



Published in final edited form as:

Clin Infect Dis. 2026 May 20; 82(5): e932–e941. doi:10.1093/cid/ciag059.

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

Kevin C. Ma, PhD¹, Diya Surie, MD¹, Natalie Dean, PhD², Clinton R. Paden, PhD¹, Natalie J. Thornburg, PhD¹, Fatimah S. Dawood, MD¹

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

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

Abstract

Background: SARS-CoV-2 lineage-specific COVID-19 vaccine effectiveness (VE) studies can inform decision-making on whether 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 time period-based (using variant predominance thresholds) approaches.

Methods: We first summarize benefits, challenges (including cost and timeliness), and methodologic considerations for estimating lineage-specific COVID-19 VE 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 versus sequence-based lineage assignment across different variant emergence scenarios.

Results: Our model suggests period-based analyses may sometimes underestimate differences in VE between two lineages due to lineage misclassification error. 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. Period-based VE analyses should therefore consider including sensitivity analyses to characterize robustness of VE estimates to different predominance thresholds.

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

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).

Conflicts of interest. NED reports research grants from CDC, NIH, and the Gates Foundation; consulting fees from Janssen; support for attending meetings and/or travel from CEPI; and participation on an advisory or data safety monitoring board for the International Vaccine Institute. All other authors declare no competing interests.

for robust estimation and suggest principles that may translate to other pathogens that undergo continuous antigenic drift.

Summary:

Whole-genome sequencing, predominance time periods, and genetic proxies are common lineage assignment approaches used to estimate variant-specific COVID-19 vaccine effectiveness. These approaches have varied strengths, costs, and biases that may affect the accuracy and precision of vaccine effectiveness estimates.

Keywords

SARS-CoV-2; vaccine effectiveness; COVID-19; variant; test-negative design

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 [1,2]. 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 [3].

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 [4–7]. 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 [8,9]; and 3) time period-based assignment using variant predominance periods. Importantly, these approaches 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 (see Table 1 for a summary). We 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. 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 [10], mirroring findings from a multi-site cohort in the US which also used WGS to assess VE against XBB and JN lineages [11]. 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 specific spike protein substitutions [12,13] or by genetic distance from the COVID-19 vaccine strain [14,15]. 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 [16], 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 [17]) 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 may introduce selection biases, as specimens that cannot be sequenced due to low viral load are excluded and viral load may correlate with variant, disease severity, or vaccination status [18,19]. Sensitivity analyses can be conducted comparing overall VE with VE among patients with sequenced specimens to assess the degree of bias [20]. 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 [21]. 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 for sequencing. 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 [8]. 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 [22], and has since alternated in occurrence with the emergence of some subsequent variants, including Delta, Omicron, and various Omicron descendants [9,23]. Targeted RT-PCRs for specific spike mutations have also been previously used to detect known variants of concern [24], and further work is needed to develop new scalable and accurate proxies [25].

The use of S-gene target failure (SGTF) or presence (SGTP) as a lineage marker can enable rapid estimation of lineage-specific VE. 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 [26]. 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 [27] and JN lineage infection in 2024 [28]. 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 [9,23]. However, SGTP/F cannot discriminate between specific mutations or other sub-lineage variation that may confer 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 [23,29]. 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 [9,27]. Andeweg et al. restricted to time periods when SGTP/F had >85% positive predictive value for variants of interest to address this issue [30], 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.

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% was common, but higher thresholds such as 60%, 70%, 80%, and 90% have also been used (see Supplementary Methods for examples). A key advantage of this assignment approach is that it enables estimation of lineage-specific VE from platforms without individual-level WGS or PCR laboratory data, potentially enabling larger sample sizes and representativeness. 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 [31]. 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 [32]. 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 vaccination. Extensions of period-based analyses to incorporate variant proportion as a continuous covariate may expand utility to a broader range of co-circulation scenarios [33].

Variant prevalence data for defining predominance periods can be derived using sequencing data from clinical specimens collected from the study population or from external data sources. External sources include global sequence repositories 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 [32]. Sequencing data availability and capacity varies substantially globally [34], 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, and decreasing counts of published sequences in some national surveillance platforms 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 [35]. Defining site-specific predominance periods in multicenter analyses may better account for regional differences in variant circulation [31], provided that sufficient regional sequencing data are available.

Other study designs: Some investigators have used a hierarchy of data sources to classify variants [36–39]. For instance, Adams et al. and Luring et al. conducted analyses in a multi-site TND network using periods defined by both 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 [37,38]. Tartof et al. used a hierarchy of approaches beginning with WGS when available, SGTF, and finally period-based classification (using an 80% or higher threshold), maximizing the number of included case-patients [39].

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, the adjusted odds of receipt of updated COVID-19 vaccination are compared between two lineage groups of interest, which can provide 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 that later TND assessments corroborated [40]. Both case-only study designs and traditional TNDs can also support early detections of changes in COVID-19 clinical severity as new variants emerge [11,40,41].

Potential impact of lineage misclassification errors

Methods: We conducted a modeling study to assess the potential impact of lineage misclassification errors using period-based versus sequence-based approaches 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 an average 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 parametric control over the extent of co-circulation (Supplementary Methods, see Table 2 for a summary of parameters) [42]. We assessed other strategies for generating co-circulation patterns, such as varying the standard deviation of the variant 2 epidemic curve while holding peak timing constant (Supplementary Figure 2), but these approaches did not produce prolonged (>30 weeks) co-circulation.

Time-invariant VE against COVID-19 hospitalization was assumed to be 50% against variant 1 and 20% against variant 2, representing approximate differences in VE observed empirically against recent vaccine strain-matched and divergent variants, respectively [10,11,43,44]. These estimates of COVID-19 VE represent additional protection beyond existing levels conferred by prior vaccination or infection, which has been increasingly difficult to ascertain. We explored the impact of time-varying VE using a piecewise function where VE increases from $\tau = 0$ to a peak at $\tau = 4$ weeks. The peak time-varying VE was defined to be 50% against variant 1 and 20% against variant 2 to correspond to the time-invariant setting. We then allowed VE to wane to 0% at an identical rate for the two variants to represent natural waning of immunologic protection over time (Supplementary Methods, Supplementary Figure 3). These parameters represent one illustrative scenario and are not meant to be best estimates of current rates of waning of COVID-19 VE.

Vaccine coverage was assumed to increase linearly from the start of the season before plateauing at 20% [45]. 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 across 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–F, Supplementary Figure 4). These scenarios correspond to the 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 [32].

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 decreased (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 was 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, predominance periods could not be defined or 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). When the modeled difference in VE between the two variants was reduced from 30% to 15% in a sensitivity analysis, the impact of this variant misclassification bias on absolute VE point estimate error was smaller, though relative trends were still similar (Supplementary Figure 5). Increasing the population-wide vaccination coverage level enhanced precision of estimates but did not substantially affect point estimates or the nature of bias observed (Supplementary Figure 6).

These findings suggested lineage misclassification error may be mitigated in scenarios in which a new variant quickly overtakes an old one. Such scenarios are often of epidemiologic and scientific interest as they suggest a fitness advantage for the new variant, but rapid estimation of VE as variant 2 is emerging is critical for early assessments of risk and public health impact. Therefore, we modified the scenario where variant takeover was rapid (Figure 1E) by estimating VE weekly following variant 2 reaching >50% prevalence, mirroring challenges of limited data availability early in emerging variant growth trajectories. Using a sequence-based approach, estimates were unbiased and initially imprecise, but increasingly gained precision as data accumulated (Figure 3A, Supplementary Table 2). In contrast, a 50% period-based analysis substantially underestimated the difference in VE between lineages early, though both precision and bias also 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 delayed until the prevalence of variant 2 reached the higher thresholds of 70% and 90%, respectively (Figure 3C, Figure 3D, Supplementary Table 2). Allowing VE to wane by time since vaccination yielded similar findings: VE estimation bias was highest for the 50% period-based analysis and lowest for the 90% period-based analysis during weeks when estimation was possible (Supplementary Figure 7).

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 at the expense of sample size, reducing precision or precluding estimation under some scenarios. Period-based VE analyses should therefore consider including sensitivity analyses to characterize robustness of VE estimates to different predominance thresholds. Sequence-based analyses do not face this tradeoff but

can introduce selection bias (not included in this model) as not all case-patients will have successful specimen sequencing results [19,20].

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 warrant evaluation (Table 1). Under real-world settings, observed decreases in VE over time may be attributable to multiple factors beyond lineage replacement, including natural waning of immunologic protection due to antibody decay and spurious measurement of waning due to depletion of susceptible individuals or other biases [46]. Disentangling the impact of specific factors in the TND can be challenging; one strategy is to jointly stratify COVID-19 VE estimates by variant and by time since vaccination categories [6]. Other analytic considerations in the booster era include increases in baseline levels of immunity, with high population levels of anti-SARS-CoV-2 seropositivity resulting from a combination of infection and prior vaccination. One simulation study suggests that failing to account for prior infection can exert a modest downward bias on COVID-19 VE estimates from a TND [47]. 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. Empirical comparisons of lineage assignment approaches, ideally within a single study platform, remain important to inform interpretation of VE estimates and inform methodological best practices [48].

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, to focus on the effect of lineage misclassification error, we used a basic infection model that did not account for risk heterogeneity, prior SARS-CoV-2 infection or COVID-19 vaccination, or depletion of susceptible individuals [46]. In addition, we did not model other general sources of bias in VE TNDs like confounding or misclassification of vaccination status, which have been reviewed and evaluated elsewhere [4–7,49,50]. Third, we limited assessment to two distinct lineages and did not model cross-reactive immunity or differential rates of waning VE between variants or account for stochastic effects that may cause variant extinction. Finally, based on empirical trends in the US, we assumed fairly rapid uptake of COVID-19 vaccine at the start of the circulation period with minimal increases in uptake after the initial weeks.

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 [3]. 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 highlight how period-based VE analyses can benefit from assessing VE using different predominance thresholds, and add to the growing number of analytic considerations for TNDs [5], including accounting for prior infection [46], correlated vaccination behaviors, calendar time [6,11], and other potential confounders and biases [5,49]. Findings from this analysis can support 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding.

This research received no external funding.

Data Availability Statement.

Simulation results are available in supplementary tables.

References

1. 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. [PubMed: 34002089]
2. 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. [PubMed: 38169466]
3. 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. [PubMed: 36807683]
4. Meah S, Shi X, Fritsche LG, Salvatore M, Wagner A, Martin ET, et al. Design and analysis heterogeneity in observational studies of COVID-19 booster effectiveness: A review and case study. *Sci Adv.* 2023;9: eadj3747.
5. 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. [PubMed: 38072756]
6. Varma A, Andrews NJ, Carazo S, Walter K, Kissling E, Skowronski DM, et al. Analytical approaches and examples of addressing time-varying factors in COVID-19 vaccine effectiveness studies: Report from a meeting of the World Health Organization. *Vaccine.* 2025;62: 127567.
7. Feikin DR, Higdon MM, Andrews N, Collie S, Deloria Knoll M, Kwong JC, et al. Assessing COVID-19 vaccine effectiveness against Omicron subvariants: Report from a meeting of the World Health Organization. *Vaccine.* 2023;41: 2329–2338. [PubMed: 36797097]
8. 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. [PubMed: 36347878]
9. 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
10. 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.
11. Ma KC, Surie D, Luring 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.
12. Ma KC, Webber A, Luring AS, Bendall E, Papalambros LK, Safdar B, et al. Effectiveness of 2024–2025 COVID-19 vaccination against COVID-19 hospitalization and severe in-hospital

- outcomes - IVY network, 26 hospitals, September 1, 2024-April 30, 2025. medRxiv. 2025. p. 2025.08. 29.25334612. doi:10.1101/2025.08.29.25334612
13. Feder KA, Patel A, Vepachedu VR, Dominguez C, Keller EN, Klein L, et al. Association of E484K spike protein mutation with severe acute respiratory syndrome Coronavirus 2 (SARS-CoV-2) infection in vaccinated persons: Maryland, January-May 2021. *Clin Infect Dis*. 2022;74: 2053–2056. [PubMed: 34473242]
 14. 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. [PubMed: 35710987]
 15. Kwon J, Li K, Warren JL, Pandya S, Hahn AM, Pitzer VE, et al. Utilizing virus genomic surveillance to predict vaccine effectiveness. *Infectious Diseases (except HIV/AIDS)*. medRxiv; 2025. Available: <https://www.medrxiv.org/content/10.1101/2025.06.20.25329795v1.full.pdf>
 16. Rolland M, Gilbert PB. Sieve analysis to understand how SARS-CoV-2 diversity can impact vaccine protection. *PLoS Pathog*. 2021;17: e1009406.
 17. 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.
 18. 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. [PubMed: 39225478]
 19. Van Goethem N, Serrien B, Vandromme M, Wyndham-Thomas C, Catteau L, Brondeel R, et al. Conceptual causal framework to assess the effect of SARS-CoV-2 variants on COVID-19 disease severity among hospitalized patients. *Arch Public Health*. 2021;79: 185. [PubMed: 34696806]
 20. 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 July 2021 [cited 31 May 2024]. Available: https://www.who.int/publications/i/item/WHO-2019-nCoV-vaccine_effectiveness-variants-2021.1
 21. 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. [PubMed: 38169466]
 22. 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. [PubMed: 34879193]
 23. 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.
 24. 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.
 25. 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. [PubMed: 35616102]
 26. 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. [PubMed: 35189624]
 27. 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
 28. 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. [PubMed: 38300853]

29. 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. [PubMed: 38399725]
30. 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. [PubMed: 35961956]
31. 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.
32. Ma KC, Castro J, Lambrou AS, Rose EB, Cook PW, Batra D, et al. 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: 938–945. [PubMed: 39446667]
33. 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
34. 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. [PubMed: 36385137]
35. 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. [PubMed: 37319011]
36. 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. [PubMed: 35438175]
37. 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.
38. 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.
39. 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.
40. 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
41. 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.
42. 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. [PubMed: 26732691]
43. 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>
44. Link-Gelles R, Rowley EAK, Irving SA, Klein NP, Grannis SJ, Ong TC, et al. Estimated 2023–2024 COVID-19 Vaccine Effectiveness in Adults. *JAMA Netw Open*. 2025;8: e2517402–e2517402.

45. COVID-19 Vaccination Coverage and Vaccine Confidence Among Adults. 23 Jan 2024 [cited 25 July 2024]. Available: <https://www.cdc.gov/vaccines/imz-managers/coverage/covidvaxview/interactive/adults.html>
46. 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. [PubMed: 37771045]
47. Wiegand RE, Fireman B, Najdowski M, Tenforde MW, Link-Gelles R, Ferdinands JM. Bias and negative values of COVID-19 vaccine effectiveness estimates from a test-negative design without controlling for prior SARS-CoV-2 infection. *Nat Commun*. 2024;15: 10062. [PubMed: 39567531]
48. 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. [PubMed: 36408616]
49. Shi X, Li KQ, Mukherjee B. Current Challenges With the Use of Test-Negative Designs for Modeling COVID-19 Vaccination and Outcomes. *Am J Epidemiol*. 2023;192: 328–333. [PubMed: 36446573]
50. Andrews LIB, Halloran ME, Neuzil KM, van der Laan L, Huang Y, Andriesen J, et al. Evaluating the test-negative design for COVID-19 vaccine effectiveness using randomized trial data: A secondary cross-protocol analysis of 5 randomized clinical trials. *JAMA Netw Open*. 2025;8: e2512763.

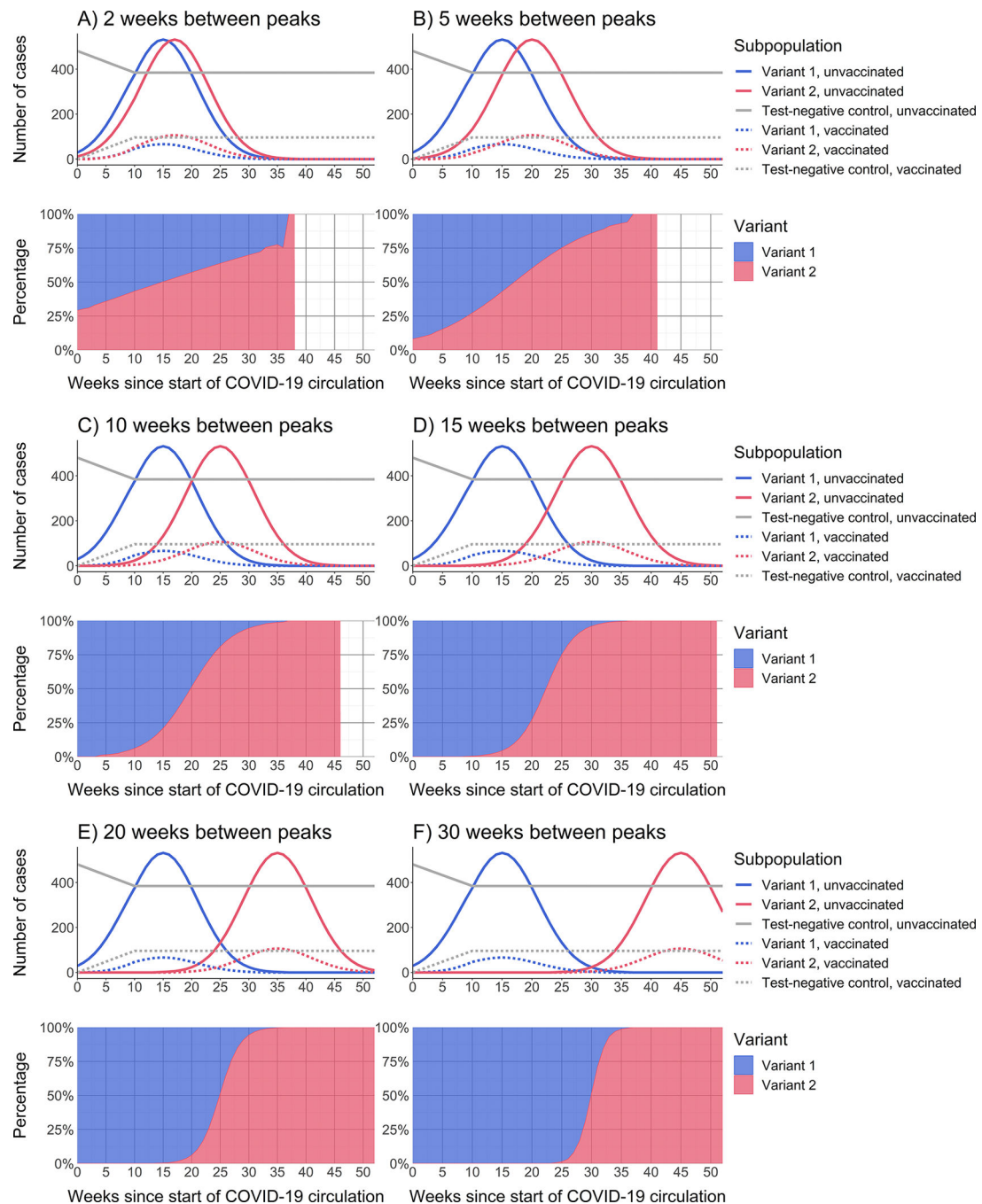


Figure 1. Counts of simulated cases and controls by variant and vaccination status (top inset) and proportions of variants (bottom inset) over time for a range of SARS-CoV-2 co-circulation scenarios (A-F).

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. Time-invariant 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.

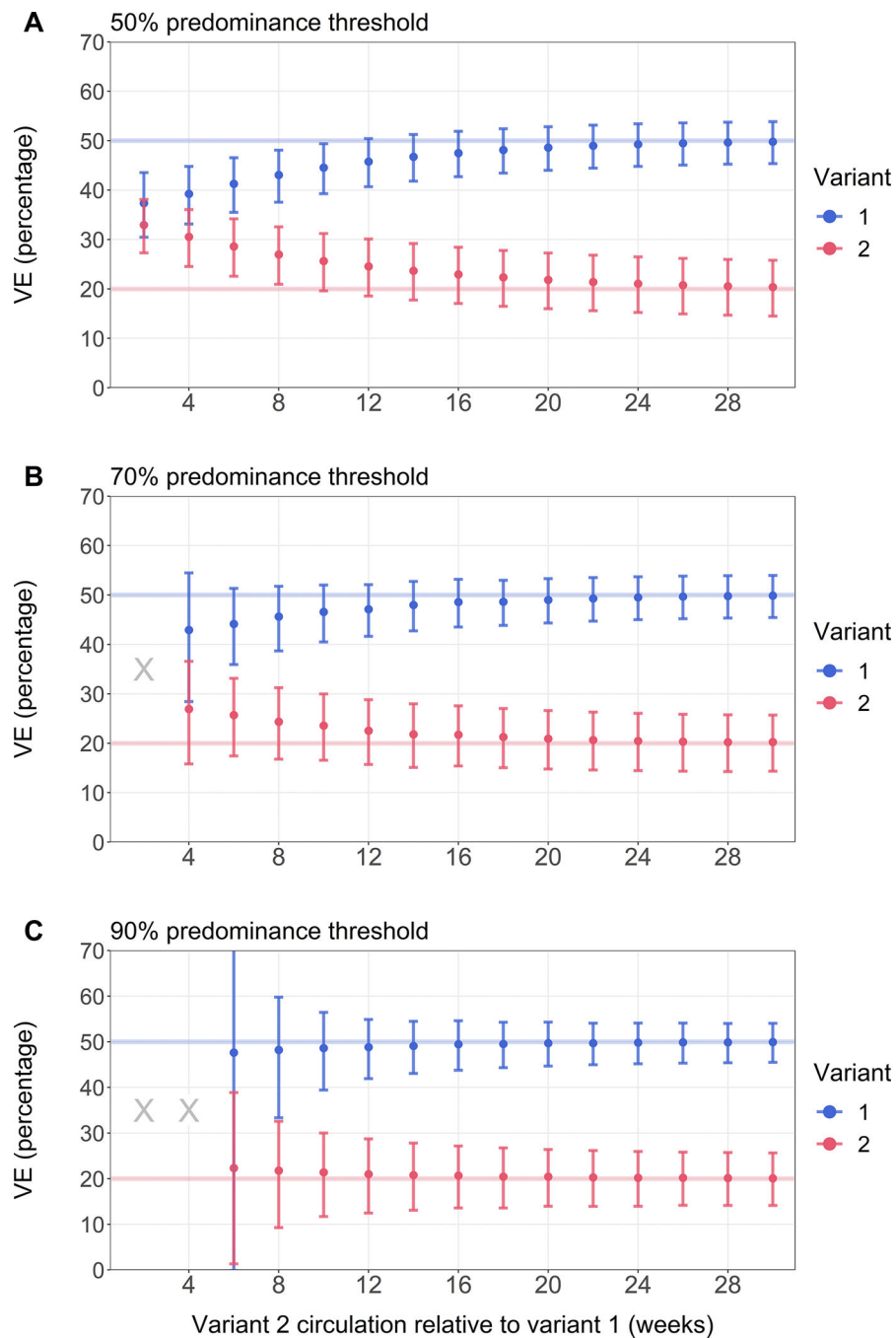


Figure 2. Point estimates and 95% confidence intervals for period-based COVID-19 vaccine effectiveness estimates across different variant co-circulation scenarios and predominance thresholds.

X = scenarios during which a predominance threshold was not met or the number of COVID-19 cases was <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. Time-invariant 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 we estimated lineage-specific VE using a period-based analysis with different predominance thresholds.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

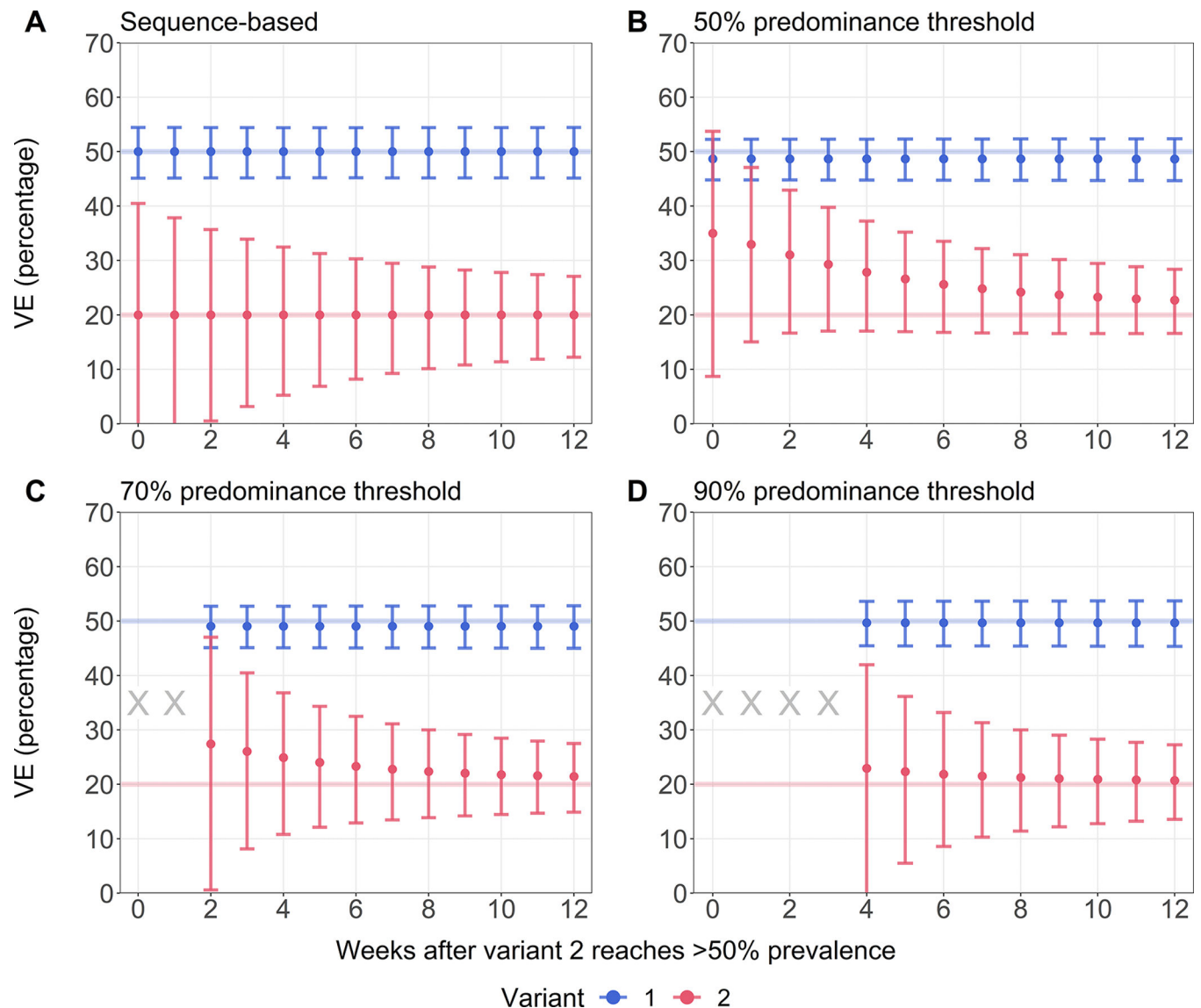


Figure 3. Point estimates and 95% confidence intervals for (A) sequence-based and (B-D) period-based COVID-19 vaccine effectiveness estimates across different predominance thresholds by week after variant 2 reaches 50% prevalence.

X = scenarios during which a predominance threshold was not met or the number of COVID-19 cases was <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. Time-invariant 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 we

estimated lineage-specific VE using a period-based analysis with different predominance thresholds.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Methods, advantages, and considerations for estimating lineage-specific COVID-19 vaccine effectiveness 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 $\Delta 69-70$ 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 [9]. 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 comparing overall VE with VE among sequenced patients to assess degree of bias [20] Sequencing data lags need to be minimized for timely VE estimation to inform public health decision-making	Cannot resolve specific spike protein substitutions of interest or sublineage evolution Requires continued validation that proxies correlate well with lineage categories identified using whole-genome sequencing	Subject to lineage misclassification bias which could result in underestimation of the difference in VE by lineage 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 [14,15]	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, including incorporating empirical variant proportions as a covariate in regression models [33] or imputing infecting variant probabilistically

Table 2.

Parameters used in lineage-specific COVID-19 vaccine effectiveness simulations.

Parameter	Value or distribution	Rationale 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 [45]
Time-invariant 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 [10,11,43]
Time-varying lineage-specific VE against hospitalization ($VE_1(\tau)$ and $VE_2(\tau)$)	VE starts at 0% and peaks 4 weeks after vaccination at 50% and 20% against variant 1 and 2, respectively; VE wanes after peaking (Supplementary Figure 3)	Waning vaccine effectiveness was defined using a piecewise function approach (Supplementary Methods)
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. [42]
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 (σ)	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 [11]
Predominance thresholds α used in period-based analyses	50%, 70%, 90%	Range of values was selected to include thresholds used in previous analyses, typically spanning from 50% to 90%