Article

Impact of unequal testing on vaccine effectiveness estimates across two study designs: a simulation study

Korryn Bodner [1] ✉, Linwei Wang[1], Rafal Kustra [2,3], Jeffrey C. Kwong [4,5,6,7,8], Beate Sander[4,6,9,10], Hind Sbihi [11,12], Michael A. Irvine[11,13] & Sharmistha Mishra [1,4,5,9,14,15] ✉

Observational studies are essential for measuring vaccine effectiveness. Recent research has raised concerns about how a relationship between testing and vaccination may affect estimates of vaccine effectiveness against symptomatic infection (symptomatic VE). Using an agent-based network model and SARS-CoV-2 as an example, we investigated how differences in the likelihood of testing by vaccination could influence estimates of symptomatic VE across two common study designs: retrospective cohort and test-negative designs. First, we measured the influence of unequal testing on symptomatic VE estimates across study designs and sampling periods. Next, we investigated whether the magnitude of bias in VE estimates from unequal testing was shaped by immune escape (vaccine efficacy against susceptibility and against infectiousness) and underlying epidemic potential (probability of transmission). We found that unequal testing led to larger bias in the cohort design than the test-negative design and that bias was largest with lower efficacy against susceptibility. We also found the magnitude of bias was moderated by the study's selected sampling period, efficacy against infectiousness, and probability of transmission, with these moderating effects more pronounced in the test-negative design. Our study illustrates that VE estimates across study designs require careful interpretation, especially in the presence of epidemic and immunological heterogeneity.

In real-world settings, observational studies are used to infer causal effects of how well vaccines protect against infection[1,2]. Accurate estimates of vaccine effectiveness (VE) are important, as they are used to evaluate the impact of vaccines at a population level and can affect vaccine confidence[3]. Most VE observational studies use population-level data collected through passive surveillance systems that rely on tests conducted to screen for or diagnose infections. However, testing for a given infection depends on many factors: variability in

[1]MAP Centre for Urban Health Solutions, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada. [2]Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada. [3]Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada. [4]ICES, Toronto, ON, Canada. [5]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada. [6]Public Health Ontario, Toronto, ON, Canada. [7]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada. [8]University Health Network, Toronto, ON, Canada. [9]Institute of Health Policy, Management and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada. [10]Toronto Health Economics and Technology Assessment (THETA) Collaborative, University Health Network, Toronto, ON, Canada. [11]British Columbia Centre for Disease Control, Vancouver, BC, Canada. [12]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada. [13]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada. [14]Institute of Medical Science, University of Toronto, Toronto, ON, Canada. [15]Division of Infectious Diseases, Department of Medicine, University of Toronto, Toronto, ON, Canada. ✉e-mail: kbodner@uoguelph.ca; sharmistha.mishra@utoronto.ca

recommendations across populations and risk factors for testing and treatment; access and uptake shaped in part by healthcare engagement; underlying burden of alternate etiologies of symptoms that drive testing. When some of these factors also influence vaccination, they can create unequal rates of testing by vaccination status, introducing bias into VE estimates such as residual confounding (due to unmeasured confounders), or selection bias (when restricting study samples to those tested)[4–8]. Concerns surrounding bias in VE estimates due to unequal testing by vaccination status were brought to the forefront during the SARS-CoV-2 pandemic, based on empirical data that revealed associations between vaccination status and testing[9,10]. Quantifying the direction and magnitude of bias in VE estimates due to unequal testing by vaccination, and the conditions under which this bias is most influential, could inform the interpretation of VE estimates and study design to reduce bias.

Two observational study designs commonly used to estimate VE are the retrospective cohort (e.g., refs. [11–15]) and retrospective test-negative designs (e.g., refs. [16–20]). Cohorts use vaccination status to assign exposure and estimate the relative risks of infection (between the vaccinated and unvaccinated cohorts) that may occur over a given sampling period. Classification of the outcome (infected vs. not infected) depends on testing for infection. Therefore, unmeasured confounders that lead to unequal testing by vaccination status can cause residual confounding in VE estimates from the cohort design[21–23].

Test-negative designs restrict the study sample to individuals who received a test for the infection in an attempt to eliminate unmeasured confounding[24,25]. At a given point in the epidemic (sampling period), cases (individuals who tested positive) and controls (individuals who tested negative) are compared to determine if the distribution of the exposure (vaccination) differs between cases and controls, using odds ratios[2]. However, as described by Westreich and Hudgens[6] and Sullivan, Tchetgen Tchetgen, and Cowling[7] through a directed acyclic graph (DAG), restricting the study sample to tested individuals may introduce a selection bias via selecting on a collider (also referred to as collider bias) in a situation where the exposure (vaccination) and outcome (infection) both independently affect the likelihood of being tested. Selecting for testing in this scenario distorts the association between vaccination and infection, thereby creating biased VE estimates[6,7].

Studies often focus on symptomatic infection as the outcome, estimating the vaccine's protection against symptomatic infection (hereafter referred to as symptomatic VE) (e.g., refs. [16, 26, 27]). When measuring symptomatic VE, these studies restrict their sample to only those individuals who exhibit symptoms and have been tested. Selecting for symptomatic cases may appear to mitigate the collider bias caused by selecting on testing by blocking the association between infection and testing. However, we posit that selecting on both testing and being symptomatic may induce artificial associations (e.g., between healthcare engagement and infection), allowing a bias to persist.

Various studies have qualitatively described the mechanisms by which unequal testing by vaccination could lead to residual confounding and/or selection bias[6–8]. However, fewer studies have measured the direction and magnitude of bias in VE estimates due to unequal testing by vaccination in cohort vs. test-negative designs (but see Suah et al.[28]); and, importantly, the epidemic and immunological properties under which bias in VE estimates are amplified or dampened in the presence of unequal testing.

Epidemic and immunological properties have the potential to shape biases in VE estimates over time and across epidemics. Epidemic properties such as the force of infection (incidence per susceptible) and the proportion of the population that is susceptible due to infection-acquired immunity can vary over the course of an epidemic and thus can influence the degree of bias in VE estimates over time. Epidemic and immunological properties that determine the overall size and length of an epidemic may also vary with changing contexts

and evolving pathogens, potentially shaping biases across epidemics. One epidemic condition that can vary is the underlying epidemic potential (probability of transmission), which could differ depending on contact rates and/or biological properties of the virus (e.g., increased transmissibility of SARS-CoV-2 variants[29,30]). Immunological properties such as the biological efficacy of vaccines against susceptibility (individual-level protection against acquiring infection[31]) and/or against infectiousness (individual-level reduction in transmitting virus if infected[31]) can also vary across epidemics, as seen with the increased immune escape of SARS-CoV-2 variants[32]. As such, epidemic and immunological properties that determine the overall size and length of an epidemic may moderate the magnitude of bias observed in a given epidemic.

In this study, we developed an agent-based network model, using SARS-CoV-2 as an example, to examine the magnitude of bias in symptomatic VE estimates due to unequal testing by vaccination status across retrospective cohort and test-negative designs. Here, our causal relationship of interest is the effect of vaccination on symptomatic SARS-CoV-2 infection (Fig. 1). First, using our simulated data, we compared the magnitude of this bias between the two study designs at the same sampling period and then examined how it varied across studies with different sampling periods within each study design (Objective 1). Next, we assessed the extent to which the magnitude of bias was moderated by immunological properties (vaccine efficacy against susceptibility and vaccine efficacy against infectiousness) (Objective 2) and by the underlying epidemic potential (probability of transmission) (Objective 3).

Results
Comparison of the magnitude of bias from testing differences across study designs and sampling periods

Using an agent-based network model and our default parameters (Fig. 2; Table 1), we simulated three testing scenarios that varied the magnitude of the relationship between testing and vaccination ("iii" in Fig. 1): an equal testing scenario (equal probability of testing by vaccination status); a moderately unequal testing scenario (vaccinated with 1.76 times higher testing); and a highly unequal testing scenario (vaccinated with 2.36 times higher testing). We assumed testing was limited to individuals with symptoms from SARS-CoV-2 infection or with COVID-like symptoms from alternate etiologies. We then measured the magnitude of bias in VE estimates—specifically in estimates derived from the cohort design (VE_{RR}) and the test-negative design (VE_{OR})—and compared them across study designs and across studies with different sampling periods. We ran the three testing scenarios using both a higher and lower efficacy against susceptibility (0.55 and 0.1), which were equivalent to higher and lower true symptomatic VE (see "Methods" section for details). We also tested the sensitivity of our results to variations in test sensitivity, the time since the start of the study (i.e., calendar time in observational studies), and, for the test-negative design, to differences in the prevalence of COVID-like symptoms from alternate etiologies.

Under equal testing, the cohort design produced unbiased VE_{RR}, which remained stable over the successive sampling periods (Fig. 3a, b). Under equal testing, the test-negative design produced unbiased VE_{OR} for studies limited to early sampling periods (e.g., day 30), which captured only the beginning of the epidemic (similar to VE_{RR}) (Fig. 3e, f). As the epidemic progressed, cumulative symptomatic infections increased, especially among unvaccinated individuals (Fig. 3c, d). Later in the epidemic, this disproportionate rise among the unvaccinated led to a disproportionate decrease in the cumulative proportion of unvaccinated individuals exhibiting COVID-like symptoms from alternate etiologies who had no prior symptomatic SARS-CoV-2 infection (Supplementary Fig. 1). Consequently, the test-negative design overestimated true symptomatic VE for longer studies that captured later stages of the epidemic.
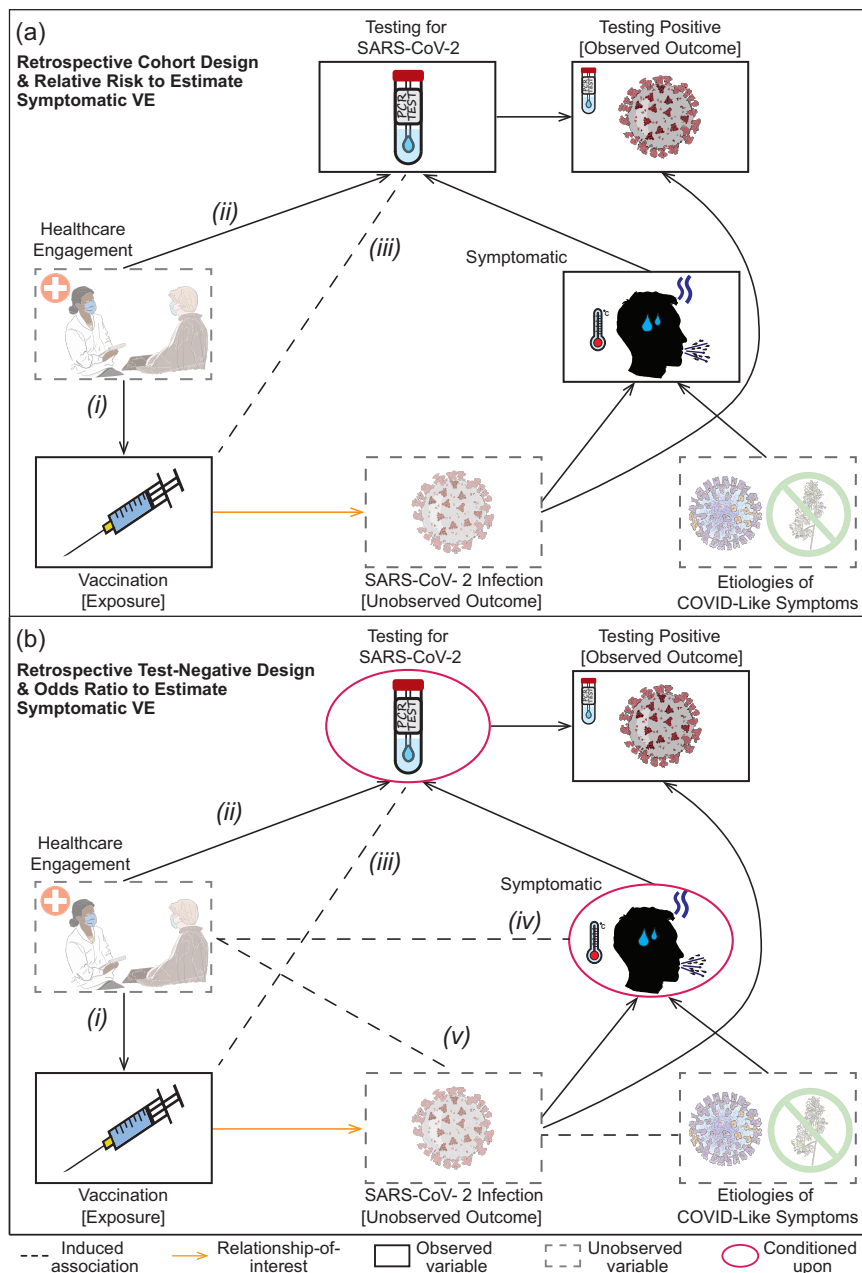
Fig. 1 | Directed acyclic graphs (DAGs) for estimating vaccine effectiveness against symptomatic SARS-CoV-2 infection across two study designs. DAGs are shown for the retrospective cohort design using relative risk (**a**) and the retrospective test-negative design using an odds ratio (**b**). The DAGs depict causal relationships (arrows) and induced associations (dashed lines). Because the vaccine does not decrease the likelihood of developing symptoms once infected, effectiveness against symptomatic infection is equivalent to effectiveness against infection (yellow arrow). Causal relationships between healthcare engagement and vaccination (i), and between healthcare engagement and testing for SARS-CoV-2 (ii) induce an association between vaccination and testing for SARS-CoV-2 (iii) and create a biased relationship between vaccination and testing positive for SARS-CoV-2 in the cohort design (depicted in (**a**)). In the test-negative design (depicted in (**b**)), conditioning on testing induces an association between healthcare engagement and being symptomatic (iv). Conditioning additionally on being symptomatic induces an association between healthcare engagement and SARS-CoV-2 infection (v). This induced relationship (v), along with the relationship between healthcare engagement and vaccination (i), introduces confounding by healthcare engagement in the relationship between vaccination and SARS-CoV-2 infection.

Under unequal testing, the cohort design produced VE_{RR} that underestimated the true symptomatic VE (Fig. 3a, b). Over the course of the epidemic (and thus for studies that captured longer sampling periods), the magnitude of the underestimate decreased and then stabilised (Fig. 3a, b). This decrease coincided with unvaccinated individuals being tested later than vaccinated individuals (Supplementary 1: Timing of Testing; Supplementary Table 1). Stabilisation of the pattern occurred at the end of the epidemic when there were no more individuals with SARS-CoV-2 available for testing (Fig. 3c, d). In sensitivity analyses, when we accounted for the influence of time using a Cox model, the temporal pattern of the bias due to unequal testing was substantially reduced, especially with lower vaccine efficacy (0.1) (Supplementary Fig. 2; Supplementary Tables 2–3). The magnitude of the underestimates was larger in the highly unequal testing scenario than in the moderately unequal testing scenario (Fig. 3a, b). Varying the test sensitivity had no effect on the magnitude of the bias (Supplementary Fig. 3).

Under unequal testing, the test-negative design also produced VE_{OR} that underestimated the true symptomatic VE, with larger biases occurring in the highly unequal testing scenario
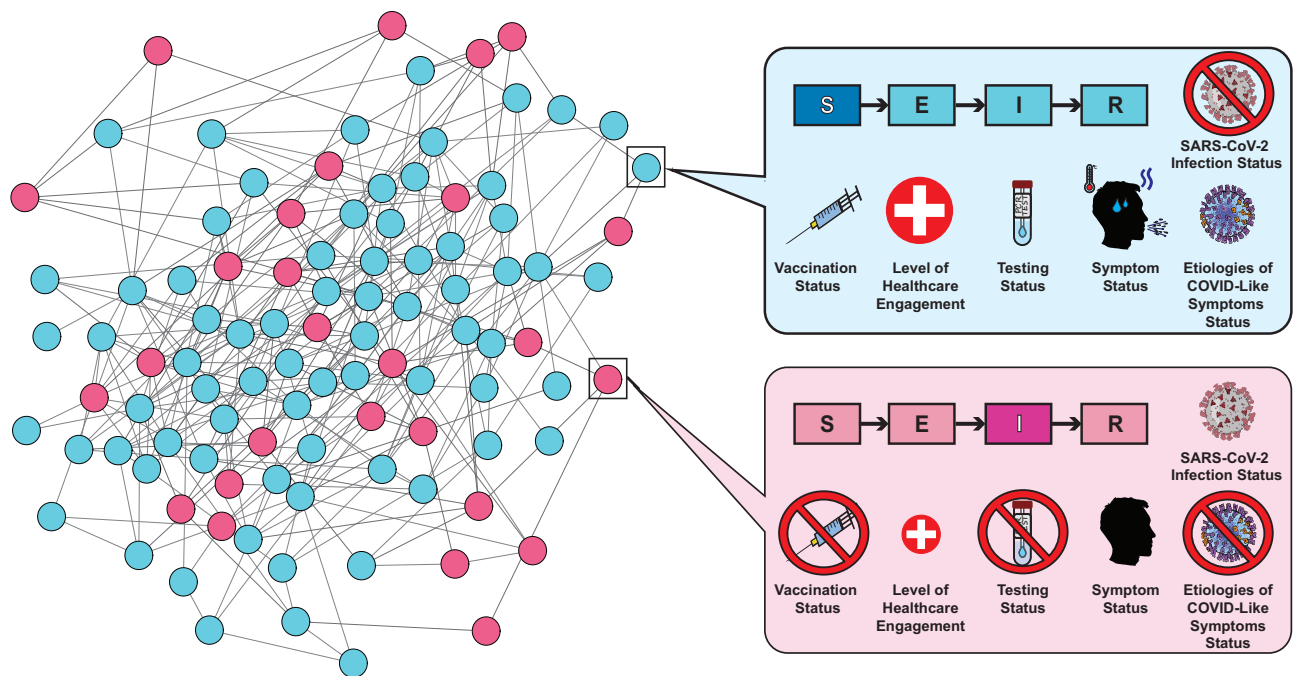
Fig. 2 | Example of the agent-based network model used to generate SARS-CoV-2 transmission dynamics. Nodes represent individuals, either vaccinated (blue) or unvaccinated (pink); edges represent contacts through which respiratory transmission can occur. Each node has individual-level attributes, including SARS-CoV-2 infection status (Susceptible [never infected], Exposed, Infectious, Recovered [previously infected]); vaccination status (vaccinated, unvaccinated); level of healthcare engagement (low, high); testing status (untested, tested positive for SARS-CoV-2, tested negative for SARS-CoV-2); symptom status (symptomatic, asymptomatic/no symptoms); and status with respect to alternate etiologies of COVID-like symptoms (exposed or unexposed).

(Fig. 3e, f). However, the test-negative design produced smaller biases than the cohort design, with the difference in the magnitude of bias between study designs most pronounced when true symptomatic VE (i.e., efficacy against susceptibility) was low.

In contrast to the cohort design, bias in the test-negative design increased for studies with longer sampling periods that included later epidemic stages, especially when true symptomatic VE was low. A substantial contributor to this pattern was the greater differential outcome misclassification by vaccination status, which accumulated over the course of the epidemic (Supplementary Fig. 4). This misclassification, driven by missed detection, increased with longer sampling periods because, as time progressed, individuals had more opportunities to both (1) become infected with symptomatic SARS-CoV-2 and be undetected (i.e., missed) and (2) develop COVID-like symptoms from alternate etiologies—sometimes more than once—and be erroneously classified as "negative" when they were undetected cases (Supplementary 1: Missed Detection Misclassification Example; Supplementary Fig. 5). In sensitivity analyses, adjusting for time since the start of a study using a logistic model reduced the magnitude of bias in studies with shorter sampling periods, particularly with lower vaccine efficacy (0.1); however, the overall pattern of increasing bias with longer sampling periods remained unchanged (Supplementary Fig. 6; Supplementary Tables 4–5).

Varying test sensitivity had a minor effect on the magnitude of bias due to unequal testing for the test-negative design, with the size of this effect slightly increasing when the study sampling period extended to the epidemic peak (Supplementary Fig. 7). Overall, the impact of test sensitivity remained consistently small ($|\Delta| < 0.045$). The direction of the test sensitivity's effect on the magnitude of bias was inconsistent and depended on the level of efficacy against susceptibility. At higher efficacy (0.55), decreasing test sensitivity increased the magnitude of bias, mirroring patterns observed across VE estimates when testing was equal (Supplementary Fig. 7a, c; Supplementary Fig. 8); at lower efficacy (0.1), decreasing test sensitivity reduced the magnitude of the bias (Supplementary Fig. 7b, d). This reduction

occurred because, at lower efficacy, the false negatives from decreased test sensitivity among the vaccinated made up a larger proportion of the vaccinated individuals classified as "negative" than was the case for unvaccinated individuals. This pattern reversed when efficacy was higher or when testing was equal (see Supplementary 1: Effects of Testing Sensitivity for details).

In the test-negative design, the magnitude of bias was also amplified in the moderately and highly unequal testing scenarios when there was a higher prevalence of COVID-like symptoms from alternate etiologies (Supplementary Fig. 9). This increase in the magnitude of bias occurred because, as the prevalence of COVID-like symptoms increased, unvaccinated individuals experienced a higher proportional increase in the number of individuals being tested and classified as "negative" than did vaccinated individuals (see Supplementary 1: Proportional Increase Calculations by Vaccination Status with COVID-like Symptoms due to Alternate Etiologies). The reason for this higher proportional increase is that, at higher prevalence of COVID-like symptoms from alternate etiologies, more individuals became symptomatic during the sampling period—with some experiencing repeated episodes—providing unvaccinated individuals who were unlikely to be tested more opportunities to test and be classified as "negative". Varying the prevalence of COVID-like symptoms from alternate etiologies had no influence on the magnitude of bias in estimates derived from the cohort design.

Influence of immunological properties on the magnitude of bias from testing differences

To determine whether varying immunological properties could shape the magnitude of bias due to testing differences, we repeated the above analyses but now varied the levels of efficacy against susceptibility (0.1 to 0.9) and efficacy against infectiousness (0.1 to 0.9). Here, we opted to compare studies conducted at a single time point (the point of highest positive SARS-CoV-2 epidemic growth [i.e., the first inflection point]) so that VE estimates could be easily compared across levels of vaccine efficacies.

Table 1 | Parameter values for agent-based model, including default values, reference(s) for the parameter values and the range of sensitivity values for those parameters of interest

Parameter	Default[a] value, and range (if applicable)	Data notes and sources	Distribution
Average contacts per person per day	6	Average estimate from prior empirical studies in the United States and the United Kingdom during the pandemic[60].	Poisson
Time from SARS-CoV-2 exposure to infectiousness	4 days	74	Poisson
Duration of SARS-CoV-2 infectiousness	10 days	75	Poisson
Probability of SARS-CoV-2 transmission per contact per day	0.1 (0.03–0.11)	Default value estimated assuming an $R_0 = 6$ using parameter values defined above with an unvaccinated population and random mixing; range values based on estimates found for ancestral SARS-CoV-2, and the Delta and Omicron variants;[33–36].	Bernoulli
Proportion of individuals who are asymptomatic if infected with SARS-CoV-2	0.35	76	Bernoulli
Proportion of individuals who are "fully" vaccinated with a COVID-19 vaccine	0.75	Based on coverage during the emergence of the SARS-CoV-2 Omicron variant in many high-income countries;[65].	
Vaccine efficacy against susceptibility	0.1, 0.55 (0.1–0.9)	Based on an average estimate and an estimate taken 9 months after vaccination for the SARS-CoV-2 Omicron variant;[66,67].	
Vaccine efficacy against infectiousness	0.2 (0.1–0.9)	Based on an average estimate for the SARS-CoV-2 Omicron variant;[68].	
Proportion (prevalence) of individuals with COVID-like symptoms from alternate etiologies	0.152	Average prevalence of symptoms of seasonal "common cold" in North America pre-COVID;[62].	
Duration of COVID-like symptoms from alternate etiologies	10 days	Average duration of symptoms with the "common cold" and influenza;[63,64]	Poisson
Testing scenarios	Ratio of testing: vaccinated/ unvaccinated		
Equal testing	1.0	No relationship between the level of healthcare engagement and the likelihood of vaccination. This was implemented by setting the proportion of vaccinated and unvaccinated individuals with high healthcare engagement to be equal (0.55). The testing probabilities for high and low healthcare engagement were set to be 0.85 and 0.15, respectively. See Supplementary 2: Testing Scenarios for implementation details	
Moderately unequal testing	1.76	A moderately strong relationship was assumed between the level of healthcare engagement and the likelihood of vaccination ("i" in Fig. 1). This was implemented by setting the proportion of vaccinated individuals with high healthcare engagement to be 2.5 times higher than that of unvaccinated individuals (0.55 vs. 0.22), following empirical survey data from Glasziou et al.[10]. The testing probabilities for high and low healthcare engagement were set to be 0.85 and 0.15, respectively. See Supplementary 2: Testing Scenarios for implementation details.	
Highly unequal testing	2.36	A strong relationship was assumed between the level of healthcare engagement and the likelihood of vaccination ("i" in Fig. 1). This was implemented by setting the proportion of vaccinated individuals with high healthcare engagement to be 5 times higher than that of unvaccinated individuals (0.55 vs. 0.11). The testing probabilities for high and low healthcare engagement were set to be 0.85 and 0.15, respectively. See Supplementary 2: Testing Scenarios for implementation details.	

[a]The parameter value used in all analyses unless otherwise specified.

We found that efficacy against susceptibility had a large influence on the magnitude of bias in both study designs (Fig. 4). Both designs were also capable of producing negative estimates of the true symptomatic VE when efficacy was low, even with moderately unequal testing (Fig. 3e, f). The moderating effect of efficacy against susceptibility was more pronounced as testing by vaccination status changed from moderately to highly unequal.

Efficacy against infectiousness also had a moderating effect on the magnitude of underestimates caused by unequal testing in both study designs (Fig. 5). The moderating effect was less pronounced than that of efficacy against susceptibility. The moderating effect of efficacy against infectiousness was present across both higher and lower levels of efficacy against susceptibility.

Under unequal testing, the moderating effect of each mechanism of efficacy (efficacy against susceptibility, efficacy against infectiousness) also varied by study design. With differences in efficacy against susceptibility, the moderating effect was more pronounced with the cohort design than the test-negative design (noting also that the absolute magnitude of the bias was much smaller with the test-negative design) (Fig. 4a, c). With differences in efficacy against infectiousness, the moderating effect was more pronounced with the test-negative design than the cohort design and was most evident under scenarios of high efficacy against infectiousness (Fig. 5a, b, e, f).

Efficacy against susceptibility had a larger influence than efficacy against infectiousness and a different influence on the magnitude of bias by study design because of the differential effect that these efficacies had on the cumulative numbers of symptomatic infections by vaccination status—the main components of our VE calculations. Although higher efficacy against susceptibility and higher efficacy against infectiousness resulted in longer epidemic periods with smaller epidemic sizes (Supplementary Figs. 10–12), changes in efficacy against susceptibility led to differential reductions in the cumulative proportions of symptomatic infections by vaccination status (Fig. 4b; Supplementary Fig. 10), while changes in efficacy against infectiousness led to more proportionate reductions by vaccination status (Fig. 5c, d; Supplementary Figs. 11–12). Specifically,
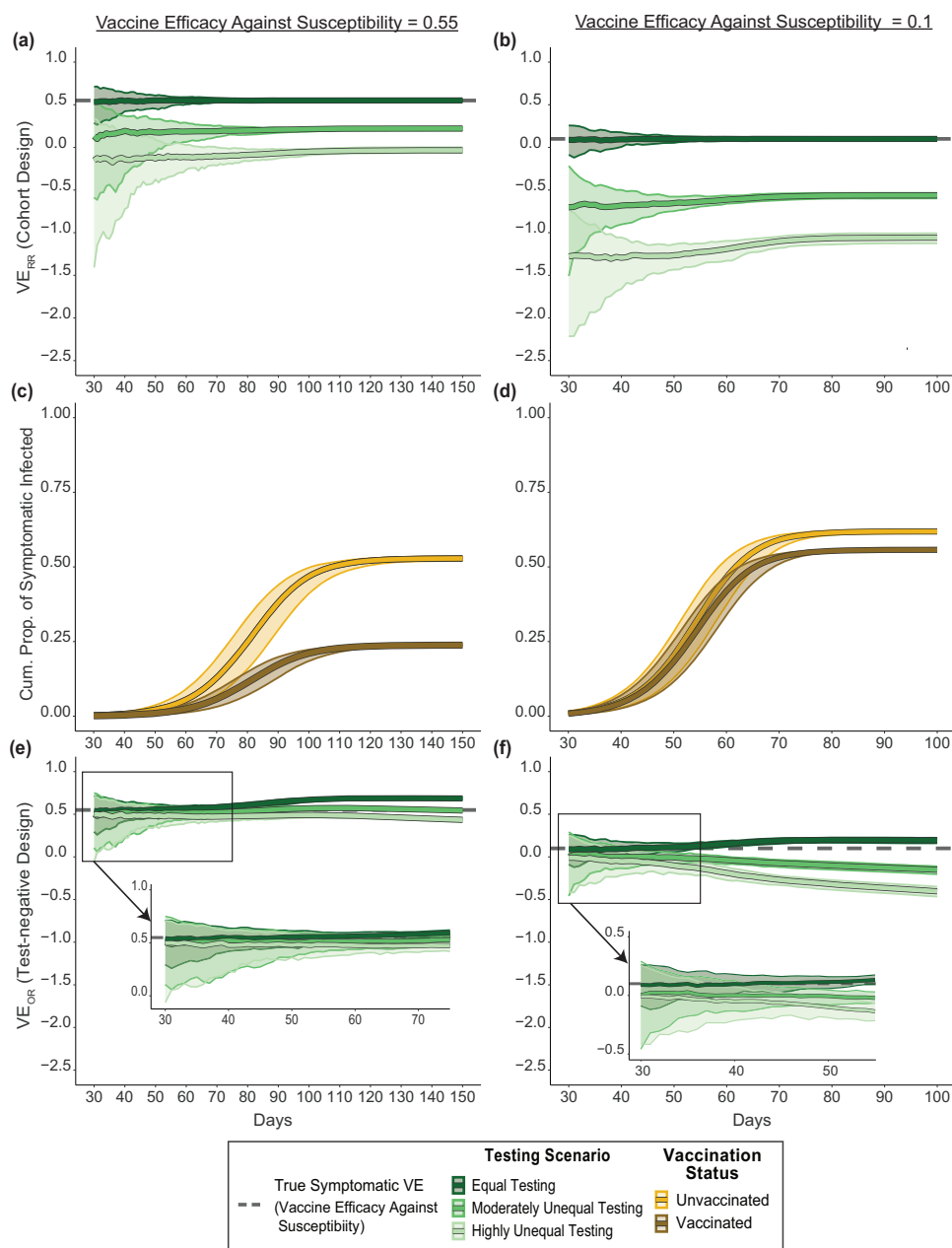
Fig. 3 | Symptomatic vaccine effectiveness (symptomatic VE) estimates across testing scenarios, study designs, and sampling periods, and cumulative proportions of symptomatic infections by vaccination status. Symptomatic VE estimates derived from the retrospective cohort design (VE_{RR}, (**a**) and (**b**)) and the retrospective test-negative design (VE_{OR}, (**e**) and (**f**)) are shown, along with the cumulative proportions of symptomatic infections by vaccination status ((**c**) and (**d**)). Panels show results under two levels of vaccine efficacy against susceptibility: higher (0.55) and lower (0.1), which represent the true symptomatic VE (grey dashed line). VE_{RR} and VE_{OR} were calculated across three testing scenarios (green): equal testing by vaccination status, moderately unequal testing (vaccinated with 1.76 times higher testing), and highly unequal testing (vaccinated with 2.36 times higher testing). Each time point corresponds to symptomatic VE estimates from studies conducted at that time during the epidemic. The cumulative proportions of symptomatic infection represent the cumulative level of symptomatic infection for unvaccinated (yellow) and vaccinated (dark gold). Central lines depict the median, and shaded regions indicate the 2.5th–97.5th percentile range across $n = 100$ epidemic realisations.

increasing efficacy against susceptibility caused symptomatic infections among the vaccinated to shrink more than among the unvaccinated, such that the absolute difference between these two groups (vaccinated versus unvaccinated) tended to monotonically increase until reaching very high levels of efficacy (Fig. 4b; Supplementary Fig. 10). The absolute difference was amplified when testing was more unequal. In contrast, changes in efficacy against infectiousness caused a similar reduction in the cumulative proportions of symptomatic infections by vaccination status as vaccinated and unvaccinated individuals both benefited from the reduction of infectiousness provided by those who were vaccinated (Fig. 5c, d; Supplementary Figs. 11–12). The consequence of this on the calculations of relative risk and odds ratios explains the differences in the moderating effects by study design and are detailed in the Supplementary Information with examples (Supplementary 1: Moderating Effect Calculations; Supplementary Figs. 13–16).
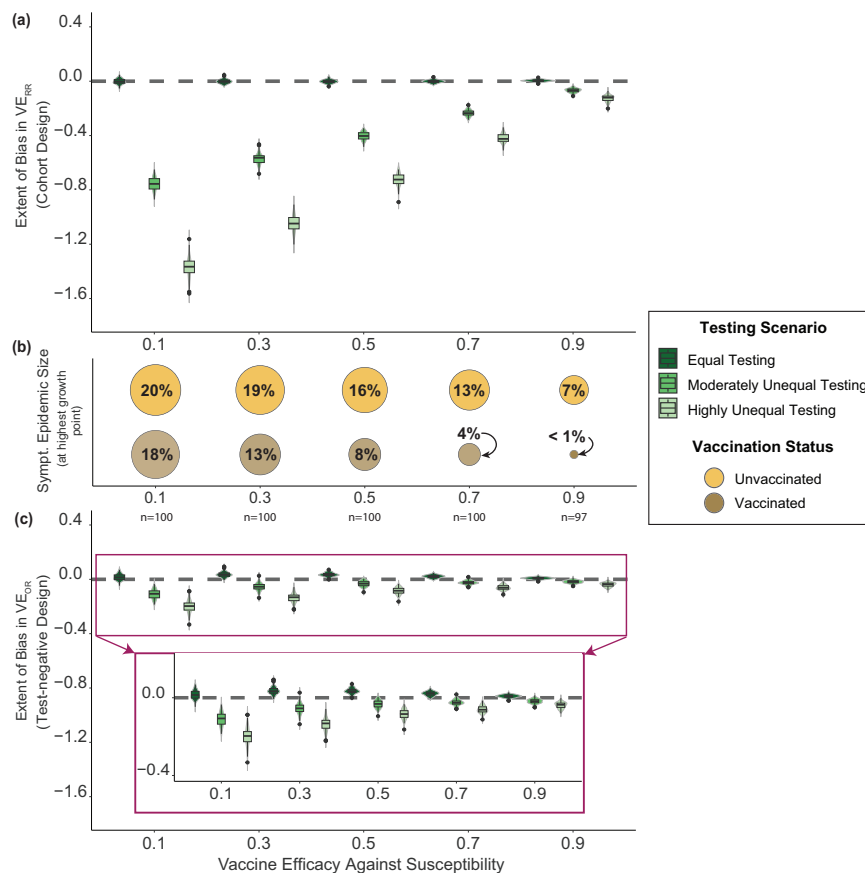
Fig. 4 | Extent of bias in symptomatic vaccine effectiveness (symptomatic VE) estimates and the epidemic size by testing scenarios and levels of vaccine efficacy against susceptibility (VE_susc). The extent of bias−how much symptomatic VE estimates diverged from the vaccine's "true" protection against symptomatic infection−is shown for estimates derived from the retrospective cohort design (VE_RR; (**a**)) and the retrospective test-negative design (VE_OR; (**c**)). VE_RR and VE_OR were calculated from studies conducted at the point of highest positive epidemic growth for each epidemic (i.e., the first inflection point), across three testing scenarios (green): equal testing by vaccination status, moderately unequal testing (vaccinated with 1.76 times higher testing), and highly unequal testing (vaccinated with 2.36 times higher testing). The epidemic size−the cumulative proportions of

individuals with symptomatic SARS-CoV-2 who were unvaccinated (yellow) and vaccinated (dark gold) (**b**)−was also recorded at that time point. Simulated epidemics with fewer than 0.1% cumulative infections were removed. Each boxplot shows the median (centre line), the interquartile range (IQR; the box spans the 25th and 75th percentiles), and whiskers that extend to the most extreme data points within 1.5 times the IQR. The violin plot outlines illustrate the kernel probability density. Boxplots and violin plots are based on $n = 100$ epidemic realisations per testing scenario and level of VE_susc, except when VE_susc = 0.9 ($n = 97$). The median time to the point of highest positive epidemic growth by the level of VE_susc was $t = 50$ for 0.1, $t = 58$ for 0.3, $t = 68$ for 0.5, $t = 92$ for 0.7, and $t = 185$ for 0.9.

Influence of underlying epidemic potential (probability of transmission) on the magnitude of bias from testing differences

To determine whether changing epidemic properties could shape the magnitude of bias due to testing differences, we varied the underlying epidemic potential−the probability of transmission (0.033 to 0.11[33–36])−and assessed the effect on the extent of bias in VE estimates. We again compared studies conducted at a single time point (the highest SARS-CoV-2 epidemic growth point). Similar to the effect of efficacy against infectiousness, decreasing the probability of transmission led to a similar, proportionate reduction in the cumulative proportions of symptomatic infections by vaccination status (Fig. 6c, d; Supplementary Figs. 17–18). As such, the moderating influence of the probability of transmission demonstrated a similar pattern to that of efficacy against infectiousness. In the context of high transmission, changes in transmission had a negligible influence on the magnitude of bias produced in both study designs (similar to the pattern observed when efficacy against infectiousness was low) (Fig. 6a, b, e, f). However, in the context of low transmission, small changes in transmission led to larger changes in the magnitude of bias, but only in the test-negative design (similar to the pattern observed when efficacy against infectiousness was high) (Fig. 6e, f).

The difference in the direction and magnitude of bias by study design occurred because of how the ratio of the numerators−the cumulative number of symptomatic SARS-CoV-2 cases in vaccinated versus unvaccinated individuals−changed with the probability of transmission in the cohort design (i.e., in the relative risk), and how both the ratio of cumulative symptomatic cases and the ratio of cumulative symptomatic controls in vaccinated versus unvaccinated individuals (i.e., the numerators and denominators, respectively) changed in the test-negative design (i.e., in the odds ratio) (Supplementary Figs. 19–20). These patterns again are similar to those observed with efficacy against infectiousness. A detailed explanation and examples illustrating how these calculations vary with the probability of transmission and differ by study design are provided in the Supplementary Information (Supplementary 1: Moderating Effect Calculations; Supplementary Figs. 19–20).

Discussion

Using simulation modelling and SARS-CoV-2 as the motivating example, we examined the magnitude of bias in symptomatic VE due to unequal testing by vaccination status and explored how epidemic and immunological properties could shape its magnitude across two
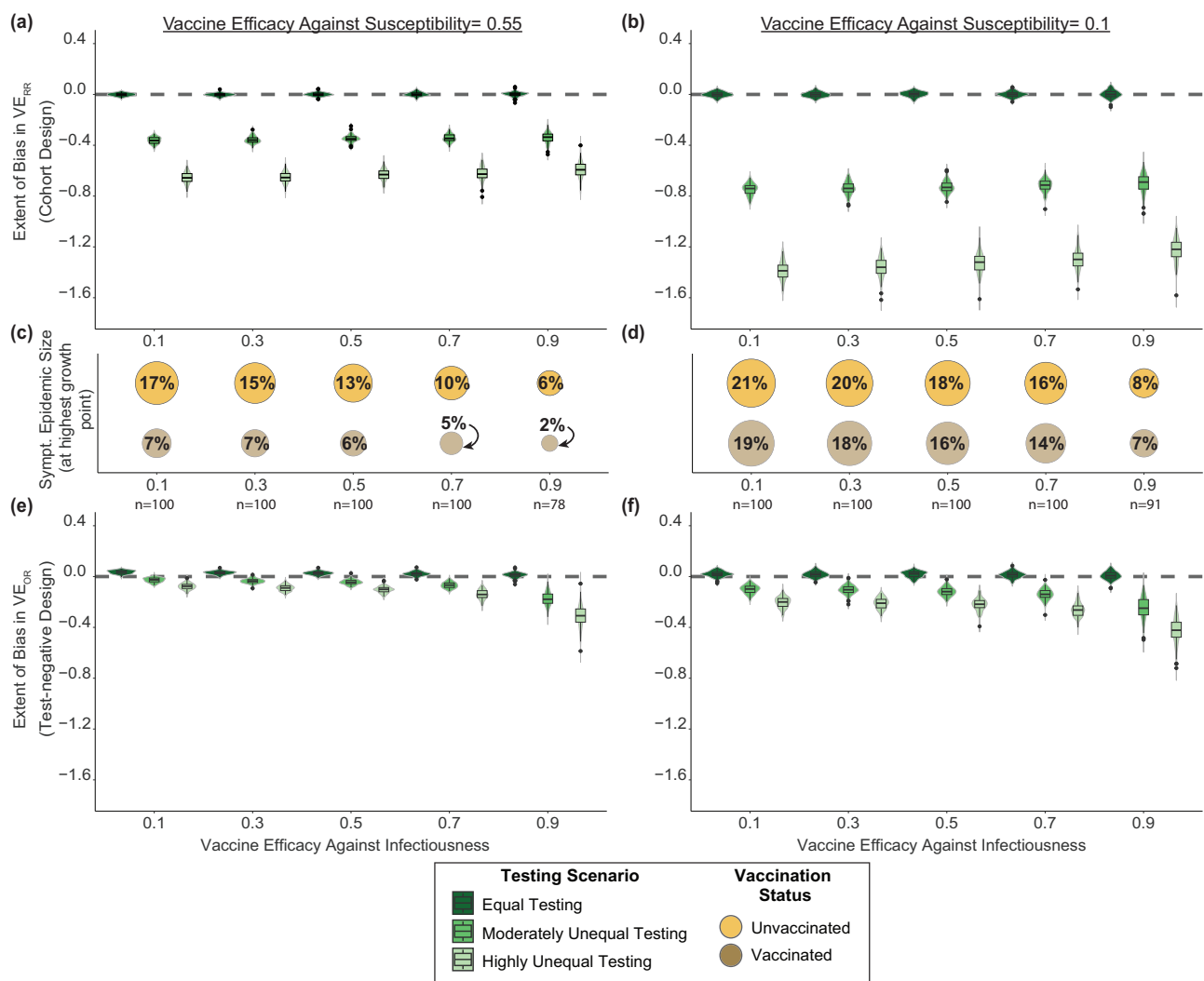
Fig. 5 | Extent of bias in symptomatic vaccine effectiveness (symptomatic VE) estimates and the epidemic size by testing scenarios and levels of vaccine efficacy against infectiousness (VE_infect) given higher and lower vaccine efficacy against susceptibility (VE_susc; 0.55 and 0.1). The extent of bias−how much symptomatic VE estimates diverged from the vaccine's "true" protection against symptomatic infection−is shown for estimates derived from the retrospective cohort design (VE_RR) ((a) and (b)) and the retrospective test-negative design (VE_OR) ((e) and (f)). VE_RR and VE_OR were calculated from studies conducted at the point of highest positive epidemic growth for each epidemic (i.e., the first inflection point), across three testing scenarios (green): equal testing by vaccination status, moderately unequal testing (vaccinated with 1.76 times higher testing), and highly unequal testing (vaccinated with 2.36 times higher testing). The epidemic size−the cumulative proportions of individuals with symptomatic SARS-CoV-2 who were unvaccinated (yellow) and vaccinated (dark gold) ((c) and (d))−was also recorded at that time point. Simulated epidemics with less than 0.1% cumulative infections were removed. Each boxplot shows the median (centre line), the interquartile range (IQR; the box spans the 25th and 75th percentiles), and whiskers that extend to the most extreme data points within 1.5 times the IQR. The violin plot outlines illustrate the kernel probability density. Boxplots and violin plots are based on $n = 100$ epidemic realisations per testing scenario and level of VE_infect, except when VE_infect = 0.9 ($n = 78$ when VE_susc is higher [0.55]; $n = 91$ when VE_susc is lower [0.1]). When VE_susc was higher (0.55), the median time to the point of highest positive epidemic growth across VE_infect was $t = 69$ for 0.1, $t = 79$ for 0.3, $t = 93$ for 0.5, $t = 123$ for 0.7, and $t = 290$ for 0.9; when VE_susc was lower (0.1), the median time was $t = 48$ for 0.1, $t = 54$ for 0.3, $t = 64$ for 0.5, $t = 85$ for 0.7, and $t = 167$ for 0.9.

commonly adopted observational VE study designs. Similar to Suah et al.[28], we found that the presence of unequal testing by vaccination− where unvaccinated individuals tested less than vaccinated individuals −resulted in an underestimation of the true symptomatic VE in both the cohort and test-negative designs, with greater testing differences resulting in larger underestimates. As expected, the cohort design consistently produced larger underestimates (larger biases) than the test-negative design. The largest bias for both study designs occurred with low efficacy against susceptibility, but the moderating effect of efficacy against susceptibility on bias was more pronounced in the cohort design compared with the test-negative design. In contrast, the magnitude of bias due to unequal testing in the test-negative design was more likely than that in the cohort design to be influenced by the study's selected sampling period, the efficacy against infectiousness, and the probability of transmission.

Our findings confirm and build on previous insights on the mechanisms by which unequal testing and healthcare engagement can create biased estimates of VE in cohort and test-negative designs[6–8,21–23]. First, our work expands on previous qualitative studies that showcased the potential for residual confounding and selection bias in VE estimates[6–8]. Specifically, we illustrated how selection bias in test-negative designs can play an important role in affecting estimates of true symptomatic VE−although the magnitude of the bias is likely to be small. Second, our work provides a detailed comparison of the magnitude of bias found across the two most commonly adopted study designs−the retrospective cohort and test-
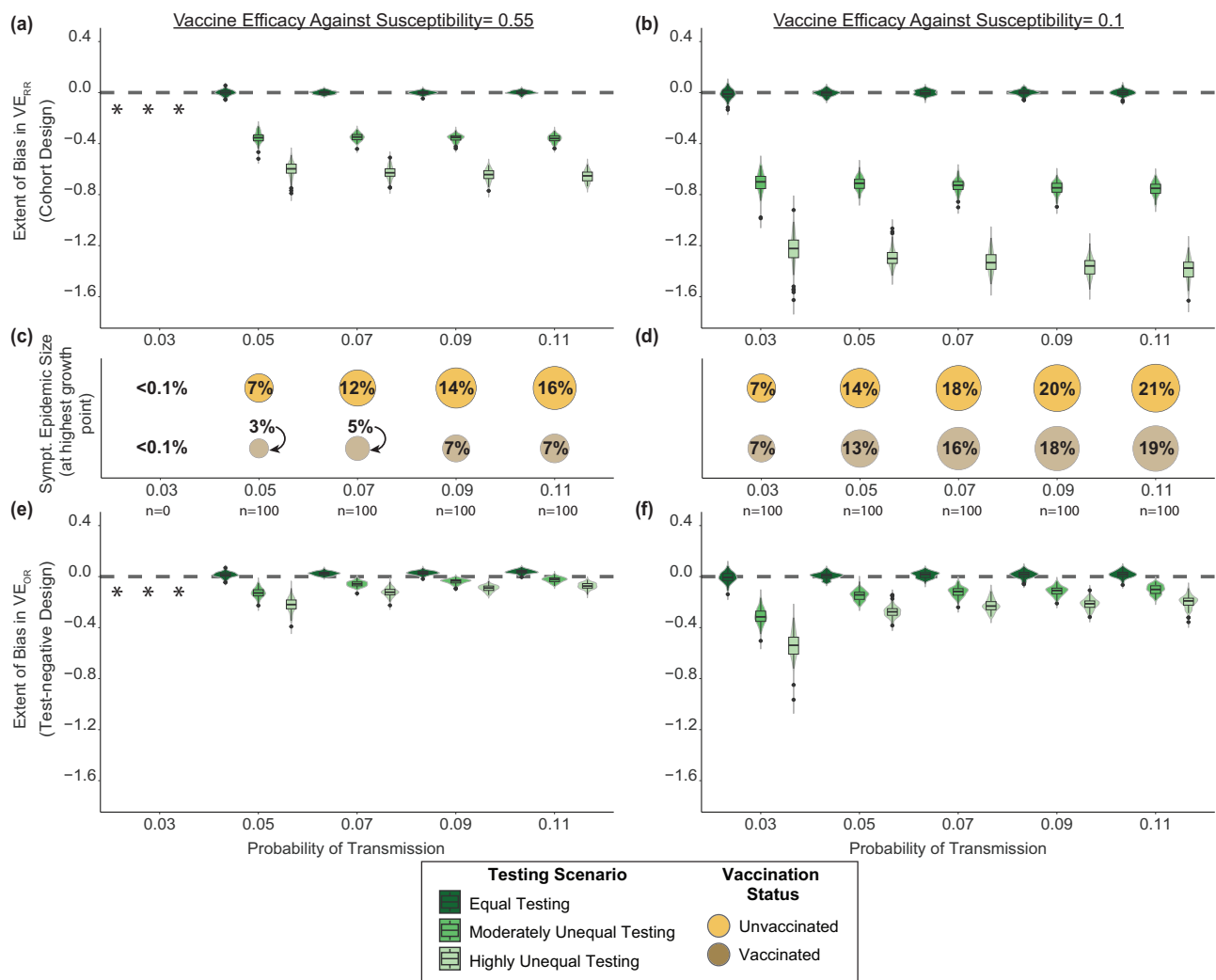
Fig. 6 | Extent of bias in symptomatic vaccine effectiveness (symptomatic VE) estimates and the epidemic size by testing scenarios and probability of transmission (P_{trans}) for higher and lower vaccine efficacy against susceptibility (VE_{susc}; 0.55 and 0.1). The extent of bias—how much symptomatic VE estimates diverged from the vaccine's "true" protection against symptomatic infection —is shown for estimates derived from the retrospective cohort design (VE_{RR}; (a) and (b)) and the retrospective test-negative design (VE_{OR}; (e) and (f)). VE_{RR} and VE_{OR} were calculated from studies conducted at the point of highest positive epidemic growth for each epidemic (i.e., the first inflection point), across three testing scenarios (green): equal testing by vaccination status, moderately unequal testing (vaccinated with 1.76 times higher testing), and highly unequal testing (vaccinated with 2.36 times higher testing). The epidemic size—the cumulative proportions of individuals with symptomatic SARS-CoV-2 who were unvaccinated (yellow) and vaccinated (dark gold) ((c) and (d))—was also recorded at that time point. Simulated epidemics with less than 0.1% cumulative infections were removed. If the extent of bias could not be calculated, estimates were replaced with an asterisk. Each boxplot shows the median (centre line), the interquartile range (IQR; the box spans the 25th and 75th percentiles), and whiskers that extend to the most extreme data points within 1.5 times the IQR. The violin plot outlines illustrate the kernel probability density. Boxplots and violin plots are based on $n = 100$ epidemic realisations per testing scenario and level of P_{trans}. When VE_{susc} was higher (0.55), the median time to the point of highest positive epidemic growth across P_{trans} was $t = 202$ for 0.05, $t = 108$ for 0.07, $t = 82$ for 0.09, and $t = 67$ for 0.11; when VE_{susc} was lower (0.1), the median time was $t = 224$ for 0.03, $t = 90$ for 0.05, $t = 65$ for 0.07, $t = 54$ for 0.09, and $t = 47$ for 0.11.

negative design. Our findings add support to the use of the test-negative design over that of the cohort design in estimating VE in the presence of unequal testing by vaccination status[24]. We further identified important epidemic and immunological properties that should be considered when interpreting VE estimates from both study designs: efficacy against susceptibility, efficacy against infectiousness, and the probability of transmission.

Efficacy against susceptibility had the largest influence on the magnitude of bias in VE estimates, with a more pronounced effect on estimates from the cohort design. Other studies have also found that efficacy against susceptibility can moderate the magnitude of bias produced by other mechanisms of bias, such as heterogeneity in contact levels by vaccination status[37] and confounding bias due to

correlated case and control vaccination behaviours[2]. The key implication from these findings is that as a virus undergoes genetic adaptations leading to immune escape, and/or waning against susceptibility occurs, the risk of larger biases in VE estimates increases. This is likely why several studies found estimates of negative VE (suggesting that the vaccine was increasing symptomatic infection) during the emergence of the Omicron variant of SARS-CoV-2 (e.g.,[26,27])—a variant known to have higher immune escape[38].

There are three important implications stemming from the influence of the study's selected sampling period, efficacy against infectiousness, and the epidemic potential (probability of transmission)—all of which were more pronounced with the test-negative design than with the cohort design. First, in the context of an emerging and

evolving outbreak, VE studies are routinely conducted over successive time periods. Thus, it becomes increasingly important in the context of unequal testing to repeat VE studies and consider the timing of sampling for interpretation of VE estimates. Early in the epidemic, a test-negative design offers a large advantage in terms of minimising the magnitude of bias due to unequal testing, but it calls for greater attention as the epidemic progresses. Even with equal testing, continued caution may still be warranted, as VE estimates can become overestimated when the epidemic is large and the study's selected sampling period is longer and includes later epidemic stages. This overestimation occurs because, as the epidemic progresses, the pool of individuals who have never been infected with symptomatic SARS-CoV-2 shrinks—especially among the unvaccinated—leading to a control group that no longer accurately reflects the population's vaccination distribution.

Second, although changes in efficacy against infectiousness play an overall smaller role in moderating the magnitude of bias from unequal testing compared to efficacy against susceptibility, their effects can still be prominent when using the test-negative design. Efficacy against infectiousness is therefore important to consider in the context of the test-negative design as efficacy against infectiousness can change as vaccines are updated and as viruses, such as influenza, develop immune escape that alters their ability to transmit[39].

Third, although more pronounced in the context of low epidemic potential, the finding that changes in epidemic potential can moderate the magnitude of bias from unequal testing underscores a need for caution when comparing and interpreting VE estimates across different contexts. Respiratory viruses can have seasonal trends[40,41] and, in cases such as SARS-CoV-2 and respiratory syncytial virus, can evolve to have different transmissibility[32,42]. This means that the epidemic potential of respiratory viruses—and potentially the magnitude of bias due to testing differences—could vary across seasons, regions, and time periods. Similarly, the prevalence of COVID-like symptoms from alternate etiologies, which also influenced the magnitude of bias in symptomatic VE estimates, may also exhibit distinct geographic, seasonal, and yearly patterns that differ from those of the infection of interest. Taken together, these considerations highlight the challenge of ensuring meaningful comparisons of symptomatic VE estimates and the potential importance of accounting for seasonal, geographic and temporal variability in these comparisons.

Taken together, when pooling or synthesising VE estimates, future work is needed to account for sources of heterogeneity that affect the magnitude of potential biases to strengthen the robustness of evidence synthesis. In the meantime, careful inclusion of nuanced DAGs (across study designs) and cautious and transparent interpretation to communicate the potential benefits of vaccination are essential. Good practices also include anticipating and recognising that observed VE may represent biased VE estimates in the presence of unequal testing by vaccination status. These practices are important given that addressing unequal testing remains challenging. Electronic health records in national databases typically have information on exposures and health outcomes, but contain limited data on confounders[43]. We therefore echo the sentiments of Lewnard et al.[44] cautioning that the test-negative design—along with other types of study designs—should not be naively applied to testing datasets collected through administrative or public health surveillance data sources. An area of future methodological work includes using externally collected data on testing by vaccination status from representative surveys (e.g., refs. 9,10) to conduct quantitative bias analysis, including simulation-based bias analysis[45,46].

Our findings also have utility beyond VE studies as they can offer insights into how other forms of unequal testing may impact estimates of protective effects, such as natural immunity. Testing differences by prior infection status have been observed[47]. Had we replaced vaccination with prior infection in our study and instead estimated the protection of natural immunity, we would expect that unequal testing by prior infection would lead to biased estimates. This expectation aligns with other studies that have noted the potential for such unequal testing to compromise protective effect estimates[48,49]. Given this, the good practices and future methodological work for interpreting and addressing potential bias from unequal testing in VE estimates can also be applied to estimates of other protective effects, such as natural immunity, that may be affected by testing differences.

Our simulation study has several limitations. Chief among them is our assumption that vaccination coverage was static. Rolling out a vaccine over time could potentially reduce the magnitude of bias due to unequal testing if, for example, individuals who have high healthcare engagement (and hence are likely to be tested) also wanted to be vaccinated but were not able to immediately receive one. However, this effect could also be reversed if interventions initially reached those with the highest healthcare engagement first before being distributed to those with lower healthcare engagement. This pattern could exacerbate unequal testing by vaccination early in the epidemic before stabilising as the vaccine is distributed to those with lower healthcare engagement.

Second is our assumption that individuals can only be infected with SARS-CoV-2 once. The presence of prior infections can contribute to underestimating symptomatic VE[50]. Therefore, including prior undetected infections before the start of each study's sampling period may have further exacerbated the underestimation of VE due to unequal testing by vaccination status. Allowing for multiple infections throughout the sampling period and estimating hybrid immunity could also have underestimated this protective effect if testing was unequal and unvaccinated individuals—with or without natural immunity—remained a reference group. The magnitude of this bias would, however, depend on how testing varied by vaccination status, prior infection status, and the interaction between the two.

Third, we had assumed that for those infected with SARS-CoV-2, the severity and duration of symptoms did not vary by vaccination status. Vaccination for SARS-CoV-2 has been shown to reduce both the severity of symptoms and symptom duration[51–53], which could lead to vaccinated individuals testing less frequently when symptomatic or having fewer days to test. This could, in turn, reduce the magnitude of bias due to unequal testing. However, the magnitude of bias could remain unchanged if vaccinated individuals were also more likely than unvaccinated individuals to get tested with milder symptoms and to test earlier.

Fourth, our scenarios were conditioned on vaccinated individuals testing more than unvaccinated individuals based on several surveys related to SARS-CoV-2 testing and vaccination (e.g., refs. 9, 10). However, unequal testing has also occurred in the opposite direction[54], which is an important consideration when comparing VE studies across different study settings, irrespective of pathogen. If unequal testing reflects higher testing among unvaccinated individuals, the direction of the bias would change, leading to an overestimation rather than an underestimation of symptomatic VE.

Finally, for simplicity, we assumed that the vaccine provided "all-or-nothing" protection rather than "leaky" protection against infection. This means the vaccine induced 100% protection for a subset of vaccinated individuals and none for the remainder rather than offering partial protection to every vaccinated individual[55]. Had we assumed leaky protection and not excluded those with past infection[56], the increasing magnitude of bias in our symptomatic VE estimates with unequal testing would likely have increased further. This is because even with equal testing, leaky protection introduces downward bias in VE estimates over the course of an epidemic as more uninfected and unvaccinated individuals who are not at risk of infection are included in the analysis[25,56].

Observational studies are key to assessing how well vaccines might be working in the real world[2]. We illustrated how unequal testing

by vaccination status could bias estimates of true symptomatic VE across two commonly used study designs. Overall, using a test-negative design rather than a cohort design can reduce the magnitude of bias due to unequal testing when the mechanisms driving unequal testing are unmeasured. However, interpretation of VE estimates from both study designs still requires caution, especially given heterogeneity in epidemic and immunological properties—with the most important property being vaccine efficacy against susceptibility.

Methods
Mechanisms of interest
The causal relationship we are interested in is the effect of vaccination on symptomatic SARS-CoV-2 infection (Fig. 1). For simplification, we assumed that vaccination does not reduce symptoms in vaccinated individuals who acquire SARS-CoV-2, and thus the effect of vaccination on infection is the same as its effect on symptomatic infection. If infected, an individual may develop symptoms, which leads to an opportunity to engage in SARS-CoV-2 testing. Both infectious and non-infectious etiologies other than SARS-CoV-2 can lead to symptoms similar to those caused by SARS-CoV-2 (referred to as etiologies of COVID-like symptoms, Fig. 1)[57].

We assumed that the association between vaccination and testing (pathway marked as "iii" in Fig. 1, which reflects unequal testing by vaccination status) is induced by a relationship between the level of healthcare engagement and vaccination (pathway marked as "i" in Fig. 1) and a relationship between the level of healthcare engagement and testing (pathway marked as "ii" in Fig. 1).

In the cohort design, everyone was included in the study sample. Figure 1a depicts how the assumed relationships between healthcare engagement, vaccination, and testing can lead to residual confounding in VE estimates when using the cohort design. Specifically, the existence of relationships between the level of healthcare engagement and vaccination ("i" in Fig. 1a) and between the level of healthcare engagement and testing ("ii" in Fig. 1a) creates a biasing pathway between vaccination and testing positive for SARS-CoV-2.

In the test-negative design, the study sample was restricted to individuals who were symptomatic and tested for SARS-CoV-2 (depicted by the two "conditioned upon" circles in Fig. 1b). Conditioning on testing aims to block the back-door path between vaccination and testing positive for SARS-CoV-2 that exists due to unmeasured healthcare engagement; however, this conditioning can also introduce a selection bias[8]. Selecting additionally on being symptomatic aims to block the selection bias, but it can also lead to another bias. Specifically, conditioning on testing introduces a relationship between healthcare engagement and being symptomatic ("iv" in Fig. 1b), while additionally conditioning on being symptomatic induces an association between healthcare engagement and SARS-CoV-2 infection ("v" in Fig. 1b). This induced association ("v" in Fig. 1b), together with the existing causal relationship between healthcare engagement and vaccination ("i" in Fig. 1b), results in confounding that biases the relationship between vaccination and SARS-CoV-2 infection (see Supplementary 2: Selection Bias for details).

To focus our examination on the mechanism of bias from unequal testing by vaccination status, we assumed no other sources of unmeasured confounding.

Model overview
We used an agent-based model to simulate SARS-CoV-2 dynamics and to model the individual-level variations in vaccination and testing (Fig. 2). We chose an agent-based model as it is a flexible approach that can capture emergent phenomena arising from complex interactions in a heterogeneous population[58]. In our model, each individual (i.e., agent) was assigned attributes to capture health states related to SARS-CoV-2: Susceptible, Exposed, Infectious, and Recovered (where individuals are assumed to be fully protected

from reinfection). Other individual-level attributes that vary over time include: symptom status (symptomatic or asymptomatic), SARS-CoV-2 testing (tested or not tested); SARS-CoV-2 diagnosis if tested (tested positive vs. negative). Attributes that do not vary over time include: healthcare engagement (low or high) and vaccination status (vaccinated or unvaccinated). The model updates each individual's attributes daily. Table 1 summarises the parameters and their corresponding data sources.

Contact patterns
Transmission occurs in daily time-steps through contacts in a homogenous and static network generated using an Erdős-Rényi random graph model[59]. The network is based on an average of 6 daily contacts[60]. Given the network is sparse, the distribution of contacts across the population follows a Poisson distribution[61]. The original placement of edges (representing contacts) is homogeneous by vaccination status, and all nodes (individuals) are connected to the main network (Supplementary 2: Contact Patterns).

SARS-CoV-2 transmission
Based on available data from the literature, we assigned a default value for the biological probability of transmission per contact per day, the duration of the exposure state, the duration of infectiousness, and the proportion asymptomatic if infected with SARS-CoV-2 (Table 1). Each of these processes is stochastic and was implemented using draws from statistical distributions (Supplementary 2: SARS-CoV-2 Transmission Dynamics).

COVID-like symptoms
At any given point in time, a stable proportion (default value of 0.152 in our primary analysis) of the simulated population experienced COVID-like symptoms from alternate etiologies[62]. We assumed symptoms lasted on average for 10 days, reflecting the average duration of symptoms from the "common cold"[63] and influenza[64] (Supplementary 2: COVID-like Symptoms). Individuals could experience COVID-like symptoms from alternative etiologies more than once.

Vaccination (coverage and mechanisms of protection)
We applied a fixed 75% vaccination coverage[65] such that 75% of the simulated population was fully vaccinated prior to the start of the epidemic, with no further vaccination thereafter. This reflects a scenario wherein a variant may emerge, but a new or additional vaccination has not yet been rolled out.

We simulated vaccine efficacy against susceptibility to acquisition of SARS-CoV-2 and vaccine efficacy against infectiousness[55]. Efficacy against susceptibility was implemented as "all-or-nothing" protection. That is, the vaccine induced 100% protection for a subset of vaccinated individuals (equal to the efficacy against susceptibility) and zero protection for the remainder (i.e., 1 - efficacy)[55]. Efficacy against infectiousness was implemented as a reduced probability of SARS-CoV-2 transmission for vaccinated individuals infected with the virus, given contact with a susceptible individual. We assumed that vaccination did not reduce the probability of becoming symptomatic and that vaccine efficacies did not wane over time.

Most of our analyses used a default value for higher and lower efficacy against susceptibility (0.55, informed by an average estimate of VE against Omicron infection[66]; and 0.1, reflecting VE estimates taken 9 months after vaccination[67]). For efficacy against infectiousness, we used a default value of 0.2[68].

Symptomatic VE estimation
To mimic the retrospective cohort design in studies using population-level health administrative data, we assumed that symptomatic individuals with SARS-CoV-2 who were not tested would be misclassified as not infected when estimating true symptomatic VE using the cohort

design (Eq. 1):

$$VE_{RR}(t) = 1 - RR(t) \qquad (1)$$

with:

$$RR(t) = \frac{\frac{CI_V(t)}{N_V}}{\frac{CI_U(t)}{N_U}} \qquad (2)$$

where $RR(t)$ is the relative risk for a study conducted at time t; $CI_V(t)$ and $CI_U(t)$ are the cumulative numbers of vaccinated and unvaccinated individuals, respectively, who are symptomatic and infected with SARS-CoV-2, and have been tested and diagnosed (tested positive) for a study conducted at time t; and N_V and N_U are the total numbers of vaccinated and unvaccinated individuals, respectively.

To assess the potential influence of time in the cohort design, additional sensitivity analyses were conducted, in which true symptomatic VE was estimated using a Cox model with vaccination status as a covariate (see Supplementary 2: Symptomatic VE Estimation for details). If an individual tested more than once during their symptomatic SARS-CoV-2 infection, only the first positive test was selected. Individuals who did not test positive were censored at the end of each study's sampling period.

To mimic the test-negative design, we sampled from the simulated cohort individuals who were both symptomatic and tested. Using those sampled individuals, we estimated the true symptomatic VE using the test-negative design (Eq. 3):

$$VE_{OR}(t) = 1 - OR(t) \qquad (3)$$

with

$$OR(t) = \frac{\frac{CI_V(t)}{CU_V(t)}}{\frac{CI_U(t)}{CU_U(t)}} \qquad (4)$$

where $OR(t)$ is the odds ratio for a study conducted at time t; and $CU_V(t)$ and $CU_U(t)$ are the cumulative numbers of vaccinated and unvaccinated individuals, respectively, who are symptomatic due to alternate etiologies that cause COVID-like symptoms, who tested negative for SARS-COV-2, and had no prior positive tests for symptomatic SARS-COV-2 for a study conducted at time t. Each individual included in a study receives only one classification.

To assess the potential influence of time in the test-negative design, additional sensitivity analyses were conducted, in which true symptomatic VE was estimated using a logistic model with vaccination status and time since the start of the study (i.e., calendar day) as covariates (see Supplementary 2: Symptomatic VE Estimation for details). Individuals contributed only a single positive or negative test to a study: if an individual had at least one positive test, only the first positive test was selected; if an individual had multiple negative tests and no positive tests, one negative test was randomly selected.

Testing scenarios
We simulated three testing scenarios, each with a varying strength of the relationship between testing and vaccination ("iii" in Fig. 1): equal testing scenario (equal probability of testing by vaccination status); moderately unequal testing scenario (vaccinated with 1.76 times higher testing); and highly unequal testing scenario (vaccinated with 2.36 times higher testing). Here, the direction of the testing differences (i.e., vaccinated individuals testing more than unvaccinated) and the magnitude of the differences in the moderately unequal testing scenario are based on empirical survey results[10]. Varying "iii" also affected

the timing of testing by vaccination status. Thus, in our testing scenarios, vaccinated individuals tested earlier than unvaccinated individuals when testing was unequal, with the difference in the timing growing as unequal testing went from moderately unequal to highly unequal (Supplementary 1: Timing of Testing).

The magnitude of "iii" is influenced by the likelihood of vaccination by the level of healthcare engagement ("i" in Fig. 1); and the likelihood of testing by the level of healthcare engagement ("ii" in Fig. 1). We varied values of "i" to simulate the three testing scenarios above (see Supplementary 2: Testing and Healthcare Engagement and Testing Scenarios for details). The magnitude of "iii" remained constant throughout each epidemic realisation.

Simulations and analyses
Each scenario consisted of 100 epidemic realisations and each epidemic realisation used a population size of 100,000. We selected these values because in scenarios using our default parameters, estimates of symptomatic VE had stabilised by the point of highest positive epidemic growth for SARS-CoV-2 (i.e., the first inflection point). All realisations were initialised with 10 randomly assigned individuals on day 1 of their infectiousness (see Supplementary 2: Event Scheduling). In our analyses, we estimated symptomatic VE using a cohort design and test-negative design (VE_{RR} and VE_{OR}) under the three testing scenarios. As previously noted, vaccination coverage, immune escape, and differences in testing were all assumed to be constant over time and no additional time-related confounders were included in the model.

The magnitude of bias was measured as the difference between efficacy against susceptibility (the model input) and the estimates of symptomatic VE produced by the cohort design (VE_{RR}) and the test-negative design (VE_{OR}). We assessed the magnitude of bias using efficacy against susceptibility as it represents the true symptomatic VE. Efficacy against susceptibility is equivalent to true symptomatic VE because (1) the vaccine provides all-or-nothing protection[55]; and (2) the vaccine does not reduce the likelihood of becoming symptomatic once infected. We performed model verification to confirm the accuracy of our model and our symptomatic VE estimates (Supplementary 2: Model Verification; Supplementary Figs. 21–22). Results are presented as medians with variability summarised using either 2.5th–97.5th percentile ranges or boxplots.

To explore how estimates of symptomatic VE vary across study designs and sampling periods (Objective 1), we calculated VE_{RR} and VE_{OR} (cumulative measurements as shown in formulas 1 and 2) across studies with varying sampling periods under the three testing scenarios. Specifically, we began on day 30 (representing a study with a 30-day sampling period) and then repeated the calculations at each subsequent time point (representing studies with progressively larger sampling periods) until the end of the epidemic. We conducted these analyses using both higher and lower default values of efficacy against susceptibility (0.55 and 0.1, respectively). To assess the potential influence of time on estimates of symptomatic VE, we performed additional sensitivity analyses using models that accounted for time since the start of the study. We also varied the prevalence of etiologies of COVID-like symptoms (0.05–0.2) and test sensitivity (0.5–1) to explore their potential influences on the magnitude of bias under the three testing scenarios. The influence of test sensitivity was assessed by considering only each individual's first positive test within an infection period. As the effects of test sensitivity are known to change with the ratio of true disease cases to controls[69], we evaluated its influence in studies conducted at two distinct stages of the epidemic: (1) the point of highest positive epidemic growth (i.e., the first inflection point); and (2) the epidemic peak.

After estimating symptomatic VE across studies with different sampling periods, we then calculated VE_{RR} and VE_{OR} (with default

parameters as per Table 1) under a range of epidemic and immunological properties to determine whether varying these properties could shape the magnitude of the bias due to unequal testing (Objectives 2 and 3). First, we varied efficacy against susceptibility (0.1–0.9) and efficacy against infectiousness (0.1–0.9). We then varied the underlying epidemic potential—the probability of transmission (0.033–0.11[33–36]).

Results for Objective 1 were reported across a range of study sampling periods (30–150 and 30–100 days for higher and lower efficacy against susceptibility, respectively); results for Objectives 2 and 3 were reported as the difference between VE_{RR} and VE_{OR} and true symptomatic VE (i.e., the level of efficacy against susceptibility) for studies conducted at a single time point—the point of highest positive epidemic growth for SARS-CoV-2 (i.e., the first inflection point). A single time point was used for Objectives 2 and 3 so that symptomatic VE estimates could be easily compared across different levels of epidemic potential and immunological protection.

All simulations were conducted using R[70] (version 4.3.1) with computations performed on the Niagara supercomputer at the SciNet HPC Consortium[71,72] (Supplementary 2: Computing Resources).

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
The simulated data required to reproduce the main and Supplementary Figures, tables, and analyses are available as Source Data and on GitHub: https://github.com/mishra-lab/testing_bias/releases/tag/v1.0.0[73]. Source data are provided with this paper.

Code availability
The code used to generate the simulation data and to produce the main and Supplementary Figures, tables, and analyses is available on GitHub: https://github.com/mishra-lab/testing_bias/releases/tag/v1.0.0[73].

References
1. Chua, H. et al. The use of test-negative controls to monitor vaccine effectiveness: a systematic review of methodology. *Epidemiology* **31**, 43 (2020).
2. Doll, M. K., Pettigrew, S. M., 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.* **75**, e564–e571 (2022).
3. Crowcroft, N. S. & Klein, N. P. A framework for research on vaccine effectiveness. *Vaccine* **36**, 7286–7293 (2018).
4. Hungerford, D. et al. Mitigating bias in observational vaccine effectiveness studies using simulated comparator populations: Application to rotavirus vaccination in the UK. *Vaccine* **36**, 6674–6682 (2018).
5. Li, K. Q., Shi, X., Miao, W. & Tchetgen, E. T. Double negative control inference in test-negative design studies of vaccine effectiveness. *J. Am. Stat. Assoc.* **119**, 1859–1870 (2024).
6. Westreich, D. & Hudgens, M. G. Invited commentary: beware the test-negative design. *Am. J. Epidemiol.* **184**, 354–356 (2016).
7. Sullivan, S. G., Tchetgen Tchetgen, E. J. & Cowling, B. J. Theoretical basis of the test-negative study design for assessment of influenza vaccine effectiveness. *Am. J. Epidemiol.* **184**, 345–353 (2016).
8. Lipsitch, M., Jha, A. & Simonsen, L. Observational studies and the difficult quest for causality: lessons from vaccine effectiveness and impact studies. *Int. J. Epidemiol.* **45**, 2060–2074 (2016).
9. Wratil, P. R. et al. Vaccine-hesitant individuals accumulate additional COVID-19 risk due to divergent perception and behaviors related to SARS-CoV-2 testing: a population-based, cross-sectional study. *Infection* **51**, 909–919 (2023).
10. Glasziou, P. et al. Testing behaviour may bias observational studies of vaccine effectiveness. *J. Assoc. Med. Microbiol. Infect. Dis. Can.* **7**, 242–246 (2022).
11. Simpson, C. R., Ritchie, L. D., Robertson, C., Sheikh, A. & McMenamin, J. Effectiveness of H1N1 vaccine for the prevention of pandemic influenza in Scotland, UK: a retrospective observational cohort study. *Lancet Infect. Dis.* **12**, 696–702 (2012).
12. Huang, Z. et al. Effectiveness of inactivated COVID-19 vaccines among older adults in Shanghai: retrospective cohort study. *Nat. Commun.* **14**, 2009 (2023).
13. Rennert, L., Ma, Z., McMahan, C. S. & Dean, D. Covid-19 vaccine effectiveness against general SARS-CoV-2 infection from the omicron variant: A retrospective cohort study. *PLoS Glob. Public Health* **3**, e0001111 (2023).
14. Ono, S. et al. Comparative effectiveness of BNT162b2 and mRNA-1273 booster dose after BNT162b2 primary vaccination against the Omicron variants: a retrospective cohort study using large-scale population-based registries in Japan. *Clin. Infect. Dis.* **76**, 18–24 (2023).
15. Wolff Sagy, Y. et al. Real-world effectiveness of a single dose of mpox vaccine in males. *Nat. Med.* **29**, 748–752 (2023).
16. Cerqueira-Silva, T. et al. Duration of protection of CoronaVac plus heterologous BNT162b2 booster in the Omicron period in Brazil. *Nat. Commun.* **13**, 4154 (2022).
17. Tseng, H. F. et al. Effectiveness of mRNA-1273 vaccination against SARS-CoV-2 omicron subvariants BA.1, BA.2, BA.2.12.1, BA.4, and BA.5. *Nat. Commun.* **14**, 189 (2023).
18. van Doorn, E. et al. Influenza vaccine effectiveness estimates in the Dutch population from 2003 to 2014: the test-negative design case-control study with different control groups. *Vaccine* **35**, 2831–2839 (2017).
19. Buchan, S. A. et al. Estimated effectiveness of COVID-19 vaccines against Omicron or Delta symptomatic infection and severe outcomes. *JAMA Netw. Open* **5**, e2232760 (2022).
20. Jackson, M. L. et al. Influenza vaccine effectiveness in the United States during the 2015–2016 season. *N. Engl. J. Med.* **377**, 534–543 (2017).
21. Ostropolets, A. & Hripcsak, G. COVID-19 vaccination effectiveness rates by week and sources of bias: a retrospective cohort study. *BMJ Open* **12**, e061126 (2022).
22. Ioannidis, J. P. A. Factors influencing estimated effectiveness of COVID-19 vaccines in non-randomised studies. *BMJ Evid. Based Med.* **27**, 324–329 (2022).
23. Fukushima, W. & Hirota, Y. Basic principles of test-negative design in evaluating influenza vaccine effectiveness. *Vaccine* **35**, 4796–4800 (2017).
24. Jackson, M. L. & Nelson, J. C. The test-negative design for estimating influenza vaccine effectiveness. *Vaccine* **31**, 2165–2168 (2013).
25. Lewnard, J. A., Tedijanto, C., Cowling, B. J. & Lipsitch, M. Measurement of vaccine direct effects under the test-negative design. *Am. J. Epidemiol.* **187**, 2686–2697 (2018).
26. Andrews, N. et al. Covid-19 vaccine effectiveness against the Omicron (B.1.1.529) variant. *N. Engl. J. Med.* **386**, 1532–1546 (2022).
27. Chemaitelly, H. et al. Duration of mRNA vaccine protection against SARS-CoV-2 Omicron BA.1 and BA.2 subvariants in Qatar. *Nat. Commun.* **13**, 3082 (2022).
28. Suah, J. L., Bar-Zeev, N. & Knoll, M. D. How important are study designs? A simulation assessment of vaccine effectiveness estimation bias with time-varying vaccine coverage, and heterogeneous testing and baseline attack rates. Preprint at *medRxiv* https://doi.org/10.1101/2022.08.25.22279235 (2022)

29. Campbell, F. et al. Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021. *Eurosurveillance* **26**, 2100509 (2021).

30. Jalali, N. et al. Increased household transmission and immune escape of the SARS-CoV-2 Omicron compared to Delta variants. *Nat. Commun.* **13**, 5706 (2022).

31. Halloran, M. E., Haber, M. & Longini, I. M. Interpretation and esti-mation of vaccine efficacy under heterogeneity. *Am. J. Epidemiol.* **136**, 328–343 (1992).

32. Markov, P. V. et al. The evolution of SARS-CoV-2. *Nat. Rev. Microbiol.* **21**, 361–379 (2023).

33. Wu, J., Scarabel, F., McCarthy, Z., Xiao, Y. & Ogden, N. H. A window of opportunity for intensifying testing and tracing efforts to prevent new COVID-19 outbreaks due to more transmissible variants. *Can. Commun. Rep.* **47**, 329–338 (2021).

34. Gabriele-Rivet, V. et al. Modelling the impact of age-stratified public health measures on SARS-CoV-2 transmission in Canada. *R. Soc. Open Sci.* **8**, 210834 (2021).

35. Cui, Q. et al. Dynamic variations in COVID-19 with the SARS-CoV-2 Omicron variant in Kazakhstan and Pakistan. *Infect. Dis. Poverty* **12**, 18 (2023).

36. Yuan, P. et al. Assessing the mechanism of citywide test-trace-isolate Zero-COVID policy and exit strategy of COVID-19 pandemic. *Infect. Dis. Poverty* **11**, 104 (2022).

37. Bodner, K., Knight, J., Hamilton, M. A. & Mishra, S. Testing whether higher contact among the vaccinated can be a mechanism for observed negative vaccine effectiveness. *Am. J. Epidemiol.* **192**, 1335–1340 (2023).

38. Lyngse, F. P. et al. Household transmission of the SARS-CoV-2 Omicron variant in Denmark. *Nat. Commun.* **13**, 5573 (2022).

39. Park, A. W. et al. Quantifying the impact of immune escape on transmission dynamics of influenza. *Science* **326**, 726–728 (2009).

40. Moriyama, M., Hugentobler, W. J. & Iwasaki, A. Seasonality of respiratory viral infections. *Annu. Rev. Virol.* **7**, 83–101 (2020).

41. Townsend, J. P. et al. Seasonality of endemic COVID-19. *mBio* **14**, e01426-23 (2023).

42. Koltai, M. et al. Determinants of RSV epidemiology following sup-pression through pandemic contact restrictions. *Epidemics* **40**, 100614 (2022).

43. Gail, M. H. et al. Design choices for observational studies of the effect of exposure on disease incidence. *BMJ Open* **9**, e031031 (2019).

44. Lewnard, J. A. et al. Theoretical framework for retrospective studies of the effectiveness of SARS-CoV-2 vaccines. *Epidemiol. Camb. Mass* **32**, 508–517 (2021).

45. Lash, T. L. et al. Good practices for quantitative bias analysis. *Int. J. Epidemiol.* **43**, 1969–1985 (2014).

46. Bodner, K., Irvine, M. A., Kwong, J. C. & Mishra, S. Observed nega-tive vaccine effectiveness could be the canary in the coal mine for biases in observational COVID-19 studies. *Int. J. Infect. Dis.* **131**, 111–114 (2023).

47. Freedman, A. S. et al. Inferring COVID-19 testing and vaccination behavior from New Jersey testing data. *Proc. Natl Acad. Sci. USA* **121**, e2314357121 (2024).

48. Nordström, P., Ballin, M. & Nordström, A. Risk of SARS-CoV-2 rein-fection and COVID-19 hospitalisation in individuals with natural and hybrid immunity: a retrospective, total population cohort study in Sweden. *Lancet Infect. Dis.* **22**, 781–790 (2022).

49. Pilz, S., Theiler-Schwetz, V., Trummer, C., Krause, R. & Ioannidis, J. P. A. SARS-CoV-2 reinfections: overview of efficacy and duration of natural and hybrid immunity. *Environ. Res.* **209**, 112911 (2022).

50. Tsang, T. K. et al. Prior infections and effectiveness of SARS-CoV-2 vaccine in test-negative studies: a systematic review and meta-analysis. *Am. J. Epidemiol.* https://doi.org/10.1093/aje/kwae142 (2024).

51. Al Kaabi, N. et al. Effect of 2 inactivated SARS-CoV-2 vaccines on symptomatic COVID-19 infection in adults: a randomized clinical trial. *JAMA* **326**, 35–45 (2021).

52. Zhu, X. et al. Dynamics of inflammatory responses after SARS-CoV-2 infection by vaccination status in the USA: a prospective cohort study. *Lancet Microbe* **4**, e692–e703 (2023).

53. Mohr, N. M. et al. Presence of symptoms 6 weeks after COVID-19 among vaccinated and unvaccinated US healthcare personnel: a prospective cohort study. *BMJ Open* **13**, e063141 (2023).

54. Kuitunen, I., Uimonen, M., Seppälä, S. J. & Ponkilainen, V. T. COVID-19 vaccination status and testing rates in Finland—a potential cause for bias in observational vaccine effectiveness analysis. *Influenza Other Respir. Viruses* **16**, 842–845 (2022).

55. Shim, E. & Galvani, A. P. Distinguishing vaccine efficacy and effec-tiveness. *Vaccine* **30**, 6700–6705 (2012).

56. Kahn, R., Schrag, S. J., Verani, J. R. & Lipsitch, M. Identifying and alleviating bias due to differential depletion of susceptible people in postmarketing evaluations of COVID-19 vaccines. *Am. J. Epide-miol.* **191**, 800–811 (2022).

57. Public Health Agency of Canada. COVID-19 signs, symptoms and severity of disease: a clinician guide. https://www.canada.ca/en/public-health/services/diseases/2019-novel-coronavirus-infection/guidance-documents/signs-symptoms-severity.html (2022).

58. Venkatramanan, S. et al. Using data-driven agent-based models for forecasting emerging infectious diseases. *Epidemics* **22**, 43–49 (2018).

59. Erdös, P. & Rényi, A. On random graphs I. *Publ. Math. Debr.* **6**, 290–297 (1959).

60. Liu, C. et al. Rapid review of social contact patterns during the COVID-19 pandemic. *Epidemiol. Camb. Mass* **32**, 781–791 (2021).

61. Durrett, R. *Random Graph Dynamics* (Cambridge Univ. Press, Cambridge, U.K., 2007).

62. National Health Interview Survey. QuickStats: percentage of per-sons who had a cold in the past 2 weeks, by age group and calendar quarter — National Health Interview Survey, United States. *MMWR Morb. Mortal. Wkly. Rep.* **69**, 429 (2020).

63. Centers for Disease Control and Prevention. Suffering from a cold? *Centers for Disease Control and Prevention - Treatment for Common Illnesses* https://www.cdc.gov/antibiotic-use/colds.html (2023).

64. HealthLink BC. Facts about influenza (the flu). https://www.healthlinkbc.ca/healthlinkbc-files/facts-about-influenza-flu (2022).

65. Jeter, L. COVID-19 vaccines rollout nationwide: December 2021. *National Association of Attorneys General* https://www.naag.org/attorney-general-journal/current-status-of-covid-19-vaccines-rollout-nationwide-december-2021-update-for-the-attorney-general-community/ (2021).

66. Zeng, B., Gao, L., Zhou, Q., Yu, K. & Sun, F. Effectiveness of COVID-19 vaccines against SARS-CoV-2 variants of concern: a systematic review and meta-analysis. *BMC Med.* **20**, 200 (2022).

67. Menegale, F. et al. Evaluation of waning of SARS-CoV-2 vaccine–induced immunity: a systematic review and meta-analysis. *JAMA Netw. Open* **6**, e2310650 (2023).

68. Braeye, T. et al. Vaccine effectiveness against transmission of alpha, delta and omicron SARS-COV-2-infection, Belgian contact tracing, 2021–2022. *Vaccine* **41**, 3292–3300 (2023).

69. Endo, A., Funk, S. & Kucharski, A. J. Bias correction methods for test-negative designs in the presence of misclassification. *Epidemiol. Infect.* **148**, e216 (2020).

70. R Core Team. *R: A Language and Environment for Statistical Com-puting* (R Foundation for Statistical Computing, 2023).

71. Loken, C. et al. SciNet: lessons learned from building a power-efficient top-20 system and data centre. *J. Phys. Conf. Ser.* **256**, 012026 (2010).

72. Ponce, M. et al. Deploying a top-100 supercomputer for large parallel workloads: the Niagara Supercomputer. In *Proc. Practice and Experience in Advanced Research Computing on Rise of the Machines (Learning)* 1–8 (Association for Computing Machinery, New York, NY, USA, 2019).

73. Bodner, K. et al. Code and simulated data for "Impact of unequal testing on vaccine effectiveness estimates across two study designs: a simulation study". *GitHub* https://github.com/mishra-lab/testing_bias/releases/tag/v1.0.0 (2025).

74. Guo, Z. et al. Comparing the incubation period, serial interval, and infectiousness profile between SARS-CoV-2 Omicron and Delta variants. *J. Med. Virol.* **95**, e28648 (2023).

75. Li, R. et al. Clinical characteristics and risk factors analysis of viral shedding time in mildly symptomatic and asymptomatic patients with SARS-CoV-2 Omicron variant infection in Shanghai. *Front. Public Health* **10**, 1073387 (2022).

76. Sah, P. et al. Asymptomatic SARS-CoV-2 infection: a systematic review and meta-analysis. *Proc. Natl Acad. Sci. USA* **118**, e2109229118 (2021).

Author contributions

K.B., L.W. and S.M. contributed to the conceptual development of the project. K.B. contributed to data simulation, analysis and figure creation. K.B., L.W., R.K., J.C.K., B.S., H.S., M.A.I. and S.M. contributed to the description and interpretation of results. All authors contributed to manuscript editing and have approved the final submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-59768-1.

Correspondence and requests for materials should be addressed to Korryn Bodner or Sharmistha Mishra.

Peer review information *Nature Communications* thanks Neil Pearce, Chad Wells and the other anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

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