganisms. 2024;12: 321.

46. Davies NG, Jarvis CI, CMMID COVID-19 Working Group, Edmunds WJ, Jewell NP, Diaz-Ordaz K, et al. Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. Nature. 2021;593: 270–274.

47. Link-Gelles R, Levy ME, Natarajan K, Reese SE, Naleway AL, Grannis SJ, et al. Estimation of COVID-19 mRNA Vaccine Effectiveness and COVID-19 Illness and Severity by Vaccination Status During Omicron BA.4 and BA.5 Sublineage Periods. JAMA Netw Open. 2023;6: e232598.

48. Adams K, Rhoads JP, Surie D, Gaglani M, Ginde AA, McNeal T, et al. Vaccine effectiveness of primary series and booster doses against covid-19 associated hospital admissions in the United States: living test negative design study. BMJ. 2022;379: e072065.

49. Plumb ID, Feldstein LR, Barkley E, Posner AB, Bregman HS, Hagen MB, et al. Effectiveness of COVID-19 mRNA Vaccination in Preventing COVID-19-Associated Hospitalization Among Adults with Previous SARS-CoV-2 Infection - United States, June 2021-February 2022. MMWR Morb Mortal Wkly Rep. 2022;71: 549–555.

50. Schrag SJ, Verani JR, Dixon BE, Page JM, Butterfield KA, Gaglani M, et al. Estimation of COVID-19 mRNA vaccine effectiveness against medically attended COVID-19 in pregnancy during periods of Delta and Omicron variant predominance in the United States. JAMA Netw Open. 2022;5: e2233273.

51. Laniece Delaunay C, Mazagatos C, Martínez-Baz I, Túri G, Goerlitz L, Domegan L, et al. COVID-19 Vaccine Effectiveness in autumn and winter 2022 to 2023 among older Europeans. JAMA Netw Open. 2024;7: e2419258.

52. Carazo S, Skowronski DM, Brousseau NM, Guay C-A, Sauvageau C, Racine É, et al. Monovalent mRNA XBB.1.5 vaccine effectiveness against COVID-19 hospitalization in Quebec, Canada: impact of variant replacement and waning protection during 10-month follow-up. medRxiv; 2024. Available: https://www.medrxiv.org/content/10.1101/2024.11.13.24317190v1.full.pdf

53. Antunes L, Rojas-Castro M, Lozano M, Martínez-Baz I, Leroux-Roels I, Borg M-L, et al. Effectiveness of the XBB.1.5 COVID-19 vaccines against SARS-CoV-2 hospitalisation among adults aged ≥ 65 years during the BA.2.86/JN.1 predominant period, VEBIS hospital study, Europe, November 2023 to May 2024. Influenza Other Respi Viruses. 2025;19: e70081.

54. Nguyen JL, Mitratza M, Volkman HR, de Munter L, Tran TMP, Marques C, et al. Effectiveness of the BNT162b2 XBB.1.5-adapted vaccine against COVID-19 hospitalization related to the JN.1 variant in Europe: a test-negative case-control study using the id.DRIVE platform. EClinicalMedicine. 2025;79: 102995.

55. Oliveira EA, Oliveira MCL, Colosimo EA, Simões E Silva AC, Mak RH, Vasconcelos MA, et al. Vaccine effectiveness against SARS-CoV-2 variants in adolescents from 15 to 90 days after second dose: A population-based test-negative case-control study. J Pediatr. 2023;253: 189-196.e2.

56. Tartof SY, Slezak JM, Puzniak L, Frankland TB, Ackerson BK, Jodar L, et al. Effectiveness of BNT162b2 XBB Vaccine against XBB and JN.1 Sub-lineages. medRxiv. 2024. p. 2024.05.04.24306875. doi:10.1101/2024.05.04.24306875

57. Rose AM, Nicolay N, Sandonis Martín V, Mazagatos C, Petrović G, Baruch J, et al. Vaccine effectiveness against COVID-19 hospitalisation in adults (≥ 20 years) during Omicron-dominant circulation: I-MOVE-COVID-19 and VEBIS SARI VE networks, Europe, 2021 to 2022. Euro Surveill. 2023;28: 2300187.

58. Caffrey AR, Appaneal HJ, Lopes VV, Puzniak L, Zasowski EJ, Jodar L, et al. Effectiveness of BNT162b2 XBB vaccine in the US Veterans Affairs Healthcare System. medRxiv. 2024. p. 2024.04.05.24305063. doi:10.1101/2024.04.05.24305063

59. Britton A, Fleming-Dutra KE, Shang N, Smith ZR, Dorji T, Derado G, et al. Association of COVID-19 vaccination with symptomatic SARS-CoV-2 infection by time since vaccination and Delta variant predominance. JAMA. 2022;327: 1032–1041.

60. Ma KC. Genomic surveillance for SARS-CoV-2 variants: Circulation of Omicron XBB and JN.1 lineages — United States, May 2023–September 2024. MMWR Morb Mortal Wkly Rep. 2024;73. doi:10.15585/mmwr.mm7342a1

61. Antunes L, Nunes B, Núñez O, Martínez-Baz I, Dockx Y, Borg M-L, et al. Modelling approaches for estimating vaccine effectiveness of consecutive SARS-CoV-2 variant sublineages in the absence of study-specific genetic sequencing data, VEBIS hospital network, Europe, 2023/24. medRxiv. 2025. p. 2025.06.24.25330183. doi:10.1101/2025.06.24.25330183

62. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health: Data, Disease and Diplomacy. Global Chall. 2017;1: 33–46.

63. Ma KC, Shirk P, Lambrou AS, Hassell N, Zheng X-Y, Payne AB, et al. Genomic Surveillance for SARS-CoV-2 Variants: Circulation of Omicron Lineages - United States, January 2022-May 2023. MMWR Morb Mortal Wkly Rep. 2023;72: 651–656.

64. Chen Z, Azman AS, Chen X, Zou J, Tian Y, Sun R, et al. Global landscape of SARS-CoV-2 genomic surveillance and data sharing. Nat Genet. 2022;54: 499–507.

65. Brito AF, Semenova E, Dudas G, Hassler GW, Kalinich CC, Kraemer MUG, et al. Global disparities in SARS-CoV-2 genomic surveillance. Nat Commun. 2022;13: 7003.

66. CDC. Published SARS-CoV-2 Sequences. In: Centers for Disease Control and Prevention [Internet]. 28 Mar 2020 [cited 16 Apr 2025]. Available: https://covid.cdc.gov/covid-data-tracker

67. Tartof SY, Slezak JM, Puzniak L, Hong V, Frankland TB, Ackerson BK, et al. BNT162b2 vaccine effectiveness against SARS-CoV-2 omicron BA.4 and BA.5. Lancet Infect Dis. 2022;22: 1663–1665.

68. Brown J, Mook P, Vanhaverbeke M, Gimma A, Hagan J, Singini I, et al. Case-only analysis of routine surveillance data: detection of increased vaccine breakthrough infections with SARS-CoV-2 variants in Europe. Epidemiol Infect. 2025;153: e16.

69. Moustsen-Helms IR, Bager P, Larsen TG, Møller FT, Vestergaard LS, Rasmussen M, et al. Relative vaccine protection, disease severity, and symptoms associated with the SARS-CoV-2 omicron subvariant BA.2.86 and descendant JN.1 in Denmark: a nationwide observational study. Lancet Infect Dis. 2024. doi:10.1016/S1473-3099(24)00220-2

70. Lewnard JA, Mahale P, Malden D, Hong V, Ackerson BK, Lewin BJ, et al. Immune escape and attenuated severity associated with the SARS-CoV-2 BA.2.86/JN.1 lineage. Nat Commun. 2024;15: 8550.

71. Quinot C, Kirsebom F, Andrews N, Stowe J, Ramsay M, Dabrera G, et al. Severity of COVID-19 sub-lineages XBB/XBB 1.5/XBB1.16, EG.5.1. and JN.1. in England. The Lancet Regional Health – Europe. 2024;43. doi:10.1016/j.lanepe.2024.100975

72. COVID-19 Vaccination Coverage and Vaccine Confidence Among Adults. 23 Jan 2024 [cited 25 Jul 2024]. Available: https://www.cdc.gov/vaccines/imz-managers/coverage/covidvaxview/interactive/adults.html

73. Lambrou AS, South E, Ballou ES, Paden CR, Fuller JA, Bart SM, et al. Early detection and surveillance of the SARS-CoV-2 variant BA.2.86 — worldwide, July–October 2023. Morbidity and Mortality Weekly Report. 2023;72: 1162–1167.

74. Mésidor M, Liu Y, Talbot D, Skowronski DM, De Serres G, Merckx J, et al. Test negative design for vaccine effectiveness estimation in the context of the COVID-19 pandemic: A systematic methodology review. Vaccine. 2024;42: 995–1003.

75. Kahn R, Feikin DR, Wiegand RE, Lipsitch M. Examining bias from differential depletion of susceptibles in vaccine effectiveness estimates in settings of waning. Am J Epidemiol. 2024;193: 232–234.

76. Wiegand R, Fireman B, Najdowski M, Tenforde M, Link-Gelles R, Ferdinands J. Bias and negative values of COVID-19 vaccine effectiveness estimates from a test-negative design without controlling for prior SARS-CoV-2 infection. 2024. Available: https://www.researchsquare.com/article/rs-4802667/latest

77. Doll MK, Pettigrew SM, Ma J, Verma A. Effects of Confounding Bias in Coronavirus Disease 2019 (COVID-19) and Influenza Vaccine Effectiveness Test-Negative Designs Due to Correlated Influenza and COVID-19 Vaccination Behaviors. Clin Infect Dis. 2022;75: e564–e571.

78. Dean NE, Halloran ME, Longini IM Jr. Temporal Confounding in the Test-Negative Design. Am J Epidemiol. 2020;189: 1402–1407.

79. Bond HS, Sullivan SG, Cowling BJ. Regression approaches in the test-negative study design for assessment of influenza vaccine effectiveness. Epidemiol Infect. 2016;144: 1601–1611.

80. Cates JE, Amin AB, Tate JE, Lopman B, Parashar U. Do Rotavirus Strains Affect Vaccine Effectiveness? A Systematic Review and Meta-analysis. Pediatr Infect Dis J. 2021;40: 1135–1143.

81. Maurel M, Pozo F, Pérez-Gimeno G, Buda S, Sève N, Oroszi B, et al. Influenza vaccine effectiveness in Europe: Results from the 2022-2023 VEBIS (Vaccine Effectiveness, Burden and Impact Studies) primary care multicentre study. Influenza Other Respi Viruses. 2024;18: e13243.

82. Flannery B, Kondor RJG, Chung JR, Gaglani M, Reis M, Zimmerman RK, et al. Spread of Antigenically Drifted Influenza A(H3N2) Viruses and Vaccine Effectiveness in the United States During the 2018-2019 Season. J Infect Dis. 2020;221: 8–15.

83. CDC. COVID Data Tracker Variant Proportions. [cited 16 May 2024]. Available: https://covid.cdc.gov/covid-data-tracker/#variant-proportions

Table 1. Methods, advantages, and considerations for estimating lineage-specific COVID-19 VE using different lineage assignment methods.

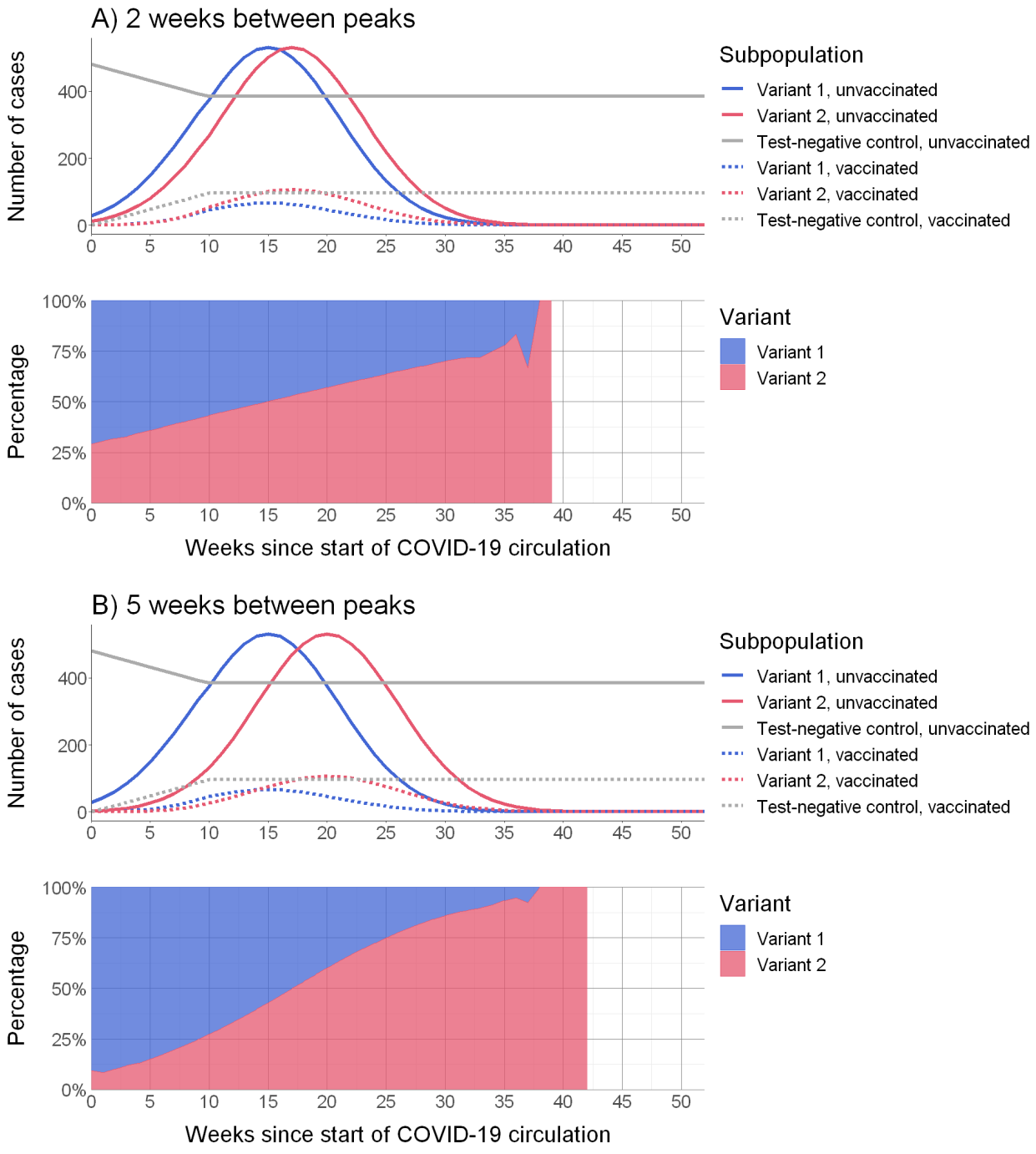
	Sequence-based	Proxy-based	Period-based
Method for lineage assignment	Whole-genome sequencing of SARS-CoV-2 positive cases	Use of a proxy genetic marker (such as spike Δ69–70 correlating with S-gene PCR target presence or failure) for a lineage of interest	Lineage assignment based on variant predominance time periods from internal or external data sources, such as national genomic surveillance or public sequence data repositories
Requirements and costs	Requires funding and capacity for whole-genome sequencing, including sample collection and storage; most expensive approach and sample sizes could be reduced by cost constraints	Less costly compared to whole-genome sequencing and can take advantage of existing clinical laboratory testing infrastructure	If external data sources are used, does not require direct funding or capacity for whole-genome sequencing of specimens, but still dependent on representativeness and timeliness of external sources
Advantages	Full genomic information allows identification of sublineages and mutations of interest even during periods of prolonged variant co-circulation	Compared to sequence-based analyses, less costly, higher throughput, and faster, and retains accuracy for categorizing broad variant groups compared to whole-genome sequencing data [15]. May have higher sample sizes compared to sequence-based analyses	Does not require collection or sequencing of specimens, enabling study platforms without laboratory capacity to also generate lineage-specific estimates. May have higher sample sizes compared to sequence-based analyses
Disadvantages and potential biases	Specimens that cannot be sequenced due to viral load may be dropped out; sensitivity analyses can be conducted	Cannot resolve specific spike protein substitutions of interest or sublineage evolution	Subject to lineage misclassification bias which could result in underestimation of the difference in VE by lineage

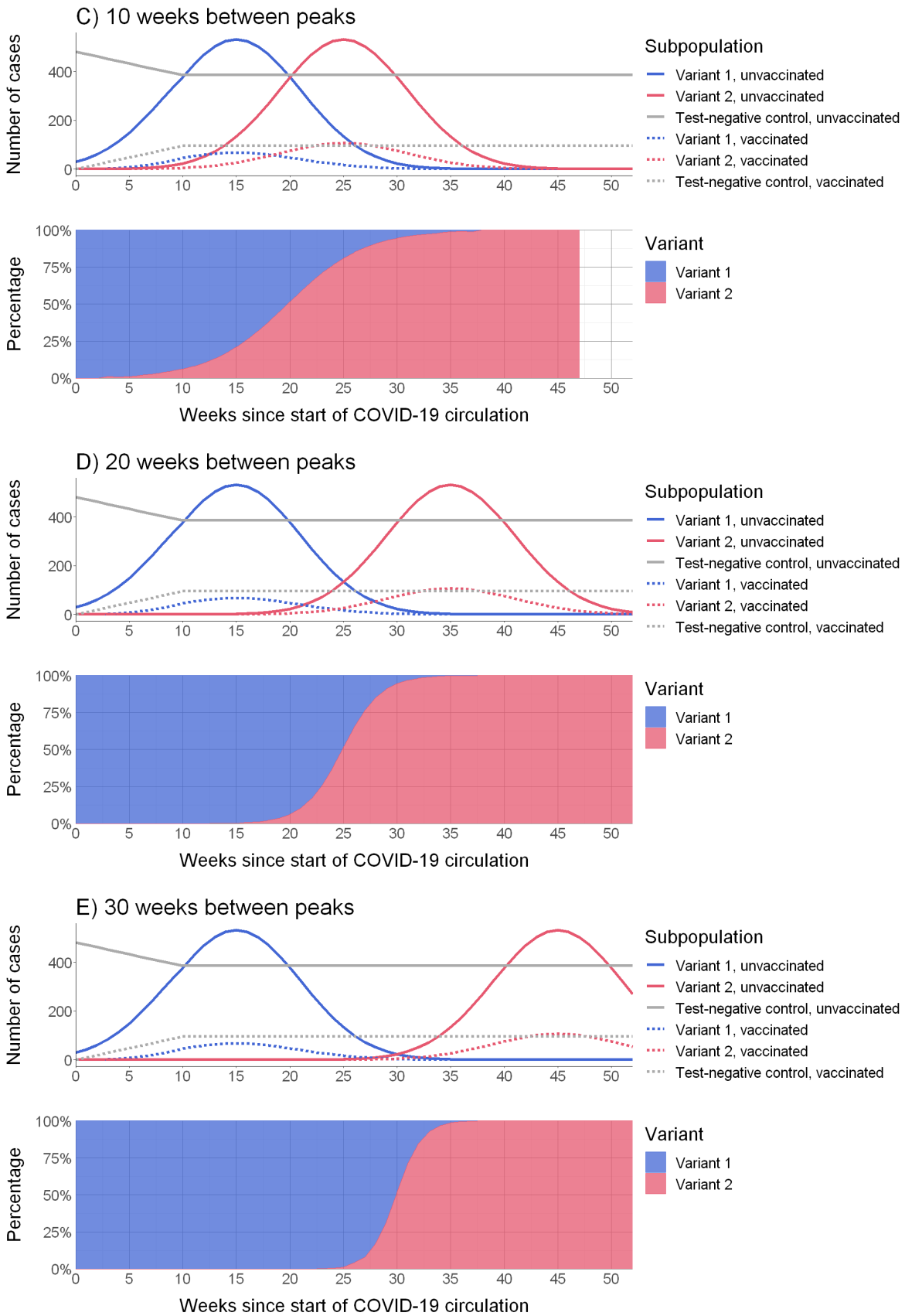
	comparing overall VE with VE among sequenced patients to assess degree of bias [25] Sequencing data lags need to be minimized for timely VE estimation to inform public health decision-making	Requires continued validation that proxies correlate well with lineage categories identified using whole-genome sequencing	Requires setting a threshold to define predominance, which has varied in the literature (50–90%) and could affect the extent of misclassification Cannot estimate lineage-specific VE during prolonged periods with multiple co-circulating variants Can be challenging to disentangle effect of variants from other time-varying confounders
Analytic opportunities	Leveraging the full extent of genomic information available to assess VE by spike protein substitutions or genetic distance	Characterization of the impact of proxy misclassification error and how model-based corrections can be used to improve estimates	Development of new analytic approaches to estimate VE during periods of prolonged co-circulation (e.g. using variant proportion as a covariate in regression models [61])

Table 2. Parameters used in lineage vaccine effectiveness simulations.

Parameter	Values or distribution	Notes and references
Population size (N)	1,000,000	
Duration	52 weeks	
Vaccination coverage	Linear increase from 0 to 20% by 10 weeks	Derived from empirical trends in vaccination coverage 2023–24 and 2024–25 in the US [72]
Lineage-specific VE against hospitalization (VE_1 and VE_2)	$VE_1 = 50\%$ $VE_2 = 20\%$	Derived from VE against XBB and JN.1 lineage hospitalization during administration of 2023–24 COVID-19 XBB.1.5 vaccines [13,19,53,59]
Variant epidemic time series	$P_1 \sim N(T_1, \sigma^2)$ $P_2 \sim N(T_2, \sigma^2)$	Model framework adapted from Bond et al. [79]
Incidence of non-SARS-CoV-2 acute respiratory illness pathogens	Uniform	Numbers of enrolled test-negative control-patients are assumed to accrue at a constant rate over time
Variant epidemic time series standard deviation ($\boldsymbol{\sigma}$)	6 weeks	Derived from empirical trends in variant circulation (Supplementary Methods). Corresponds to 95% of cases for a given variant occurring within a 24-week period
Variant 1 peak (T_1)	15 weeks	
Variant 2 peak (T_2)	17 to 45 weeks (2 to 30 weeks after variant 1 peak)	Range of values was selected to generate variety in co-circulation patterns
Percent of cases sequenced in sequence-based analysis	60%	Derived from empirical data from a multisite inpatient TND network [13]
Predominance thresholds α used in period-based analyses	50%, 70%, 90%	Range of values was selected to include thresholds used in previous analyses, such as 50% [10,17,48–51], 60% [11,52–54], 70% [55], 75% [56], 80% [12,57–59], and 90% [60]

Figure 1. Counts of simulated cases and controls by variant and vaccination status (top) and proportions of variants (bottom) over time for a range of co-circulation scenarios.

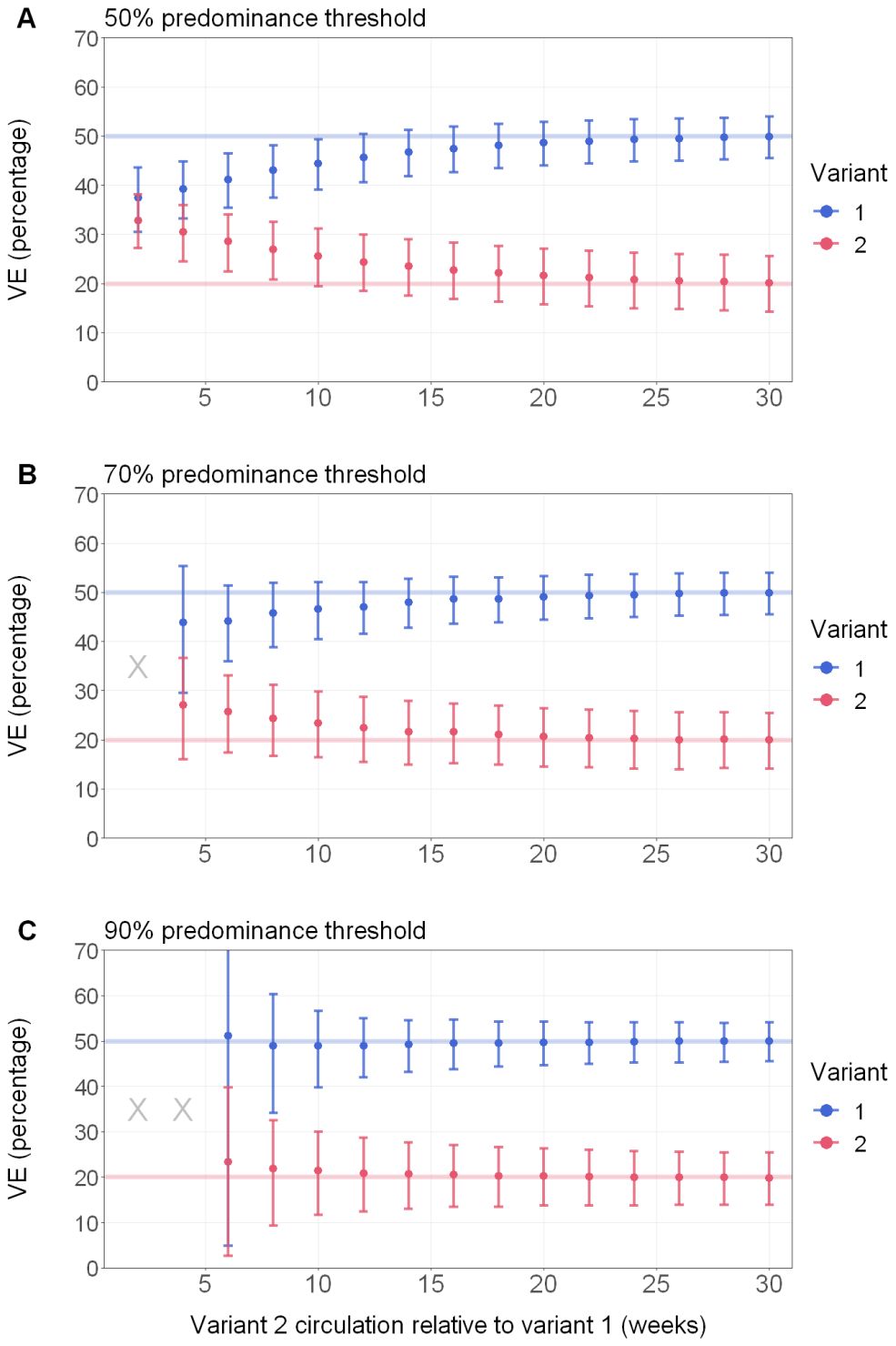



C) 10 weeks between peaks
Number of cases
Subpopulation
Variant 1, unvaccinated
Variant 2, unvaccinated
Test-negative control, unvaccinated
Variant 1, vaccinated
Variant 2, vaccinated
Test-negative control, vaccinated
Percentage
Variant
Variant 1
Variant 2
Weeks since start of COVID-19 circulation
D) 20 weeks between peaks
Number of cases
Subpopulation
Variant 1, unvaccinated
Variant 2, unvaccinated
Test-negative control, unvaccinated
Variant 1, vaccinated
Variant 2, vaccinated
Test-negative control, vaccinated
Percentage
Variant
Variant 1
Variant 2
Weeks since start of COVID-19 circulation
E) 30 weeks between peaks
Number of cases
Subpopulation
Variant 1, unvaccinated
Variant 2, unvaccinated
Test-negative control, unvaccinated
Variant 1, vaccinated
Variant 2, vaccinated
Test-negative control, vaccinated
Percentage
Variant
Variant 1
Variant 2
Weeks since start of COVID-19 circulation

Variant co-circulation was modeled using symmetric normally-distributed epidemic curves, with the peak of variant 2 circulation occurring a range of 2 to 30 weeks after variant 1. VE was assumed to be 50% against variant 1 and 20% against variant 2, representing approximate differences in VE observed empirically against recent strain-matched and divergent variants, respectively. Vaccine coverage was assumed to increase linearly from the start of the season before plateauing at 20%, derived from empirical trends in vaccination coverage 2023–24 and 2024–25 in the US.

Alt text: Multiple subfigures displaying line graphs of counts of simulated cases and controls by variant and vaccination status and area plots of proportions of variants over time for a range of co-circulation scenarios.

Figure 2. Point estimates and 95% confidence intervals for period-based vaccine effectiveness estimates across different variant co-circulation scenarios and predominance thresholds.

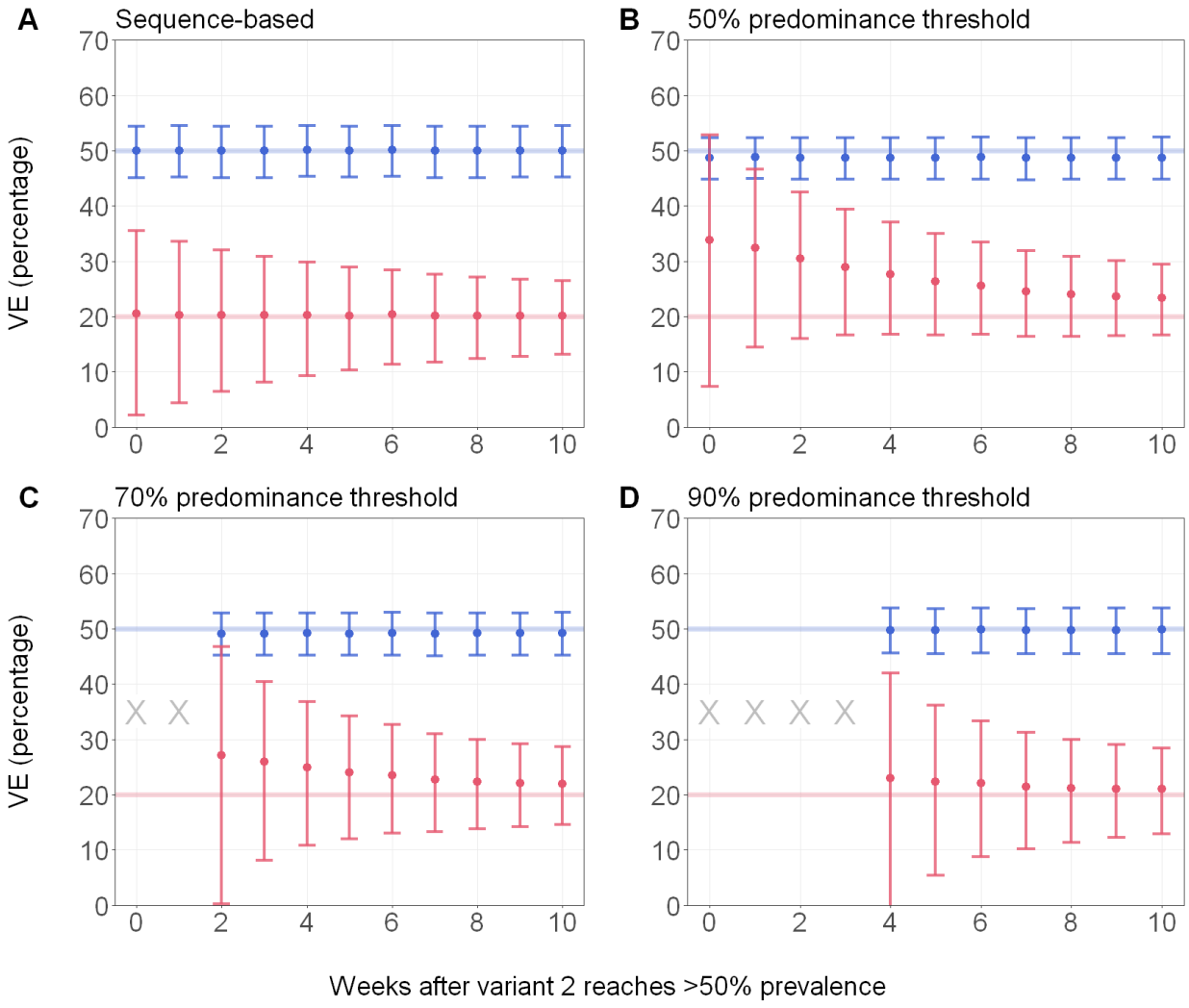


X = weeks during which the number of COVID-19 cases <100 for either variant and VE could not be estimated.

Variant co-circulation was modeled using symmetric normally-distributed epidemic curves, with the peak of variant 2 circulation occurring a range of 2 to 30 weeks after variant 1. VE was assumed to be 50% against variant 1 and 20% against variant 2, representing approximate differences in VE observed empirically against recent strain-matched and divergent variants, respectively. Vaccine coverage was assumed to increase linearly from the start of the season before plateauing at 20%, derived from empirical trends in vaccination coverage 2023–24 and 2024–25 in the US. Using a TND framework, we estimated lineage-specific VE using logistic regression adjusting for calendar time, and estimated lineage-specific VE using a period-based analysis with different predominance thresholds.

Alt text: Point estimates and 95% confidence intervals depicted using error bars for variant 1 and 2 (colors) for period-based vaccine effectiveness estimates. Estimates are shown across different variant co-circulation scenarios along the x-axis and predominance thresholds (50%, 70%, 90%) in separate subfigures.

Figure 3. Point estimates and 95% confidence intervals for (A) sequence-based and (B-D) period-based vaccine effectiveness estimates across different predominance thresholds by week after variant 2 reaches 50% prevalence.



X = weeks during which the number of COVID-19 cases <100 for either variant and VE could not be estimated.

Variant co-circulation was modeled using symmetric normally-distributed epidemic curves, with the peak of variant 2 circulation occurring a range of 2 to 30 weeks after variant 1. VE was assumed to be 50% against variant 1 and 20% against variant 2, representing approximate differences in VE observed empirically against recent strain-matched and divergent variants, respectively. Vaccine coverage was assumed to increase linearly from the start of the season before plateauing at 20%, derived from empirical trends in vaccination coverage 2023–24 and 2024–25 in the US. Using a TND framework, we estimated lineage-specific VE using logistic

regression adjusting for calendar time, and estimated lineage-specific VE using a period-based analysis with different predominance thresholds.

Alt text: Point estimates and 95% confidence intervals depicted using error bars for variant 1 and 2 (colors) for period-based vaccine effectiveness estimates. Estimates are shown by number of weeks after variant 2 reaches >50% prevalence along the x-axis and lineage assignment method (sequence-based, or period-based using 50%, 70%, 90% thresholds) in separate subfigures.

Supplementary Materials

Supplementary Methods

Variant epidemic curves

We used data from CDC's national genomic surveillance program, which aggregates biweekly SARS-CoV-2 sequence data from CDC-funded programs and public data repositories [83], to estimate COVID-19 cases attributable to variants circulating from July 2021 to September 2024. Using previously defined SARS-CoV-2 lineage groups [24,63], we included the following variants that attained 5% prevalence or greater anytime from July 20, 2021, to September 14, 2024: B.1.617.2, B.1.1.529, BA.1.1, BA.2, BA.2.12.1, BA.4, BA.5, BA.4.6, BF.7, BQ.1, BQ.1.1, XBB.1.5-like, XBB.1.16-like, EG.5-like, XBB, XBB.2.3, FL.1.5.1-like, HV.1, HK.3-like, JD.1.1, JN.1-like, KQ.1-like, KP.2-like, JN.1.16-like, KP.3, LB.1-like, and KP.3.1.1-like lineages. XBB and JN.1 lineages with identical spike residue 31 and receptor binding domain substitutions (residues 332–527) were grouped and denoted as "representative lineage-like." We estimated biweekly numbers of COVID-19 infections attributable to these variants by multiplying counts of positive SARS-CoV-2 positive test results from the National Respiratory and Enteric Virus Surveillance System (NREVSS) with biweekly variant proportions from genomic surveillance [24]. For each included variant, we restricted to biweeks in which it circulated at 0.1% share or greater, and normalized to the maximum cases per variant. No additional adjustments were made for overall decreasing detection of COVID-19 cases over time due to reductions in testing availability.

Biweekly estimates of normalized variant-attributable COVID-19 cases generally appeared to be symmetrically distributed over time (Supplementary Figure 1). Periods of sequential variant takeover and predominance, such as during early Omicron evolution from B.1.1.529/BA.1 to BA.2 to BA.4/5 in 2022, appeared to have staggered variant epidemic curves (Supplementary Figure 1A). Periods of more prolonged co-circulation, such as during the latter half of 2023 when multiple XBB descendants circulated, showed greater overlap in variant epidemic curves (Supplementary Figure 1B). The peak number of attributed cases per variant occurred a median of 12 weeks after the variant first reached 0.1% prevalence. Across the variants included, the median epidemic curve standard deviation was 6.0 weeks. Assuming normally distributed time series, this indicated that approximately 95% of cases for a given variant occurred within a 24-week (5.5 month) period.

Variant co-circulation model

Based on our empiric assessments of variant circulation, we used a simple deterministic model to represent dynamics of two SARS-CoV-2 lineages, adapted from Bond et al.'s approach of defining symmetric epidemic curves using the binomial distribution for influenza transmission [79]. This approach allows full parametric determination of the shape and height of an epidemic curve but does not assume a mechanistic model underlying transmission.

Formally, epidemic dynamics for variants 1 and 2 are assumed to each follow a normally distributed time series P_i, with T_1 and T_2 representing the weeks in which variants 1 and 2 peak, respectively. Variant 1 is assumed to peak at 15 weeks, and variant 2 is assumed to peak 2 to 30 weeks after variant 1 (Figure 1); this degree of right-shift is the main variable that is adjusted to generate different variant circulation patterns. The standard deviation σ from the peak of the epidemic time series is assumed to be 6 weeks for both variants, corresponding to 95% of cases for a given variant occurring within a 24-week period.

$$P_1 \sim N(T_1, \sigma^2) \qquad P_2 \sim N(T_2, \sigma^2)$$

We define *U(t)* and *V(t)* as the number of unvaccinated and vaccinated individuals, respectively, at time *t*. The total population size [*N* = *U(t)* + *V(t)*] is fixed with the assumption of no births, deaths, or migration in or out of the population. We assume the total population size (e.g., VE study recruitment catchment area) is 1 million individuals with no age or other demographic strata modeled. The number of vaccinated individuals is represented by a piecewise linear function that increases until a pre-determined threshold V_{max} at time T_v. We set V_{max} to be 20% of the total population and T_v to be 10 weeks, as updated 2023–2024 COVID-19 vaccination uptake in the US appeared to plateau at approximately 20% of the eligible adult population several months after updated doses were available [72].

$$V(t) = V_{max}, \quad t \geq T_v$$

$$V(t) = \frac{V_{max}}{T_v} t, \quad t < T_v$$

The numbers of unvaccinated and vaccinated individuals hospitalized following infection from variant 1 are given by $U_1(t)$ and $V_1(t)$, respectively:

$$U_1(t) = U(t)\, \phi_1\, P_1(t)$$

$$V_1(t) = V(t)\, \phi_1\, P_1(t)\, (1 - VE_1)$$

where ϕ_1 is a parameter that scales the height of the epidemic curve (representing multiple factors not individually modeled here, including infectiousness and severity of variant 1) and VE_1 represents VE specific to variant 1. We assume all hospitalized patients are recruited into the VE study. Assuming $\phi_1 = \phi_2$, the dynamics of variant 2 are similar:

$$U_2(t) = U(t)\,\phi_2\,P_2(t)$$

$$V_2(t) = V(t)\,\phi_2\,P_2(t)\,(1 - VE_2)$$

We assume VE against variant 1 (VE_1) = 50% and VE_2 = 20%, representing approximate levels of recent VE against hospitalization observed empirically against strain-matched and divergent variants, respectively [13,19,53,59]. Test-negative controls can present with infection from multiple acute respiratory illness-causing pathogens and are used to estimate vaccine uptake. Numbers of enrolled test-negative control-patients are assumed to accrue at a constant rate over time. Example time series are shown in Figure 1 for counts of patients with variant 1, variant 2, and test-negative controls stratified by vaccination status.

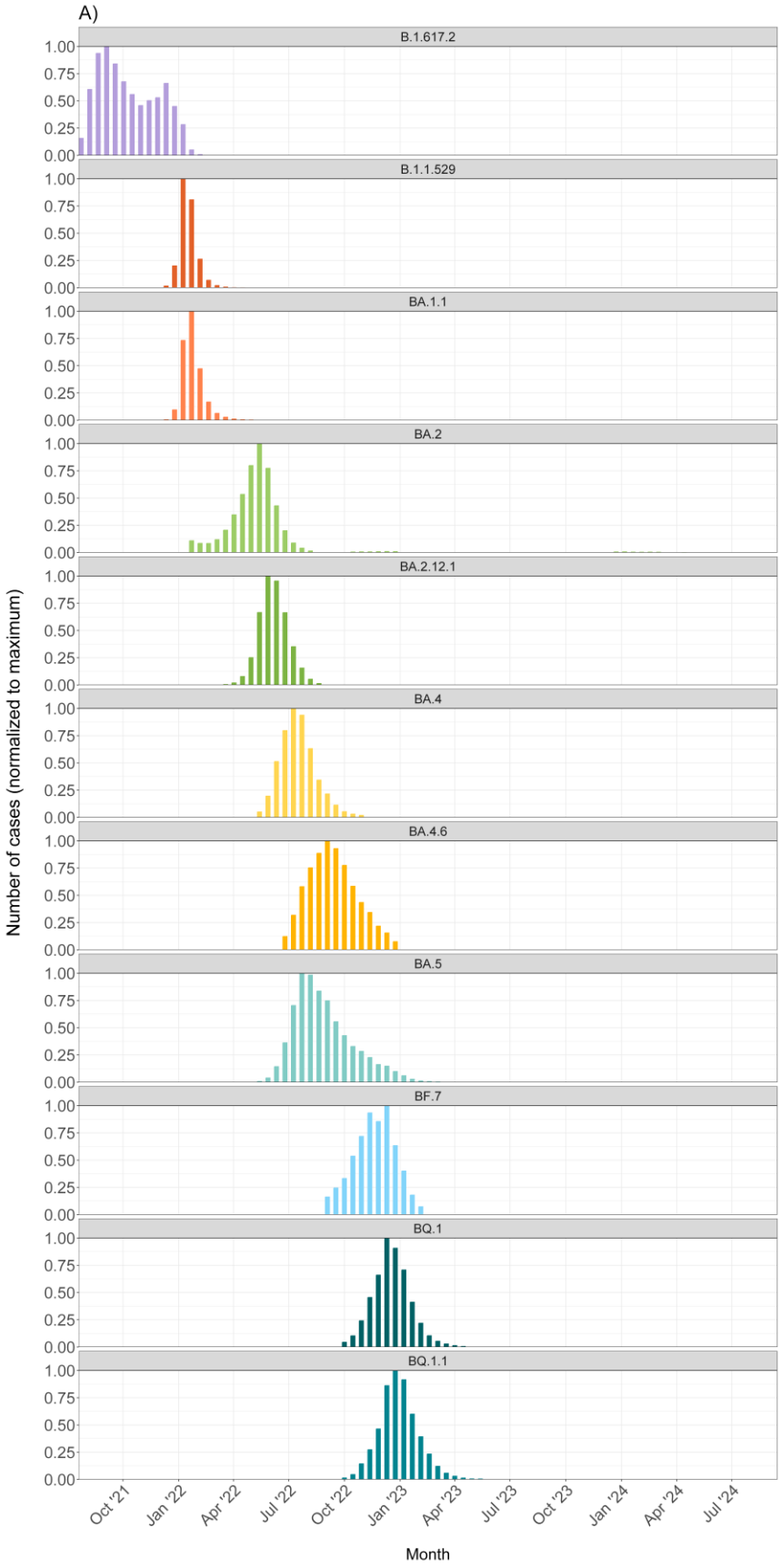
Vaccine effectiveness

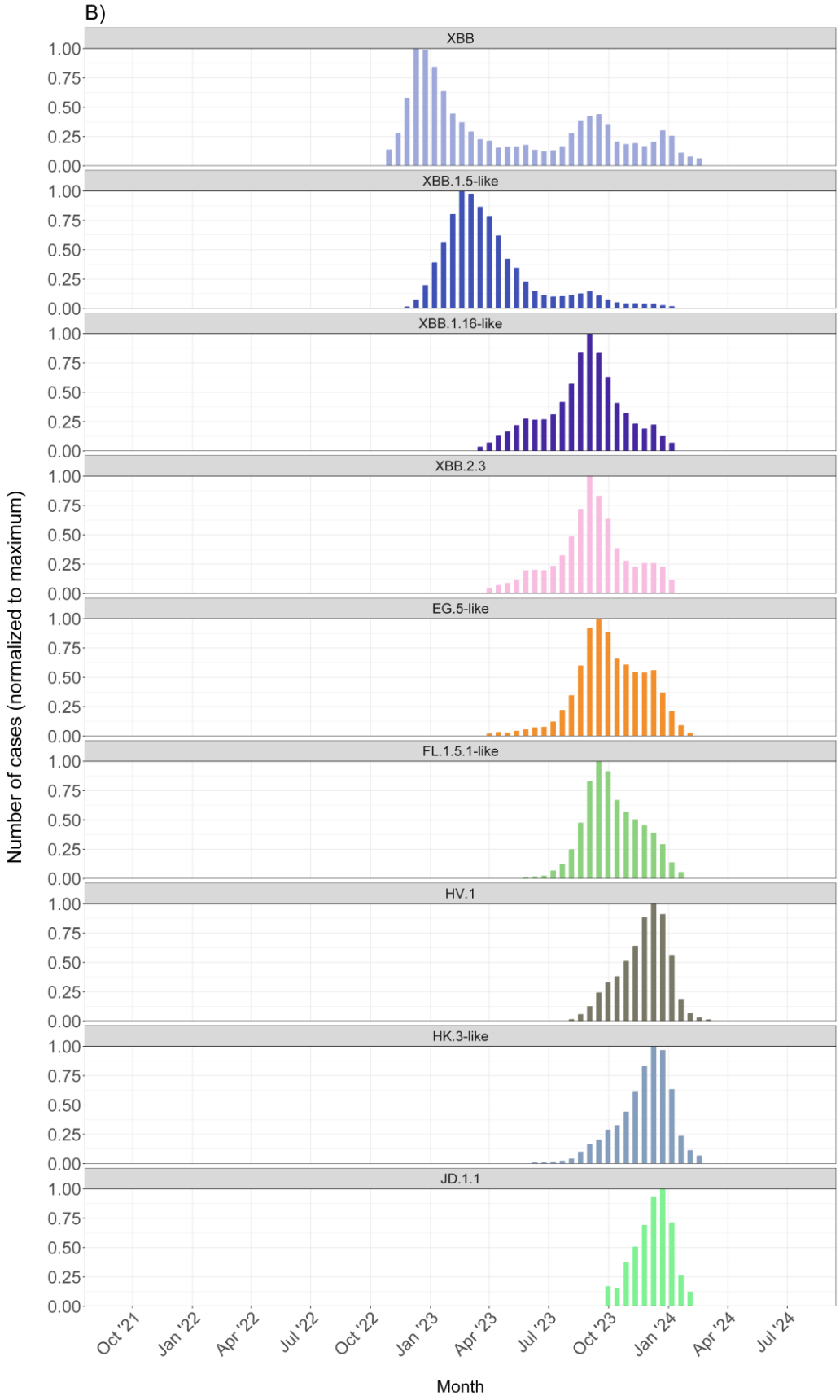
We simulate this model for 52 weeks, which produces a time series of vaccinated and unvaccinated cases for each SARS-CoV-2 lineage that is then used as input in a logistic regression model for estimating VE. The log odds of case versus control status was modeled as a linear function of vaccination status and categorical calendar week. VE was calculated as 1 minus the odds ratio. For period-based analyses with a specified threshold α, the predominance period for variant 1 was defined from the start of COVID-19 circulation (1 or more cases detected) to the last week in which variant 1's prevalence was above α. The predominance period for variant 2 spanned from the first week in which variant 2's prevalence was above α to the end of COVID-19 circulation. For sequence-based analyses, correct lineage determination was assumed to occur in 60% of patients, reflecting sample reduction from variable sample quality and other errors [13]. Our simplified model does not incorporate differences among patients with and without successful sequencing results.

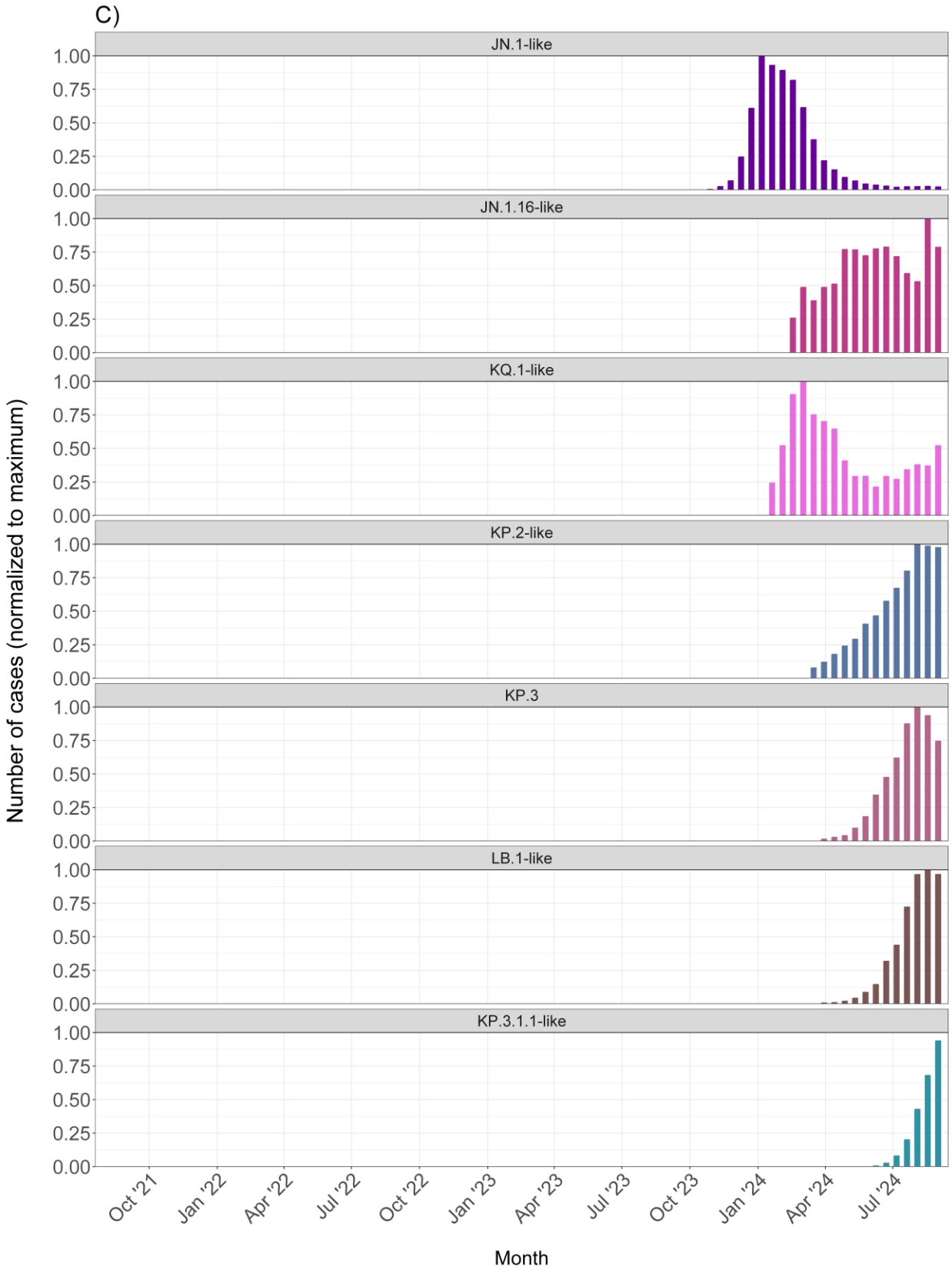
Supplementary Figure 1. Normalized estimated biweekly variant-attributed COVID-19 cases among A) early Omicron variants, B) XBB and descendants, and C) JN.1. and descendants

(Next page)

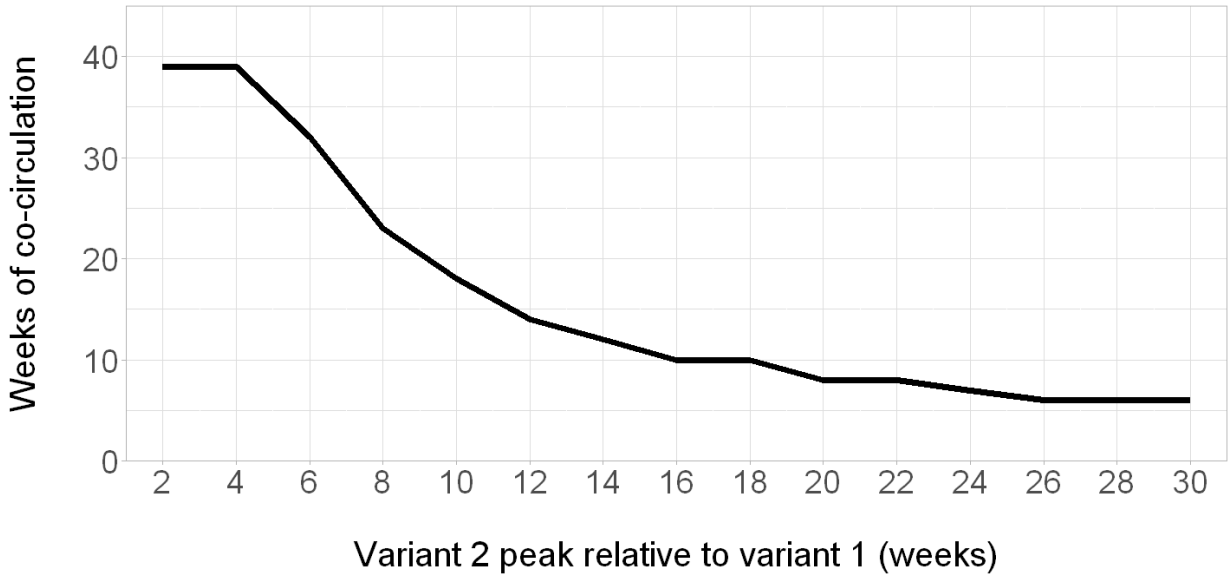
Data from CDC's national genomic surveillance program [83] were used to estimate COVID-19 cases attributable to variants circulating from July 2021 to September 2024 among previously defined SARS-CoV-2 lineage groups [24,63]. We estimated biweekly numbers of COVID-19 infections attributable to these variants by multiplying counts of positive SARS-CoV-2 positive test results from the National Respiratory and Enteric Virus Surveillance System (NREVSS) with biweekly variant proportions from genomic surveillance [24]. For each variant, we restricted to biweeks in which it circulated at 0.1% share or greater and normalized to the maximum cases per variant.


A)
B.1.617.2
B.1.1.529
BA.1.1
BA.2
BA.2.12.1
BA.4
BA.4.6
BA.5
BF.7
BQ.1
BQ.1.1
1.00
0.75
0.50
0.25
0.00
Number of cases (normalized to maximum)
Oct '21
Jan '22
Apr '22
Jul '22
Oct '22
Jan '23
Apr '23
Jul '23
Oct '23
Jan '24
Apr '24
Jul '24
Month


B)
XBB
XBB.1.5-like
XBB.1.16-like
XBB.2.3
EG.5-like
FL.1.5.1-like
HV.1
HK.3-like
JD.1.1
1.00
0.75
0.50
0.25
0.00
Number of cases (normalized to maximum)
Oct '21
Jan '22
Apr '22
Jul '22
Oct '22
Jan '23
Apr '23
Jul '23
Oct '23
Jan '24
Apr '24
Jul '24
Month


C)
JN.1-like
JN.1.16-like
KQ.1-like
KP.2-like
KP.3
LB.1-like
KP.3.1.1-like
1.00
0.75
0.50
0.25
0.00
Number of cases (normalized to maximum)
Oct '21
Jan '22
Apr '22
Jul '22
Oct '22
Jan '23
Apr '23
Jul '23
Oct '23
Jan '24
Apr '24
Jul '24
Month

Supplementary Figure 2. Delay between peak of variant 1 and 2 versus duration of co-circulation in simulated epidemic curves.



Co-circulation was defined as weeks in which both variants had a prevalence of 10% or greater. Variant co-circulation was modeled using symmetric normally-distributed epidemic curves, with the peak of variant 2 circulation occurring a range of 2 to 30 weeks after variant 1.

Supplementary Tables

Supplementary Table 1. Simulation results underlying Figure 2 (Point estimates and 95% confidence intervals for period-based weekly vaccine effectiveness estimates across different variant co-circulation scenarios and predominance thresholds); see attached

Supplementary Table 2. Simulation results underlying Figure 3 (Point estimates and 95% confidence intervals for sequence-based and period-based weekly vaccine effectiveness estimates across different predominance thresholds by week after variant 2 reaches 50% prevalence); see attached