American Journal of Epidemiology Submitted Manuscript
Special Collection: N/A

Title: Estimating protection afforded by prior infection in preventing reinfection: Applying the test-negative study design

Authors: Houssein H. Ayoub,[1†*] Milan Tomy,[1,2,3†] Hiam Chemaitelly,[2,3,4] Heba N. Altarawneh,[2,3,4] Peter Coyle,[5,6,7] Patrick Tang,[8] Mohammad R. Hasan,[8] Zaina Al Kanaani,[5] Einas Al Kuwari,[5] Adeel A. Butt,[4,5] Andrew Jeremijenko,[5] Anvar Hassan Kaleeckal,[5] Ali Nizar Latif,[5] Riyazuddin Mohammad Shaik,[5] Gheyath K. Nasrallah,[6,9] Fatiha M. Benslimane,[6,9] Hebah A. Al Khatib,[6,9] Hadi M. Yassine,[6,9] Mohamed G. Al Kuwari,[10] Hamad Eid Al Romaihi,[11] Hanan F. Abdul-Rahim,[12] Mohamed H. Al-Thani,[11] Abdullatif Al Khal,[5] Roberto Bertollini,[11] and Laith J. Abu-Raddad[2,3,4,12*]

Correspondence Address:

Dr. Houssein H. Ayoub, Mathematics Program, Department of Mathematics, Statistics, and Physics, College of Arts and Sciences, Qatar University, P.O. Box 2713, Doha, Qatar. Telephone: +(974) 4403-7543. E-mail: hayoub@qu.edu.qa.

Professor Laith J. Abu-Raddad, Infectious Disease Epidemiology Group, World Health Organization Collaborating Centre for Disease Epidemiology Analytics on HIV/AIDS, Sexually Transmitted Infections, and Viral Hepatitis, Weill Cornell Medicine - Qatar, Qatar Foundation - Education City, P.O. Box 24144, Doha, Qatar. Telephone: +(974) 4492-8321. Fax: +(974) 4492-8333. E-mail: lja2002@qatar-med.cornell.edu

Affiliations:

[1]Mathematics Program, Department of Mathematics, Statistics, and Physics, College of Arts and Sciences, Qatar University, Doha, Qatar

[2]Infectious Disease Epidemiology Group, Weill Cornell Medicine-Qatar, Cornell University, Doha, Qatar

[3]World Health Organization Collaborating Centre for Disease Epidemiology Analytics on HIV/AIDS, Sexually Transmitted Infections, and Viral Hepatitis, Weill Cornell Medicine–Qatar, Cornell University, Qatar Foundation – Education City, Doha, Qatar

[4]Department of Population Health Sciences, Weill Cornell Medicine, Cornell University, New York, New York, USA

[5]Hamad Medical Corporation, Doha, Qatar

[6]Biomedical Research Center, Member of QU Health, Qatar University, Doha, Qatar

[7]Wellcome-Wolfson Institute for Experimental Medicine, Queens University, Belfast, United Kingdom

1

⁸Department of Pathology, Sidra Medicine, Doha, Qatar
⁹Department of Biomedical Science, College of Health Sciences, Member of QU Health, Qatar University, Doha, Qatar
¹⁰Primary Health Care Corporation, Doha, Qatar
¹¹Ministry of Public Health, Doha, Qatar
¹²Department of Public Health, College of Health Sciences, Member of QU Health, Qatar University, Doha, Qatar

[†]These authors contributed equally

2

Disclaimer: NA

Abstract

The COVID-19 pandemic has highlighted the need to use infection testing databases to rapidly estimate effectiveness of prior infection in preventing reinfection (PE_S) by novel SARS-CoV-2 variants. Mathematical modeling was used to demonstrate a theoretical foundation for applicability of the test-negative, case-control study design to derive PE_S. Apart from the very early phase of an epidemic, the difference between the test-negative estimate for PE_S and true value of PE_S was minimal and became negligible as the epidemic progressed. The test-negative design provided robust estimation of PE_S and its waning. Assuming that only 25% of prior infections are documented, misclassification of prior infection status underestimated PE_S, but the underestimate was considerable only when >50% of the population was ever infected. Misclassification of latent infection, misclassification of current active infection, and scale-up of vaccination all resulted in negligible bias in estimated PE_S. The test-negative design was applied to national-level testing data in Qatar to estimate PE_S for SARS-CoV-2. PE_S against SARS-CoV-2 Alpha and Beta variants was estimated at 97.0% (95% CI: 93.6-98.6) and 85.5% (95% CI: 82.4-88.1), respectively. These estimates were validated using a cohort study design.

3

The test-negative design offers a feasible, robust method to estimate protection from prior infection in preventing reinfection.

Introduction

Estimating effectiveness of prior infection in preventing reinfection (PE_S) is essential to understand the epidemiology of a given infection. Various studies estimated PE_S for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants [1-9]. However, there are challenges in estimating PE_S using conventional epidemiologic study designs. Such designs require extensive, complete electronic health records to be feasible. Vaccination scale-up makes it difficult to disentangle immunity induced by prior infection from that induced by vaccination. Even when it is feasible to apply conventional designs, estimates can be prone to strong bias, due to misclassification of prior infection status, since many prior infections are not documented [10-12]. Effects of this bias increase with increased cumulative infection exposure in the population [13]. Emergence of the Omicron [14] (B.1.1.529) variant and its subsequent subvariants emphasized the need to estimate PE_S rapidly once a new variant/subvariant emerges.

Here, we demonstrate a robust, practical method to estimate PE_S using a test-negative, case-control study design. This is, to our knowledge, the first use of this method to estimate PE_S. While it has been used to study vaccine effectiveness [15-22], it does not appear to have been used to estimate PE_S, perhaps because of a perception that it is not applicable, as most prior and current infections are undocumented, unlike vaccinations, which are typically documented and

4

tracked in health systems. We also provide an application of this method by estimating PE_S for

SARS-CoV-2 infection in Qatar, at a time when the Alpha [14] (B.1.1.7) and Beta [14] (B.1.351)

variants dominated incidence [21-26].

This article includes two components. The first is a parsimonious mathematical modeling

component whose purpose is to motivate the test-negative design and to demonstrate that

theoretically it can be applied to provide credible estimates for PE_S despite specific sources of

bias. This modeling exercise is not intended to provide a simulation of a specific empirical study

or discuss all sources of potential bias, but to provide a theoretical foundation of the applicability

of such design to estimate PE_S. The second component is a real-world application of the test-

negative design to actual routine data to generate estimates for PE_S. This specific application

was conducted because there are already published estimates for PE_S using a cohort study

design applied to the same data, population, and duration of study [4]. Therefore, the cohort

study design provides a validation for the test-negative design, as both the cohort and test-

negative designs yielded the same results when applied to the same data source.

Methods

Test-negative case-control study design

The test-negative, case-control study design has emerged as a robust and practical method to

assess vaccine effectiveness for respiratory tract infections [15-22, 27-32]. In this design, which

resembles a case-control design though it is not strictly a case-control design, persons seeking

healthcare because of symptoms are recruited into the study [15, 16, 27, 28, 30-32]. Those

testing positive for the infection (cases) are then matched to those testing negative (controls) [15,

5

16, 27, 28, 30-32]. Matching is done to control for differences in the risk of exposure to the infection [21, 22, 33]. Vaccine effectiveness is then derived as one minus the ratio of the odds of vaccination in subjects testing positive to the odds of vaccination in subjects testing negative [15, 16]. A key strength of this design is removal of differences in healthcare-seeking behavior between vaccinated and unvaccinated persons, thereby minimizing bias [15, 16, 27-32]. Another strength is minimization of bias arising from misclassification of infection [15, 16, 27-32].

Mathematical modeling and simulation of the test-negative design

Mathematical modeling was used to demonstrate a theoretical foundation for the applicability of the test-negative, case-control study design for deriving effectiveness of prior infection in preventing reinfection (PE_S), that is, the proportional reduction in susceptibility to infection among those with prior infection versus those without [2]. Modeling was also used to investigate effects of biases on estimated PE_S. While this demonstration was done for SARS-CoV-2 infection, the approach is generic and should be broadly applicable to a range of infections. Moreover, while this demonstration was done for any SARS-CoV-2 infection, regardless of symptoms, the same approach can be applied to other outcomes such as symptomatic infection, asymptomatic infection, severe or critical COVID-19 [34], or COVID-19 death [35], as long as these outcomes are defined as specific subsets of the broad any-infection outcome or its direct disease progression.

Several models were devised to simulate SARS-CoV-2 infection transmission in the population and to investigate applicability of the test-negative design. The models were based on previously published models and their parameters for SARS-CoV-2 infection [12, 36-42]. To keep only the

6

essential details for the investigations of this study, the models were parsimonious and not structured by age nor by infection type and severity. The instantaneous prevalence at each time point, for each population compartment, was used in the analyses of these models.

The first model was the classic Susceptible Exposed Infectious Recovered (SEIR) model, but extended to allow for reinfections (Baseline Model; Figure 1A). This model was used to demonstrate applicability of the test-negative design and to investigate sources of bias. In this model and its analysis all controls were either susceptible or recovered individuals, and all cases were either infected or reinfected individuals.

Building on previous modeling studies of vaccine effectiveness and its waning [13, 43-47], the second model was an extension of the Baseline Model to incorporate scale-up of vaccination in the population (Vaccination Model; Figure 1B). This model was used to investigate whether vaccination could affect applicability of this method to estimate PE_S. Vaccine effectiveness (VE_S) was defined as the proportional reduction in susceptibility to infection among those vaccinated versus those unvaccinated [40, 41]. VE_S was set at 75%, a representative value for the range of Coronavirus Disease 2019 (COVID-19) vaccines available during times in which incidence was due to pre-Omicron variants [21, 33, 48, 49]. Duration of vaccine-induced protection was assumed to be 6 months in light of documented waning of COVID-19 vaccine protection [25, 48-52].

The third model was also an extension of the Baseline Model, incorporating gradual (linear) waning in protection offered by prior infection against reinfection (Waning of Immunity Model; Figure 1C). Time after recovery from infection was modeled as an aging process whereby the

recovered population transitions from one population compartment to the next with the average duration spent in each compartment being one month. Each one-month recovered-population compartment had a set PE_S value. PE_S was modeled to decline linearly month by month. Accordingly, the recovered population is tracked month by month after recovery to allow for test-negative-study estimation of waning of natural immunity, as is done in the literature for waning of vaccine immunity after the second or booster doses [25, 52, 53].

These models consisted of coupled nonlinear differential equations that stratified the population into compartments (groups) based on infection status (infected, reinfected, or uninfected) and vaccination status (vaccinated, unvaccinated). Susceptible individuals (vaccinated or unvaccinated) were assumed at risk of acquiring the infection at a force of infection that varied throughout the epidemic due to variation in the contact rate. Recovered individuals (vaccinated or unvaccinated) were also assumed at risk of acquiring the infection, but the force of infection was reduced by the effect of PE_S.

These models were calibrated to mimic the actual evolution of the COVID-19 epidemic in Qatar [12, 36]. The contact rate was varied to generate two major epidemic waves several months apart, as actually occurred [12, 25, 36, 54]. Parameters of the models are summarized in Table 1. Further details on these models, their equations, and their parametrization can be found in previous publications [12, 36-42]. Modeling analyses were conducted in MATLAB R2019a (Boston/MA/USA) [55].

Effectiveness of prior infection against reinfection and impact of bias

8

Applying the test-negative, case-control study design, PE_S was derived as one minus the ratio of the odds of prior infection in subjects testing positive (such as by polymerase chain reaction (PCR) testing), to the odds of prior infection in subjects testing negative for the infection. The 2-by-2 table used to derive the odds ratio is shown in Figure 2A, as expressed in terms of the Baseline Model's population variables. The mathematical expression for PE_S is also shown in Figure 2A, assuming no form of bias. An underlying assumption is that those being tested are a specific fixed proportion (random sample) of all population variables; that is, the same sampling proportion is applied for each population compartment in the model. We also assumed that those latently infected (E compartment) are diagnosable as those in acute infection (I compartment), given the wide application of the highly sensitive PCR testing for SARS-CoV-2 infection, and because of existence of large-scale routine testing in many countries, in addition to testing for symptomatic cases. A departure of the latter assumption has been investigated in a sensitivity analysis.

Several forms of bias may affect estimation of PE_S using the test-negative method. The most critical is misclassification of prior infection status. A proportion g_P of those previously infected may not have been diagnosed and may have been unaware of their infections. It is reasonable to assume that most persons with a prior infection may not have had it documented [10-12]. Here, we assumed that 75% of prior infections are undocumented, that is an ascertainment rate of only 25% (Table 1). This ascertainment rate was based on fitting epidemic models to national seroprevalence survey data in Qatar [12, 38, 56-59], and is consistent with the ascertainment rate estimated for the United States using also serological surveys [10].

9

Unlike vaccine effectiveness studies, in which records are typically available to track vaccinations [15-22, 33], most persons with prior infection could be misclassified as persons with no prior infection. Similarly, most currently active infections may not be documented. The 2-by-2 table is thus modified for this bias along with the expression for PE_S (Figure 2B). It was assumed that this bias affects both cases and controls similarly, a valid assumption considering that both cases and controls are seeking healthcare because of symptoms. This assumption is central to the test-negative design strategy [15, 16, 27, 28, 30-32].

A second source of bias is misclassification of latent infection status. A proportion g_E of those with latent infections are asymptomatic; thereby remaining untested and undiagnosed. These cases would be misclassified as controls. The 2-by-2 table is thus modified to accommodate this bias along with the expression for PE_S (Figure 2C). We assumed that $g_E = 75\%$ (Table 1). We also assumed that this bias similarly affects those with and without prior infection. This is a valid assumption considering that both are seeking healthcare for the same reason, another assumption central to the test-negative design strategy [15, 16, 27, 28, 30-32].

A proportion g_I of cases (current active infections) could be misclassified as controls, because of lack of testing or due to imperfect sensitivity of the testing method, thereby introducing bias. The 2-by-2 table is thus modified for this bias along with the expression for PE_S (Figure 2D). We assumed that $g_I = 75\%$ (Table 1). We also assumed that this bias similarly affects those with and without prior infection [15, 16, 27, 28, 30-32].

10

Estimation of PE_S may occur at a time when vaccination is being scaled up, as in the current COVID-19 pandemic. This could introduce bias as vaccination is another form of immune protection. Using the Vaccination Model, the 2-by-2 table is modified in presence of vaccination along with the expression for PE_S (Figure 2E). We assumed that vaccination is being linearly scaled up to reach the vaccine coverage attained in Qatar during the duration of the simulation. We also assumed that protection of natural immunity and of vaccine immunity act independently of each other, as suggested recently for the effect of hybrid immunity [53]. Accordingly, protection of hybrid immunity of prior infection (PE_S) and vaccination (VE_S) combines as a multiplicative protection effect [53]—hybrid immunity of prior infection and vaccination is superior to that of either prior infection or vaccination separately [53, 54, 60].

Since different forms of bias may act synergistically when present together, the impact of the above biases was also investigated by applying all of them together at the same time.

Sensitivity analyses

Four sensitivity analyses were conducted. In the first sensitivity analysis, presented analyses were repeated using the real-world detailed reference mathematical model that was used to describe the epidemic and forecast its progression in Qatar, to inform policy decision-making (the Qatar Model; Figure S1) [12, 36, 38]. This model stratified the population into compartments according to age group, infection status (uninfected, infected, reinfected), infection type (asymptomatic/mild, severe, and critical), COVID-19 disease type (severe or critical disease), and vaccination status (vaccinated, unvaccinated). The model was fitted to the national standardized, integrated, and centralized databases of SARS-CoV-2 diagnosed cases,

SARS-CoV-2 PCR and antibody testing, COVID-19 hospitalizations, and COVID-19 mortality [12], as well as to data of a series of SARS-CoV-2 epidemiological studies in Qatar [1-3, 38, 42, 57-59]. The model-fitted indicators and the measured indicators and their comparison have been published previously, as well as an array of model projections for different infection and disease outcomes [12, 36-41, 61]. Model fitting was used to estimate key epidemiologic indicators including the ascertainment rates among others. A detailed description of the model, its input data, and fitting are available in References [12, 36, 38].

The second sensitivity analysis investigated the representativeness of PE_S as derived using the test-negative study design of the true PE_S, over the full spectrum of possible PE_S values. The third sensitivity analysis investigated whether the PE_S estimate can vary by using incidence instead of instantaneous prevalence in deriving the estimate. The fourth sensitivity analysis investigated the impact on PE_S of full misclassification bias of those latently infected. That is, none of those latently infected are being diagnosed; only those in acute infection are being diagnosed.

Real-world application: Effectiveness of prior infection in preventing reinfection in Qatar

To validate the test-negative design, PE_S was estimated in Qatar using national-level routine PCR testing data. Databases include all SARS-CoV-2-related data, with no missing information since pandemic onset, such as PCR tests and vaccinations. Only persons being PCR tested for clinical suspicion of infection due to symptoms between March 8 and April 21, 2021, were eligible for inclusion in this analysis. This study duration was chosen because there are existing

12

estimates for PE_S during this time, but using a conventional, cohort study design [4]. This allows validation of the estimate generated using the test-negative design.

Prior infection was defined as a PCR-confirmed infection ≥90 days before a new PCR-positive test [2, 6]. Individuals infected during the 90 days preceding the PCR test were thus excluded. Based on existing evidence [62-64] and viral genome sequencing [3, 21], a SARS-CoV-2 Alpha variant case was defined as an *S*-gene "target failure" case using the TaqPath COVID-19 Combo Kit (Thermo Fisher Scientific, USA [65]) applying the criterion of a PCR cycle threshold (Ct) value ≤30 for both the *N* and *ORF1ab* genes, but a negative outcome for the *S* gene [3, 4, 64]. With essentially only Beta and Alpha cases identified between March 8 and April 21, 2021 [21-26], a Beta case was proxied as the complement of the Alpha criterion, that is, any case with a Ct value ≤30 for the *N*, *ORF1ab*, and *S* genes.

Only the first PCR-positive test during the study was included for each case, and only the first PCR-negative test during the study was included for each control, per established protocol for the test-negative design [21, 22, 25, 33]. No Beta-positive cases were included as Alpha-negative controls, nor vice versa. The negative controls in both the Alpha and Beta analyses were chosen from the same population of those who tested negative during the study. Alpha and Beta cases were exact-matched one-to-one to controls (PCR-negative persons) by sex, 10-year age group, nationality, and calendar week of PCR test. Matching of cases and controls was done to control for known differences in the risk of exposure to SARS-CoV-2 infection in Qatar [12, 42, 57-59]. This applied test-negative design, including these specific inclusion and exclusion criteria, was developed over a series of studies [17, 21, 22, 25, 52, 66] to minimize effects of potential bias, such as retesting after a positive test to check for clearance of infection, or to control the effect of

13

repeat testers [25]. Extensive sensitivity and additional analyses were conducted in these prior

studies to investigate effects of different kinds of potential bias in this design, including

investigating different adjustments in the analysis, different approaches for matching [67],

different approaches for factoring prior infection in the analysis, in addition to other different

study inclusion and exclusion criteria [17, 21, 22, 25, 52, 66]. The applied test-negative design is

an outcome of these analyses to optimize the design by minimizing different sources of bias in

real-world data. The design was also validated using studies that used control groups to test for

null effects [22, 25, 52, 68, 69], and also validated using cohort study designs applied to the same

population and that yielded findings similar to those of the test-negative design [21, 22, 66].

Further description of Qatar's databases and methods of analysis can be found in previous

publications [1-4, 21, 22, 25, 33, 42, 54].

Socio-demographic characteristics of study samples were described using frequency distributions

and measures of central tendency. The odds ratio, comparing odds of prior infection among cases

versus controls, and its associated 95% confidence interval (CI) were derived using conditional

logistic regression, that is factoring matching in the study design. This analytical approach is

done to minimize potential bias due to variation in epidemic phase [15, 70] and other

confounders [12, 42, 57-59, 71, 72]. PE_S and its associated 95% CI were calculated by applying

the following equation:

$PE_S = 1 - \text{odds ratio of prior infection among cases versus controls}$

Statistical analyses were conducted in STATA/SE version 17.0 [73]. The study was approved by

the Hamad Medical Corporation and Weill Cornell Medicine-Qatar Institutional Review Boards

14

with waiver of informed consent. The study was reported following the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines. The STROBE checklist is found in Table S1.

Results

Protection of prior infection using the test-negative design and impact of bias

Figure 3 shows simulated evolution of the SARS-CoV-2 epidemic in its two waves (Figure 3A), the proportion of the population ever infected (Figure 3B), and vaccine coverage (Figure 3C). Figure 4A shows the estimated PE_S using the test-negative design (labeled as $PE_S^{test-negative}$), by application of the expression in Figure 2A, compared to the true PE_S (labeled as PE_S^{true}), here assumed at 80% (Table 1). Apart from the very early phase of the epidemic, when the number of reinfections was minimal, the difference between $PE_S^{test-negative}$ and PE_S^{true} was no more than several percentage points. The difference became negligible as the epidemic progressed.

Assuming that only 25% of prior infections are documented (Table 1), Figure 5A shows the impact of misclassification of prior infection, by application of the expression in Figure 2B. This form of bias resulted in underestimation of PE_S^{true}. When the proportion of the population ever infected was below 50% (Figure 3B), $PE_S^{test-negative}$ was only few percentage points lower than that of PE_S^{true}. However, the underestimation increased to as much as 30 percentage points when the proportion of the population ever infected was ~75%. Therefore, $PE_S^{test-negative}$ would provide only a lower bound for PE_S^{true} in situations where nearly everyone is infected, such as for influenza.

15

Misclassification of latent infection (Figure 5B), misclassification of current active infection (Figure 5C), and scale-up of vaccination (Figure 5D), all resulted in negligible bias in estimated $PE_S^{test-negative}$. Application of the above forms of bias at the same time suggested that there is no synergy when biases are combined (Figure S2).

Applying the Waning of Immunity Model, Figure 4B shows $PE_S^{test-negative}$ versus PE_S^{true}, month by month after prior infection, assuming that there is a gradual linear waning in protection of prior infection against reinfection. This comparison was done after the second wave at day 600 after the virus introduction (Figure 3A). $PE_S^{test-negative}$ provided a robust approximation of PE_S^{true} and its waning month by month.

Above analyses were repeated in the first sensitivity analysis that used the real-world Qatar Model. The analysis confirmed the same findings as those of the main analysis using the parsimonious models (Figure S3). Impact of bias due to scale-up of vaccination was not investigated using the Qatar Model, as this model's fitting already factors the scale-up of vaccination in Qatar [36].

The second sensitivity analysis showed that $PE_S^{test-negative}$ reflects the value of PE_S^{true} regardless of the actual value of PE_S^{true} and over the full spectrum of possible PE_S^{true} values (Figure S4). The third sensitivity analysis showed that $PE_S^{test-negative}$ estimate using incidence is similar to that using instantaneous prevalence (Figure S5). The fourth sensitivity analysis showed that full misclassification bias of those latently infected has virtually no impact on estimated $PE_S^{test-negative}$ (Figure S6).

16

Application: Effectiveness of prior infection in preventing reinfection in Qatar

Figure 6 presents a flowchart describing the population selection process for estimating PE_S in Qatar using the test-negative design. The median age of study subjects was 32-34 years, at least half were males, and they came from diverse countries (Table 2). Study samples were broadly representative of Qatar's demographics [42, 74].

Among the 4,645 Alpha cases (PCR-positive persons), 7 had a record of prior infection, compared to 232 among their matched controls (PCR-negative persons). PE_S against Alpha was estimated at 97.0% (95% CI: 93.6-98.6). Among the 13,753 Beta cases, 124 had a record of prior infection, compared to 815 among their matched controls. PE_S against Beta was estimated at 85.5% (95% CI: 82.4-88.1).

There was a total of 239 discordant pairs and 4,406 concordant pairs in the Alpha analysis and a total of 925 discordant pairs and 12,828 concordant pairs in the Beta analysis. The analyses were conducted on large samples of paired cases and controls and should not be affected by bias due to small samples or sparse data [75].

During the study duration (March 8, 2021 to April 21, 2021), we conducted earlier two matched cohort studies to estimate PE_S for Alpha and for Beta [4]. For Alpha, cohort-study estimates were 97.6% (95% CI: 95.7-98.7%) and 96.4% (95% CI: 92.1-98.3%) [4]. For Beta, cohort-study estimates were 92.3% (95% CI: 90.3-93.8%) and 86.4% (95% CI: 82.5-89.5%) [4].

Power analysis

17

The above application for Alpha and Beta protections demonstrates an actual empirical application, but the number of cases may not be sufficient in other applications to provide a precise and meaningful estimate for PE_S. Therefore, we conducted a power analysis to provide an estimate of the sample size necessary to apply this method using Power and Sample Size software version 3.1.2 [76] following Dupont principles [77].

Assuming the proportion of controls with prior infection at 25% and a high correlation between cases and controls of 0.5 [78], an estimated sample size of 71 individuals for each of cases and controls is needed to detect an odds ratio of 0.2, that is assuming PE_S of 80%, at two-sided type I error probability of 5% and power of 80%.

Assuming an attrition of 80% due to exclusion for study ineligibility and an additional attrition of 5% from loss to matching, as informed by the above applications for Alpha and Beta protections, the required sample size would be 374 for each of cases and controls. If PE_S was 50% instead, that is an odds ratio of 0.5, the required sample size would be 1,474 for each of cases and controls.

Discussion

Study results show that the test-negative design can be used to generate rigorous estimates for protection afforded by prior infection against reinfection, even though most prior infections are undocumented. Estimates were robust despite several forms of potential bias, and even under rather extreme assumptions for these biases. The test-negative design was also applied to Qatar's routine PCR testing data, and results were validated by comparing test-negative estimates to those generated using conventional cohort study designs [4]. Application of the test-negative

18

design should be feasible in different countries as long as there are databases for infection testing that are of reasonable quality and that can be linked to documented prior infection status (and preferably to vaccination status). Such databases are available and have been used extensively in vaccine effectiveness studies using the test-negative design, such as for SARS-CoV-2 infection [17-22, 33], and recently to estimate PE_S for the Omicron variant [79]. This is a key strength for test-negative studies in that such studies are typically implemented on full eligible routine datasets where the large sample sizes optimize the statistical precision of the estimates.

Of the considered biases, only misclassification of prior infection status could have a large effect on PE_S estimation, but mainly where more than 50% of the population already had a prior infection. This situation is not likely to have been reached for SARS-CoV-2 infection before the introduction the Omicron variant in most countries [56]. Even in such situations, the direction (and magnitude) of bias is known; it underestimates PE_S. Therefore, the test-negative design can still provide a lower bound for the true PE_S, which may be sufficient to inform public health decision making, such as in relation to differential application of restrictions by prior infection status, timing of vaccination following documented infection, and protocols for isolation and quarantine. Thus, this bias may not restrict the utility of this method.

The test-negative study design has strengths that conventional designs may lack. Cohort study designs can be affected by bias resulting from different infection testing frequencies in the two arms of the study. This bias does not affect the test-negative design, as it uses only those who are tested. An example can be seen in comparing the results of the test-negative design to the results of our earlier cohort design [4]. In the cohort design, adjustment for testing frequency reduced

19

PE_S from 97.6% (95% CI: 95.7-98.7%) to 95.8% (95% CI: 92.5-97.7%) for Alpha [4], very

similar to the test-negative estimate of 97.0% (95% CI: 93.6-98.6). Similarly for Beta,

adjustment for testing frequency reduced PE_S from 92.3% (95% CI: 90.3-93.8%) to 86.5%

(95% CI: 83.0-89.2%) [4], very similar to the test-negative estimate of 85.5% (95% CI: 82.4-

88.1). Accordingly, the test-negative design may provide a more representative estimate than the

cohort design.

The test-negative design may also be preferable to the cohort design for other reasons. Cohort

designs rely on cohorts that may not be strictly comparable, and it may not be possible to control

for all differences in risk of exposure to the infection by matching and analysis adjustments. For

example, in our earlier cohort study [4], we compared those with a record of a prior PCR-

confirmed infection to those with an antibody-negative test, but these two groups may differ in

ways that cannot be controlled. Meanwhile, the test-negative design is perhaps less susceptible to

such differences, as cases and controls are selected to meet certain clinical criteria that

presumably imply the same healthcare-seeking behavior. Having said so, use of administrative

databases may still be prone to bias due to unmeasured differences in healthcare-seeking

behavior. Lastly, while the test-negative design can be biased by misclassification of prior

infection, the cohort design is perhaps more affected by this bias. The odds ratio metric in the

test-negative design is less affected by this bias than the relative risk, incidence rate ratio, or

hazard ratio metrics in the cohort design.

In regard to limitations, we used a heuristic approach to motivate the test-negative design

through mathematical modeling, but this approach may not exactly match an actual empirical

test-negative-design application. The ultimate validity and utility of this design rests on actual

20

validation studies, including comparison to results of other conventional designs. We provided two such validation studies in the present article for each of the Alpha and Beta variants. Considering the demonstrated utility of this design in providing timely results in emergent situations during the COVID-19 pandemic [53, 79-81], this article should be seen as a call for further investigation and methodological development to enhance this design and its applications.

Specific forms of bias were investigated, but other sources of bias are possible, and these may also depend on the database being analyzed [25]. There is already a volume of literature investigating other forms of bias for the test-negative design in the context of vaccine effectiveness estimation [15, 16, 27-32], some of which may also apply in the context of PE_S estimation, such as for issues relating to testing and applicability of this design for different testing modalities [25]. Yet, more studies are needed to investigate different methodological aspects of this design and other sources of bias, such as the uncertainty/power to estimate effect and validity of the assumption of proportional random sampling of the different epidemiologic classes/compartments.

While this study demonstrated use of the test-negative design to estimate PE_S, other factors need to be considered in actual application. For instance, the algorithm for matching [67, 82] needs to be developed with knowledge of the local epidemiology to ensure that matching can effectively control differences in the risk of exposure to the infection. Of note that with Qatar's young population, the estimates presented here for PE_S may not be generalizable to other countries where elderly citizens constitute a larger proportion of the total population.

21

The models used to investigate applicability of the test-negative design were not structured by age, nor by infection type and severity. However, the sensitivity analysis that used the real-world Qatar Model, with its detailed stratifications, confirmed the same findings as those of the study's parsimonious models. Moreover, the three other sensitivity analyses confirmed the applicability of the test-negative design regardless of the value of PE_S^{true}, irrespective of whether incidence is used instead of instantaneous prevalence in the estimation, and whether or not there was full misclassification bias of those latently infected.

In conclusion, the test-negative design offers a feasible and robust method to estimate protection of prior infection in preventing reinfection. This method should be considered to provide rapid, rigorous estimates of protection offered by prior infection for different variants of SARS-CoV-2, including those that emerged recently.

References

1.   Abu-Raddad, L.J., et al., *Assessment of the Risk of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Reinfection in an Intense Reexposure Setting.* Clin Infect Dis, 2021. **73**(7): p. e1830-e1840.

2.   Abu-Raddad, L.J., et al., *SARS-CoV-2 antibody-positivity protects against reinfection for at least seven months with 95% efficacy.* EClinicalMedicine, 2021. **35**: p. 100861.

3.   Abu-Raddad, L.J., et al., *Introduction and expansion of the SARS-CoV-2 B.1.1.7 variant and reinfections in Qatar: A nationally representative cohort study.* PLoS Med, 2021. **18**(12): p. e1003879.

4.   Chemaitelly, H., et al., *Efficacy of Natural Immunity against SARS-CoV-2 Reinfection with the Beta Variant.* N Engl J Med, 2021.

5.   Hansen, C.H., et al., *Assessment of protection against reinfection with SARS-CoV-2 among 4 million PCR-tested individuals in Denmark in 2020: a population-level observational study.* Lancet, 2021. **397**(10280): p. 1204-1212.

6.   Kojima, N., N.K. Shrestha, and J.D. Klausner, *A Systematic Review of the Protective Effect of Prior SARS-CoV-2 Infection on Repeat Infection.* Eval Health Prof, 2021. **44**(4): p. 327-332.

7.   Leidi, A., et al., *Risk of reinfection after seroconversion to SARS-CoV-2: A population-based propensity-score matched cohort study.* Clin Infect Dis, 2021: p. 2021.03.19.21253889.

8.   Lumley, S.F., et al., *Antibody Status and Incidence of SARS-CoV-2 Infection in Health Care Workers.* N Engl J Med, 2021. **384**(6): p. 533-540.

9.   Pilz, S., et al., *SARS-CoV-2 re-infection risk in Austria.* Eur J Clin Invest, 2021. **51**(4): p. e13520.

10.   Angulo, F.J., L. Finelli, and D.L. Swerdlow, *Estimation of US SARS-CoV-2 Infections, Symptomatic Infections, Hospitalizations, and Deaths Using Seroprevalence Surveys.* JAMA Netw Open, 2021. **4**(1): p. e2033706.

11.   Jones, J.M., et al., *Estimated US Infection- and Vaccine-Induced SARS-CoV-2 Seroprevalence Based on Blood Donations, July 2020-May 2021.* JAMA, 2021. **326**(14): p. 1400-1409.

12.   Ayoub, H.H., et al., *Mathematical modeling of the SARS-CoV-2 epidemic in Qatar and its impact on the national response to COVID-19.* J Glob Health, 2021. **11**: p. 05005.

13.   Kahn, R., et al., *Identifying and Alleviating Bias Due to Differential Depletion of Susceptible People in Postmarketing Evaluations of COVID-19 Vaccines.* Am J Epidemiol, 2022. **191**(5): p. 800-811.

14.   World Health Organization, *Tracking SARS-CoV-2 variants. Available from: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/. Accessed on: June 5, 2021.* 2021.

15.   Jackson, M.L. and J.C. Nelson, *The test-negative design for estimating influenza vaccine effectiveness.* Vaccine, 2013. **31**(17): p. 2165-8.

16.   Verani, J.R., et al., *Case-control vaccine effectiveness studies: Preparation, design, and enrollment of cases and controls.* Vaccine, 2017. **35**(25): p. 3295-3302.

17.   Lopez Bernal, J., et al., *Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant.* N Engl J Med, 2021.

18.   Sheikh, A., et al., *SARS-CoV-2 Delta VOC in Scotland: demographics, risk of hospital admission, and vaccine effectiveness.* The Lancet, 2021. **397**(10293): p. 2461-2462.

19.   Nasreen, S., et al., *Effectiveness of COVID-19 vaccines against variants of concern, Canada.* medRxiv, 2021: p. 2021.06.28.21259420.

20. Dean, N.E., J.W. Hogan, and M.E. Schnitzer, *Covid-19 Vaccine Effectiveness and the Test-Negative Design.* N Engl J Med, 2021.

21. Abu-Raddad, L.J., et al., *Effectiveness of the BNT162b2 Covid-19 Vaccine against the B.1.1.7 and B.1.351 Variants.* N Engl J Med, 2021. **385**(2): p. 187-189.

22. Chemaitelly, H., et al., *mRNA-1273 COVID-19 vaccine effectiveness against the B.1.1.7 and B.1.351 variants and severe COVID-19 disease in Qatar.* Nat Med, 2021. **27**(9): p. 1614-1621.

23. National Project of Surveillance for Variants of Concern and Viral Genome Sequencing. *Qatar viral genome sequencing data. Data on randomly collected samples. https://www.gisaid.org/phylodynamics/global/nextstrain/.* 2021; Available from: https://www.gisaid.org/phylodynamics/global/nextstrain/.

24. Hasan, M.R., et al., *Real-Time SARS-CoV-2 Genotyping by High-Throughput Multiplex PCR Reveals the Epidemiology of the Variants of Concern in Qatar.* Int J Infect Dis, 2021. **112**: p. 52-54.

25. Chemaitelly, H., et al., *Waning of BNT162b2 Vaccine Protection against SARS-CoV-2 Infection in Qatar.* N Engl J Med, 2021. **385**(24): p. e83.

26. Benslimane, F.M., et al., *One Year of SARS-CoV-2: Genomic Characterization of COVID-19 Outbreak in Qatar.* Front Cell Infect Microbiol, 2021. **11**: p. 768883.

27. Foppa, I.M., et al., *The case test-negative design for studies of the effectiveness of influenza vaccine.* Vaccine, 2013. **31**(30): p. 3104-9.

28. Sullivan, S.G., S. Feng, and B.J. Cowling, *Potential of the test-negative design for measuring influenza vaccine effectiveness: a systematic review.* Expert Rev Vaccines, 2014. **13**(12): p. 1571-91.

29. Jackson, M.L. and K.J. Rothman, *Effects of imperfect test sensitivity and specificity on observational studies of influenza vaccine effectiveness.* Vaccine, 2015. **33**(11): p. 1313-6.

30. Lewnard, J.A., et al., *Measurement of Vaccine Direct Effects Under the Test-Negative Design.* Am J Epidemiol, 2018. **187**(12): p. 2686-2697.

31. Sullivan, S.G., E.J. Tchetgen Tchetgen, and B.J. Cowling, *Theoretical Basis of the Test-Negative Study Design for Assessment of Influenza Vaccine Effectiveness.* Am J Epidemiol, 2016. **184**(5): p. 345-53.

32. Lipsitch, M., A. Jha, and L. Simonsen, *Observational studies and the difficult quest for causality: lessons from vaccine effectiveness and impact studies.* Int J Epidemiol, 2016. **45**(6): p. 2060-2074.

33. Tang, P., et al., *BNT162b2 and mRNA-1273 COVID-19 vaccine effectiveness against the SARS-CoV-2 Delta variant in Qatar.* Nat Med, 2021. **27**(12): p. 2136-2143.

34. World Health Organization (WHO), *Living guidance for clinical management of COVID-19. Aavailable from: https://www.who.int/publications/i/item/WHO-2019-nCoV-clinical-2021-2. Accessed on: February 27, 2023.* 2021.

35. World Health Organization (WHO), *International Guidelines for Certification and Classification (Coding) of COVID-19 as Cause of Death. Available from: https://www.who.int/publications/m/item/international-guidelines-for-certification-and-classification-(coding)-of-covid-19-as-cause-of-death. Accessed on: February 27, 2023.* 2020.

36. Ayoub, H.H., et al., *Epidemiological impact of prioritising SARS-CoV-2 vaccination by antibody status: mathematical modelling analyses.* BMJ Innov, 2021. **7**(2): p. 327-336.

37. Ayoub, H.H., et al., *Characterizing key attributes of COVID-19 transmission dynamics in China's original outbreak: Model-based estimations.* Glob Epidemiol, 2020. **2**: p. 100042.

24

38. Seedat, S., et al., *SARS-CoV-2 infection hospitalization, severity, criticality, and fatality rates in Qatar.* Sci Rep, 2021. **11**(1): p. 18182.

39. Mumtaz, G.R., et al., *Modeling the Impact of COVID-19 Vaccination in Lebanon: A Call to Speed-Up Vaccine Roll Out.* Vaccines (Basel), 2021. **9**(7).

40. Makhoul, M., et al., *Epidemiological Impact of SARS-CoV-2 Vaccination: Mathematical Modeling Analyses.* Vaccines (Basel), 2020. **8**(4).

41. Makhoul, M., et al., *Epidemiological Differences in the Impact of COVID-19 Vaccination in the United States and China.* Vaccines (Basel), 2021. **9**(3): p. 2021.01.07.21249380.

42. Abu-Raddad, L.J., et al., *Characterizing the Qatar advanced-phase SARS-CoV-2 epidemic.* Scientific Reports, 2021. **11**(1): p. 6233.

43. Ray, G.T., et al., *Depletion-of-susceptibles Bias in Analyses of Intra-season Waning of Influenza Vaccine Effectiveness.* Clin Infect Dis, 2020. **70**(7): p. 1484-1486.

44. Halloran, M.E., M. Haber, and I.M. Longini Jr, *Interpretation and estimation of vaccine efficacy under heterogeneity.* American Journal of Epidemiology, 1992. **136**(3): p. 328-343.

45. Halloran, M.E., et al., *Direct and indirect effects in vaccine efficacy and effectiveness.* American journal of epidemiology, 1991. **133**(4): p. 323-331.

46. Yang, Y., et al., *The transmissibility and control of pandemic influenza A (H1N1) virus.* Science, 2009. **326**(5953): p. 729-733.

47. Halloran, M.E., et al., *Containing bioterrorist smallpox.* Science, 2002. **298**(5597): p. 1428-1432.

48. International Vaccine Access Center (IVAC), *VIEW-hub COVID-19 vaccines effectiveness data. Available from: https://view-hub.org/covid-19/effectiveness-studies. Accessed on May 5, 2022.*

49. *COVID-19 Vaccine Effectiveness On Variants Of Concern. Available from: https://covid-nma.com/vaccines/os_vaccines/. Accessed on May 5, 2022.* 2022.

50. Feikin, D.R., et al., *Duration of effectiveness of vaccines against SARS-CoV-2 infection and COVID-19 disease: results of a systematic review and meta-regression.* Lancet, 2022. **399**(10328): p. 924-944.

51. Chemaitelly, H. and L.J. Abu-Raddad, *Waning effectiveness of COVID-19 vaccines.* Lancet, 2022. **399**(10327): p. 771-773.

52. Abu-Raddad, L.J., et al., *Waning mRNA-1273 Vaccine Effectiveness against SARS-CoV-2 Infection in Qatar.* N Engl J Med, 2022.

53. Altarawneh, H.N., et al., *Effects of Previous Infection and Vaccination on Symptomatic Omicron Infections.* N Engl J Med, 2022. **387**(1): p. 21-34.

54. Abu-Raddad, L.J., et al., *Association of Prior SARS-CoV-2 Infection With Risk of Breakthrough Infection Following mRNA Vaccination in Qatar.* JAMA, 2021. **326**(19): p. 1930-1939.

55. MATLAB®, *The Language of Technical Computing. The MathWorks, Inc.* 2019.

56. Ayoub, H.H., et al., *Estimates of global SARS-CoV-2 infection exposure, infection morbidity, and infection mortality rates in 2020.* Glob Epidemiol, 2021. **3**: p. 100068.

57. Jeremijenko, A., et al., *Herd Immunity against Severe Acute Respiratory Syndrome Coronavirus 2 Infection in 10 Communities, Qatar.* Emerg Infect Dis, 2021. **27**(5): p. 1343-1352.

58. Al-Thani, M.H., et al., *SARS-CoV-2 Infection Is at Herd Immunity in the Majority Segment of the Population of Qatar.* Open Forum Infect Dis, 2021. **8**(8): p. ofab221.

59. Coyle, P.V., et al., *SARS-CoV-2 seroprevalence in the urban population of Qatar: An analysis of antibody testing on a sample of 112,941 individuals.* iScience, 2021. **24**(6): p. 102646.

60. Goldberg, Y., et al., *Protection and Waning of Natural and Hybrid Immunity to SARS-CoV-2.* N Engl J Med, 2022. **386**(23): p. 2201-2212.

61. Bsat, R., et al., *Characterizing the effective reproduction number during the COVID-19 pandemic: Insights from Qatar's experience.* J Glob Health, 2022. **12**: p. 05004.

62. European Centre for Disease Prevention and Control. *Rapid increase of a SARS-CoV-2 variant with multiple spike protein mutations observed in the United Kingdom. Available from: https://www.ecdc.europa.eu/sites/default/files/documents/SARS-CoV-2-variant-multiple-spike-protein-mutations-United-Kingdom.pdf. Accessed on: February 10, 2021.* 2020.

63. Galloway, S.E., et al., *Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29, 2020-January 12, 2021.* MMWR Morb Mortal Wkly Rep, 2021. **70**(3): p. 95-99.

64. Challen, R., et al., *Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study.* BMJ, 2021. **372**: p. n579.

65. Thermo Fisher Scientific, *TaqPath™ COVID-19 CE-IVD RT-PCR Kit instructions for use. Available from: https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0019215_TaqPathCOVID-19_CE-IVD_RT-PCR%20Kit_IFU.pdf. Accessed on December 02, 2020.* 2020.

66. Butt, A.A., et al., *SARS-CoV-2 vaccine effectiveness in preventing confirmed infection in pregnant women.* J Clin Invest, 2021. **131**(23).

67. Rose, S. and M.J. Laan, *Why match? Investigating matched case-control study designs with causal effect estimation.* Int J Biostat, 2009. **5**(1): p. Article 1.

68. Abu-Raddad, L.J., et al., *Pfizer-BioNTech mRNA BNT162b2 Covid-19 vaccine protection against variants of concern after one versus two doses.* J Travel Med, 2021. **28**(7).

69. Abu-Raddad, L.J., et al., *Effectiveness of mRNA-1273 and BNT162b2 Vaccines in Qatar.* N Engl J Med, 2022.

70. Jacoby, P. and H. Kelly, *Is it necessary to adjust for calendar time in a test negative design?: Responding to: Jackson ML, Nelson JC. The test negative design for estimating influenza vaccine effectiveness. Vaccine 2013;31(April (17)):2165-8.* Vaccine, 2014. **32**(25): p. 2942.

71. Pearce, N., *Analysis of matched case-control studies.* BMJ, 2016. **352**: p. i969.

72. Rothman, K.J., S. Greenland, and T.L. Lash, *Modern epidemiology.* 3rd ed. 2008, Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins. x, 758 p.

73. StataCorp, *Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC.* 2021.

74. Planning and Statistics Authority-State of Qatar, *Qatar Monthly Statistics. Available from: https://www.psa.gov.qa/en/pages/default.aspx. Accessed on: May 26, 2020.* 2020.

75. Greenland, S., J.A. Schwartzbaum, and W.D. Finkle, *Problems due to small samples and sparse data in conditional logistic regression analysis.* American journal of epidemiology, 2000. **151**(5): p. 531-539.

76. Dupont, W. and W. Plummer, *PS: Power and Sample Size Calculation 3.1.2. Available from: https://ps-power-and-sample-size-calculation.software.informer.com/3.1/ . Accessed on March 7, 2023.* 2023.

77. Dupont, W.D., *Power Calculations for Matched Case Control Studies.* Biometrics, 1988. **44**(4): p. 1157-1168.

78. Hennessy, S., et al., *Factors influencing the optimal control-to-case ratio in matched case-control studies.* Am J Epidemiol, 1999. **149**(2): p. 195-7.

79. Altarawneh, H.N., et al., *Protection against the Omicron Variant from Previous SARS-CoV-2 Infection.* N Engl J Med, 2022.

80. Altarawneh, H.N., et al., *Protective Effect of Previous SARS-CoV-2 Infection against Omicron BA.4 and BA.5 Subvariants.* N Engl J Med, 2022. **387**(17): p. 1620-1622.

81. Chemaitelly, H., et al., *Protection against Reinfection with the Omicron BA.2.75 Subvariant.* N Engl J Med, 2023. **388**(7): p. 665-667.

82. Abu-Raddad, L.J., et al., *COVID-19 risk score as a public health tool to guide targeted testing: A demonstration study in Qatar.* PLoS One, 2022. **17**(7): p. e0271324.

83. Li, R., et al., *Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2).* Science, 2020.

84. Lauer, S.A., et al., *The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application.* Ann Intern Med, 2020.

85. Zou, L., et al., *SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients.* N Engl J Med, 2020.

86. Rothe, C., et al., *Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany.* N Engl J Med, 2020. **382**(10): p. 970-971.

87. World Health Organization, *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). Available from :https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf. Accessed on March 10, 2020.* 2020.

88. United Nations Department of Economic and Social Affairs Population Dynamics, *The 2019 Revision of World Population Prospects. Available from https://population.un.org/wpp/. Accessed on March 1st, 2020.* 2020.

27

Table 1. Model parameters and assumptions.

Parameter	Symbol	Value	Justification
Duration of latent infection		3.69 days	Based on existing estimate [83] and based on a median incubation period of 5.1 days [84] adjusted by observed viral load among infected persons [85] and reported transmission before onset of symptoms [86]
Duration of infectiousness		3.48 days	Based on existing estimate [83] and based on observed time to recovery among persons with mild infection [83, 87] and observed viral load in infected persons [85, 86]
Infection fatality rate		1.85 per 10,0000 infections	Estimate based on fitting the epidemic in Qatar [38]
Life expectancy in Qatar		80.7 years	United Nations World Population Prospects database [88]
Vaccine effectiveness in reducing susceptibility to infection	VE_S	75%	Representative value for the range of COVID-19 vaccines available at present [21, 33, 48, 49]
Duration of vaccine protection		6 months	Based on evidence on waning of vaccine protection [25, 48-52]
Model-assumed "true" effectiveness of prior infection in preventing reinfection	PE_S^{true}	80%	Informed by evidence from existing studies [1-9]
Proportion of prior infections that are undocumented	g_p	75%	Informed by evidence from existing studies [10-12, 38]
Proportion of latent infections that are undocumented	g_E	75%	Informed by evidence from existing studies [10-12, 38]
Proportion of current active infections that are undocumented	g_I	75%	Informed by evidence from existing studies [10-12, 38]

28

Table 2. Demographic characteristics of subjects in the samples used to estimate effectiveness of prior infection in preventing reinfection using the test-negative study design.

Characteristics	Cases[a] (PCR-confirmed infection with the Alpha variant) N=4,645		Controls[a] (PCR-negative) N=4,645		SMD[b]	Cases[a] (PCR-confirmed infection with the Beta variant) N=13,753		Controls[a] (PCR-negative) N=13,753		SMD[b]
Median age (IQR) — years	33 (25-40)[c]		32 (24-40)[c]		0.01[d]	34 (27-40)[c]		33 (27-40)[c]		0.01[d]
	No.	%	No.	%		No.	%	No.	%	
Age group					0.00					0.00
<20 years	868	18.7	868	18.7		1,767	12.9	1,767	12.9	
20-29 years	923	19.9	923	19.9		2,931	21.3	2,931	21.3	
30-39 years	1648	35.5	1,648	35.5		5,213	37.9	5,213	37.9	
40-49 years	871	18.8	871	18.8		2,877	20.9	2,877	20.9	
50-59 years	272	5.9	272	5.9		797	5.8	797	5.8	
60-69 years	53	1.1	53	1.1		132	1.0	132	1.0	
70+ years	10	0.2	10	0.2		36	0.3	36	0.3	
Sex					0.00					0.00
Male	2,339	50.4	2,339	50.4		9,467	68.8	9,467	68.8	
Female	2,306	49.6	2,306	49.6		4,286	31.2	4,286	31.2	
Nationality[e]					0.00					0.00
Bangladeshi	235	5.1	235	5.1		1,334	9.7	1,334	9.7	
Egyptian	358	7.7	358	7.7		990	7.2	990	7.2	
Filipino	764	16.5	764	16.5		1,610	11.7	1,610	11.7	
Indian	789	17.0	789	17.0		3,481	25.3	3,481	25.3	
Nepalese	170	3.7	170	3.7		1,283	9.3	1,283	9.3	
Pakistani	192	4.1	192	4.1		542	3.9	542	3.9	
Qatari	762	16.4	762	16.4		1,288	9.4	1,288	9.4	
Sri Lankan	125	2.7	125	2.7		538	3.9	538	3.9	
Sudanese	166	3.6	166	3.6		442	3.2	442	3.2	
Other nationalities[f]	1,084	23.3	1,084	23.3		2,245	16.3	2,245	16.3	
Prior infection[g]					0.31					0.28
No prior infection	4,638	99.8	4,413	95.0		13,629	99.1	12,938	94.1	
>90 days	7	0.2	232	5.0		124	0.9	815	5.9	

Abbreviations: IQR, interquartile range; PCR, polymerase chain reaction; SMD, standardized mean difference.

[a]Cases and controls were matched one-to-one by sex, 10-year age group, nationality, and calendar week of PCR test.

[b]SMD is the difference in the mean of a covariate between groups divided by the pooled standard deviation. An SMD<0.1 indicates adequate matching.

[c]Values are expressed as median (interquartile range).

[d]SMD is the mean difference between groups divided by the pooled standard deviation.

[e]Nationalities were chosen to represent the most populous groups in Qatar.

[f]These comprise 61 other nationalities in Qatar in the Alpha variant analysis and 78 other nationalities in the Beta variant analysis.

[g]Given our interest in quantifying differentials in the odds of exposure to prior infection between cases and controls, this variable was not included as a matching factor.

29

Figure 1. Schematic diagrams of mathematical models used in this study. A) Classic SEIR model extended to allow for reinfections (Baseline Model). B) Baseline Model extended to include vaccination (Vaccination Model). C) Baseline Model extended to include waning in protection of prior infection against reinfection (Waning of Immunity Model).

Figure 2. The 2-by-2 tables and equations used to estimate effectiveness of prior infection in preventing reinfection (PE_S) using the test-negative, case-control study design. A) PE_S estimated in absence of bias. B) PE_S estimated in presence of misclassification of prior infection. C) PE_S estimated in presence of misclassification of latent infection. D) PE_S estimated in presence of misclassification of current active infection. E) PE_S estimated in presence of vaccination scale-up.

Figure 3. Simulated SARS-CoV-2 epidemic through two epidemic waves. A) Daily number of new infections. B) Proportion of the population ever infected. C) Scale-up of vaccine coverage.

Figure 4. Estimated effectiveness of prior infection in preventing reinfection using the test-negative study design ($PE_S^{test-negative}$) compared with the true effectiveness of prior infection in preventing reinfection (PE_S^{true}). A) $PE_S^{test-negative}$ versus PE_S^{true} in presence of no waning of protection (Baseline Model). B) $PE_S^{test-negative}$ versus PE_S^{true} month by month after the prior infection in presence of gradual waning of protection against reinfection (Waning of Immunity Model). This figure was generated using the instantaneous prevalence at each time point for each population.

Figure 5. Impact of bias in estimating effectiveness of prior infection in preventing reinfection using the test-negative study design ($PE_S^{test-negative}$). A) Impact of misclassification of prior infection. B) Impact of misclassification of latent infection. C) Impact of misclassification of current active infection. D) Impact of scale-up of vaccination in the population. This figure was generated using the instantaneous prevalence at each time point for each population.

Figure 6. Flowchart describing the population selection process to estimate effectiveness of prior infection in preventing reinfection using the test-negative study design. Individuals with a PCR-confirmed infection with SARS-CoV-2 Alpha or Beta variant were exact matched on a 1:1 ratio by sex, 10-year age group, nationality, and PCR test calendar week to the first eligible PCR-negative individual. Prior infection records were retrieved for all matched individuals.
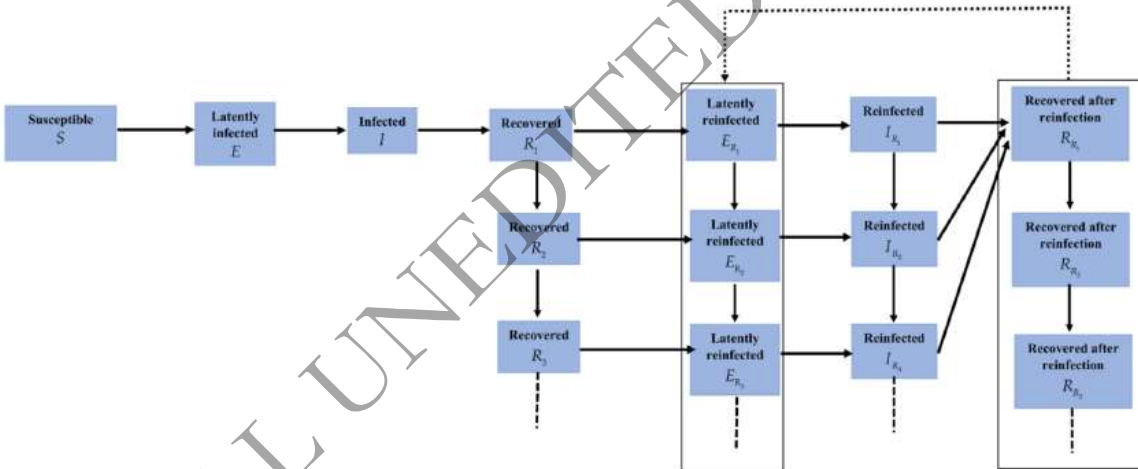
30
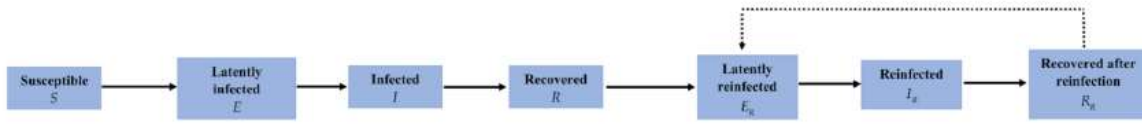
(A)   **Baseline Model**

(B)   **Vaccination Model**
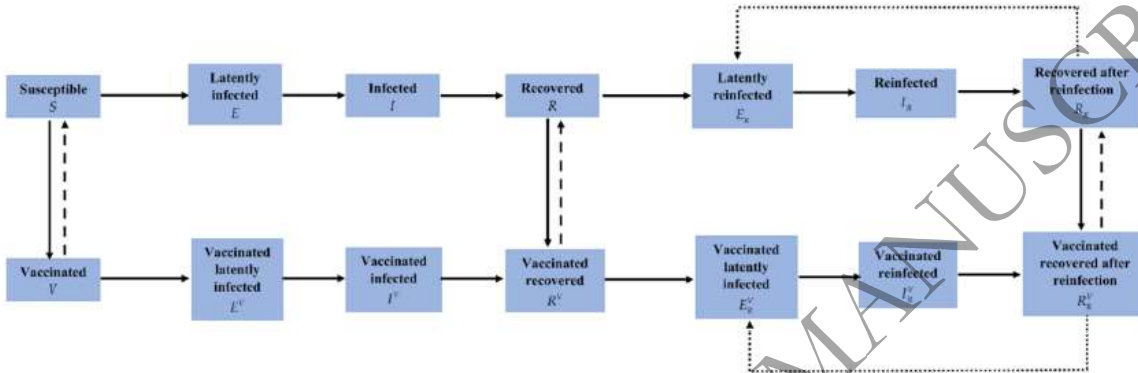
(C)   **Waning of Immunity Model**

(A)

PE_S Estimated in Absence of Bias (Baseline Model)

	Cases	Controls
Prior infection	$E_R + I_R$	$R + R_R$
No prior infection	$E + I$	S

$$PE_S = 1 - \frac{(E_R + I_R)S}{(E+I)(R+R_R)}$$

(B)

PE_S Estimated in Presence of Misclassification of Prior Infection (Baseline Model)

	Cases	Controls
Prior infection	$(E_R+I_R)\times(1-g_P)$	$(R+R_R)\times(1-g_P)$
No prior infection	$(E+I)+(E_R+I_R)\times g_P$	$S+(R+R_R)\times g_P$

$$PE_S = 1 - \frac{(E_R+I_R)\left[S+(R+R_R)\times g_P\right]}{\left[(E+I)+(E_R+I_R)\times g_P\right](R+R_R)}$$

(C)

PE_S Estimated in Presence of Misclassification of Latent Infection (Baseline Model)

	Cases	Controls
Prior infection	$E_R\times(1-g_E)+I_R$	$R+R_R+E_R\times g_E$
No prior infection	$E\times(1-g_E)+I$	$S+E\times g_E$

$$PE_S = 1 - \frac{\left[E_R\times(1-g_E)+I_R\right]\left[S+E\times g_E\right]}{\left[E\times(1-g_E)+I\right]\left[R+R_R+E_R\times g_E\right]}$$

(D)

PE_S Estimated in Presence of Misclassification of Current Active Infection (Baseline Model)

	Cases	Controls
Prior infection	$(E_R+I_R)\times(1-g_I)$	$(R+R_R)+(E_R+I_R)\times g_I$
No prior infection	$(E+I)\times(1-g_I)$	$S+(E+I)\times g_I$

$$PE_S = 1 - \frac{(E_R+I_R)\left[S+(E+I)\times g_I\right]}{(E+I)\left[(R+R_R)+(E_R+I_R)\times g_I\right]}$$

(E)

PE_S Estimated in Presence of Scale-up of Vaccination in the Population (Vaccination Model)

	Cases	Controls
Prior infection	$(E_R+I_R)+(E_R^V+I_R^V)$	$(R+R_R)+(R^V+R_R^V)$
No prior infection	$(E+I)+(E^V+I^V)$	$S+V$

$$PE_S = 1 - \frac{\left[(E_R+I_R)+(E_R^V+I_R^V)\right](S+V)}{\left[(E+I)+(E^V+I^V)\right]\left[(R+R_R)+(R^V+R_R^V)\right]}$$

g_P is the proportion of prior infections that are undocumented.
g_E is the proportion of latent infections that are undocumented.
g_I is the proportion of current active infections that are undocumented.

```
┌──────────────────────────────────────┐          ┌──────────────────────────────────────┐
│ 112,778 individuals with a PCR-        │          │ 514,763 individuals with PCR-negative  │
│ confirmed SARS-CoV-2 infection         │          │ SARS-CoV-2 tests between March 8-      │
│ between March 8-April 21, 2021         │          │ April 21, 2021 (end of study)          │
│ (end of study)                         │          │                                        │
└──────────────────────────────────────┘          └──────────────────────────────────────┘
```

66,082 individuals were excluded because they were not tested for clinical suspicion

446,698 individuals were excluded because they were not tested for clinical suspicion

7,263 individuals were excluded because they had a vaccination record before the PCR test

20,436 individuals were excluded because they had a vaccination record before the PCR test

3,522 individuals were excluded because they had a record for a prior infection within 90 days of the PCR test

1,330 individuals were excluded because they had a record for a prior infection within 90 days of the PCR test

35,911 eligible individuals with a PCR-confirmed SARS-CoV-2 infection who were tested for clinical suspicion

46,932 eligible individuals with PCR-negative SARS-CoV-2 tests who were tested for clinical suspicion

16,775 individuals were excluded because the variant status of the PCR-confirmed infection is not known

4,794 individuals with a PCR-confirmed infection with the Alpha variant

14,342 individuals with a PCR-confirmed infection with the Beta variant

149 not matched

589 not matched

42,287 not matched

33,179 not matched

4,645 matched to individuals with PCR-negative SARS-CoV-2 tests

13,753 matched to individuals with PCR-negative SARS-CoV-2 tests

4,645 matched to individuals with a PCR-confirmed infection with the Alpha variant

13,753 matched to individuals with a PCR-confirmed infection with the Beta variant

36