Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Evelyn Shi Date

Bias in Vaccine Effectiveness Estimation Under Changing Testing Probabilities When Using the Test-Negative Design

By

Evelyn Shi

Master of Science in Public Health

Department of Biostatistics and Bioinformatics

_____

Natalie Dean, Ph.D.
(Advisor)

_____

Avnika Amin, Ph.D.
(Reader)

Bias in Vaccine Effectiveness Estimation Under Changing Testing Probabilities When Using the Test-Negative Design

By

Evelyn Shi

B.S., Purdue University, 2017

Thesis Committee Chair: Natalie Dean, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in the Department of Biostatistics and Bioinformatics
2023

Abstract

Background: In recent years, the test negative design (TND) has emerged as a popular study design in estimating vaccine effectiveness (VE) for COVID-19 vaccines. While a useful and important tool, this design has major considerations that must be made before using it to estimate VE. It is vital to understand how simple changes in test reporting or differences in added protection in disease severity levels can lead to a change in VE estimates.

Objectives: We would like to characterize how sensitive VE estimates from a test-negative design are to changes in test reporting, and how it varies by disease severity and vaccine protection at different levels of disease severity.

Methods: We used a simulation to generate COVID-19 test results and demographic data and study this relationship. In particular, we looked at different patterns of testing probabilities and vaccine protection and how they would bias VE estimates in five different settings.

Results: Vaccine effectiveness for protection against infection has been found to be biased by as much as 8.64% (an estimated VE_S of 78.6% and a true value of 70%) in settings where testing probabilities were differential by disease severity and there was added vaccine protection against severe cases like hospitalization.

Conclusions: VE_S is often biased and overestimated when testing probabilities are differential by disease severity, and this is exacerbated by added vaccine protection against severe outcomes. These overestimations were due to a much greater fraction of hospitalized or severe cases represented in the test results compared to those who were mildly symptomatic or asymptomatic.

Bias in Vaccine Effectiveness Estimation Under Changing Testing Probabilities When Using the Test-Negative Design

By

Evelyn Shi

B.S., Purdue University, 2017

Thesis Committee Chair: Natalie Dean, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in the Department of Biostatistics and Bioinformatics
2023

Introduction

Coronavirus disease 2019 (COVID-19) is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). SARS-CoV-2 was first detected in Wuhan, Hubei province, in China in December 2019 (Page, 2021). The virus quickly spread worldwide, and was declared a pandemic on March 11, 2020 by the World Health Organization (WHO, 2020).

The United States began mass vaccination in December 2020 with the BNT162b2 (Pfizer-BioNTech) vaccine (Pereira, 2020). BNT162b2 was granted Emergency Use Authorization (EUA) by the Food and Drug Administration (FDA, 2020). To receive an EUA, the vaccine was rigorously tested in a multinational, placebo-controlled, and observer-blinded clinical trial, where participants 16 years or older received two doses of either BNT162b2 or the placebo (Polack et al., 2020). Primary endpoints of the study were efficacy of the vaccine against laboratory-confirmed COVID-19 and safety.

The results of the 2020 Polack et al. Pfizer study found BNT162b2 was 95% (95% CI: 90.3 - 97.6) effective against COVID-19 (Polack et al., 2020). Other vaccines available in the US include mRNA-1273 (Moderna) and Ad26.COV2.S (Janssen/Johnson & Johnson) (Edwards & Orenstein, 2023). Later in December 2020, the mRNA-1273 vaccine was granted an EUA, followed by Ad26.COV2.S a few months later in February 2021. The mRNA-1273 vaccine was found to have a vaccine efficacy of 94.1% against COVID-19 illness (Baden et al., 2021). In comparison, the Ad26.COV2.S vaccine protected against moderate to severe-critical COVID-19 at 66.1% efficacy (Sadoff et al., 2021). It was higher against severe-critical COVID-19, at 85.4% efficacy (Sadoff et al., 2021). By April, all states in the US opened vaccine eligibility to residents 16 and over. In May, the FDA gave the Pfizer

BNT162b2 vaccine full approval for adolescents aged 12 to 15, and in August full approval was given for patients 16 and older. By August, 70% of the US population had received at least one dose.

Vaccine efficacy, the result of these studies, is estimated as one minus the relative risk of vaccinated participants who developed COVID-19 compared to those who took the placebo and developed it (WHO, 2021). While clinical trials can show vaccine efficacy in a controlled setting, other factors like underlying medical conditions or age can impact how well a vaccine can work for a particular person. Vaccine effectiveness (VE), as opposed to vaccine efficacy, looks at how the vaccine performs in the real world outside of controlled trial conditions and is calculated the same way. Importantly, vaccine efficacy found in clinical trials also tends to be underpowered when looking at rare outcomes, or other subgroups (Lipsitch & Dean, 2020).

Since the completion of clinical trials, there has been widespread emergence of new SARS-CoV-2 variants. Thus, observational studies, in particular the test negative design (TND), have been a major source of information in monitoring vaccine performance outside of the controlled setting of a clinical trial. Pandemic conditions during the trials were quite different from conditions during vaccine rollout, and so it is necessary to understand what impact this could have on VE. Observational studies have been used to assess VE against severe disease and against variants such as Delta or Omicron (Polinski et al., 2022; Tang et al., 2021; Tsang et al., 2022). Both the Polinski and Tang papers, in finding these estimates, used the TND. The test negative study design has been used to estimate vaccine effectiveness for influenza and rotavirus (Bond 2016, Jackson & Nelson 2013, Schwartz 2017),

though more recently, has been used in many COVID-19 VE estimates (Lopez-Bernal 2021, Ranzani 2021, Thompson 2021).

In a test negative design, patients who seek healthcare for COVID-like symptoms go to their healthcare provider and have specimens taken in order to test for SARS-CoV-2. Those who meet the clinical case definition of COVID-19 and test positive are considered the test-positive cases, while those who tested negative but also met the clinical case definition are the test-negative group. The odds of testing positive among vaccinated participants is then compared with the odds of testing positive among unvaccinated participants, adjusting for measured confounders like age and location. VE can then be calculated from this odds ratio (OR), as VE = (1 - the adjusted OR) * 100%.

In many retrospective COVID-19 TND studies, there is no consistent symptom presentation. Since existing data does not tell us the disease status of the individual, this can result in a mix of asymptomatic, mild, moderate, and severe symptoms. Yet it is known that SARS-CoV-2 infections are heavily underreported (Lau et al., 2021). Those who seek testing are not representative of the entire infected population, as those with mild symptoms may choose not to get tested, or those who present as asymptomatic may not be known to test. This is in contrast to individuals with severe disease, who are more likely to be captured by the surveillance system (Vandenbroucke et al., 2020). Many TND studies calculate and report VE against infection or against symptomatic COVID-19, including mild disease. Furthermore, VE is often seen to vary across disease severity, and vaccines usually have higher efficacy against severe disease than milder symptoms (Polack et al., 2020). This can be observed in the

Sadoff et al. Ad26.COV2.S trial, and similar relationships have been seen for many other diseases (Jain et al., 2013; Kulkarni et al., 2017; Trach et al., 1997; Alonso et al., 2004).

The characteristics of reported tests have also changed over time. In late 2021, during the Omicron surge of COVID, access to self-administered at-home rapid antigen tests was expanded, allowing rapid, simple-to-use, and sensitive tests to be utilized to quickly diagnose infected patients (Shircliff). However, rapid, at-home tests are not required to be reported in surveillance data, which leads to further changes in test-reporting patterns – further underreporting, particularly in mildly or asymptomatic cases. While under-reporting is not a new issue, the dramatic changes in testing patterns over time may exacerbate a pre-existing issue in the TND, and complicate the interpretation of TND studies.

We hypothesize that the expanded use of at-home tests have impacted the COVID-19 case data by further decreasing the rate of mild cases or infections reported. We hypothesize that differences in test reporting and VE by disease severity can cause under-representation of milder infections, and an over-representation of severe infections, leading to an overestimation of VE against infection. We would like to know how, even without underlying changes in true VE, how changes in test reporting impact VE estimated from a TND study. How sensitive are these estimates to changes in test reporting, and how does it vary across disease severity? The rollout of at-home testing is a motivating example for this setting.

Methods

In order to examine the relationship between disease severity, reporting, and VE, we began by using a simulation to generate data for a hypothetical COVID-19 TND study to estimate VE against infection, against disease, and against hospitalization. Fixed and varied parameter values used for this data generation are defined in Tables 1 and 2 below.

Table 1. Fixed Parameters in the Simulation

Parameter	**Value**
Overall Parameters	
Population Size	10,000 people
Study End Time	1,000 units of time
Vaccination Coverage	50% of population
Percent of Population that is Older	0.4
Hazard Ratio Relating Testing Positive and Older Age	1.2
Hazard Ratio Relating Testing Negative and Older Age	2.0
Vaccine Efficacy Against Infection	0.7

Parameters Regarding Testing Positive	
Shape Parameter for Weibull Distribution for Testing Positive	1.2
Rate Parameter for Weibull Distribution for Testing Positive	0.01
Baseline Probability of a Younger Unvaccinated Person Progressing to Symptoms Given Infection	0.7
Baseline Probability of a Younger Unvaccinated Person Progressing to Severe Disease Given Symptoms	0.05
Baseline Probability of an Older Unvaccinated Person Progressing to Symptoms Given Infection	0.7
Baseline Probability of an Older Unvaccinated Person Progressing to Severe Disease Given Symptoms	0.15
Parameters Regarding Testing Negative	
Shape Parameter for Weibull Distribution for Testing Negative	0.8
Rate Parameter for Weibull Distribution for Testing Negative	0.01
Baseline Probability of a Younger Unvaccinated Person Progressing to Symptoms Given Infection	0.5
Baseline Probability of a Younger Unvaccinated Person Progressing to Severe Disease Given Symptoms	0.01
Baseline Probability of an Older Unvaccinated Person Progressing to Symptoms Given Infection	0.5
Baseline Probability of an Older Unvaccinated Person Progressing to Severe Disease Given Symptoms	0.05

Table 2. Varied Parameters in the Simulation

Parameter	**Values**
Vaccine Efficacy Against Progression to Symptoms Given Infection	0, 0.4
Vaccine Efficacy Against Progression to Severe Disease Given Symptoms	0, 0.2
Parameters Regarding Testing Probabilities	
Probability of a Vaccinated Person with no Symptoms Seeks Testing	0.1, 0.2, 0.7
Probability of a Vaccinated Person with Mild Symptoms Seeking Testing	0.5, 0.7
Probability of a Vaccinated Person with Severe Symptoms Seeking Testing	0.7, 0.9
Probability of an Unvaccinated Person with no Symptoms Seeks Testing	0.025, 0.1, 0.2, 0.7
Probability of an Unvaccinated Person with Mild Symptoms Seeking Testing	0.2, 0.35, 0.7
Probability of an Unvaccinated Person with Severe Symptoms Seeking Testing	0.3, 0.45, 0.7, 0.9

These simulations aimed to look at the effect of varying levels of vaccine efficacy and different patterns of testing probability, while holding other parameters constant.

Beyond these parameters, there are a few things that should be defined. The age of an individual is generated and considered in the simulation, and is split into “older” and “younger”. There is no set age

for older vs younger, but 50% of the population is set to be older, giving them a different hazard for testing positive or negative, and how likely they are to get infected, progress to symptoms, and/or progress to hospitalization.

There are also three levels of infection severity an individual can have: 1) infection with no symptoms, 2) non-severe COVID-like or COVID symptoms, and 3) severe COVID-like or COVID symptoms. In this paper, severe COVID-like or COVID symptoms are represented by hospitalization. As an important part of the research question, true VE is differential by disease severity. VE_S is vaccine efficacy against infection, VE_P is vaccine efficacy against progression to symptoms (mild or severe) given infection, and VE_H is vaccine efficacy against progression to severe disease given symptoms.

There are several fundamental assumptions made in this model. A major model assumption is that the COVID-19 test has 100% sensitivity and specificity. We also assume perfect knowledge of an individual's vaccination status, and there are no coinfections between SARS-CoV-2 and any other respiratory pathogens. There is only one type of vaccine, and only one circulating COVID variant. Vaccination does not affect the probability of non-COVID infections. That is, there is no cross-protection from the vaccine. Furthermore, there is no waning in vaccine protection in that protection does not fade over time.

For every individual, vaccination status and age category are randomly generated. From these two covariates, a hazard ratio for increased SARS-CoV-2 infection risk for that individual is calculated, based on their unique covariates. This hazard ratio is calculated relative to the hazard ratio for a young, unvaccinated person – if the individual is vaccinated, the coefficient is 0.3, and if they are older, the coefficient is 1.2. A rate parameter is then calculated for that individual, dependent on the previously

calculated hazard ratio, the shape parameter of 1.2, and the original rate parameter of 0.01. A COVID-19 infection time then is drawn from a Weibull distribution (shape parameter = 1.2, and using the individual's newly calculated rate parameter). Once infected, the simulation then looks to see if the individual will progress to symptoms. While both younger and older individuals will both have a probability of 0.7 in progressing to symptoms, progression from symptoms to severe disease is dependent on age. In an older person, the probability is 0.15, compared to 0.05 in a younger person. Given these probabilities, as well as vaccination status and whether or not the vaccine has any added protection against symptoms or hospitalization, VE_P or VE_H, it is then randomly determined to see if the individual progresses to symptoms or hospitalization. This determines the severity of the COVID-19 infection, and this individual's probability of testing is then based on their severity as well as their vaccination status.

Based on the individual's probability of testing, they are randomly assigned testing or not, and if they tested, those results are stored and then used to estimate VE. This imitates the reality of the TND, where only information about individuals with recorded test results are known.

A similar process is done to obtain test negative results. Like in a test positive case, an individual's hazard ratio for decreased SARS-CoV-2 infection risk is calculated depending on their age. As test negative controls are non-COVID illnesses, the vaccination should have no effect on these and thus, is not included in the hazard ratio calculation. The coefficient for age when testing negative is 2. A rate parameter is calculated from this hazard ratio, and from a shape parameter of 0.8 and original rate parameter of 0.01. An illness time is drawn from a Weibull distribution conditioning on when the last test negative test was, the individual's rate parameter, and shape parameter. Probability of progression

is done the same as the test positive process, except the baseline probability for progression to symptoms given infection is 0.5 for both old and young individuals. Younger people progress from symptoms to severe disease with a probability of 0.01, while that probability is 0.05 for older people. It is solely these probabilities that are used to determine whether or not the individual develops symptoms or progresses from symptoms to severe disease. Probability of test-seeking is dependent on vaccination status and disease severity, and the test result is only stored if the person sought testing. This process can go on for multiple times within the time before the study ends, allowing multiple non-SARS-CoV-2 infections to occur. Furthermore, this process occurs earlier in the simulation study period than the SARS-CoV-2 infection, so the individual obtaining SARS-CoV-2 infection marks the end of follow-up for this individual.

The test results are coded 1 if the individual has COVID, and 0 if otherwise. In order to obtain a VE estimate, the TND uses a logistic regression, where in this case the formula was as follows:

$$logit(Pr\ (Result = 1)) = \beta_1 VaxStatus + \beta_2\, Age + \beta_3 Time$$

In this formula, *Result* can be 1 or 0, where the former indicates a positive COVID-19 test and the latter is a negative test. *VaxStatus* is 1 or 0, where the variable is 1 if the individual is vaccinated, and 0 if not. *Age* is coded into "older" (*Age* = 1) and "younger" (*Age* = 0). *Time* is fit using a smoothing spline with df = 3.

The formula, using all simulated test results, would give the VE estimate for VE_S, which looks at vaccine effectiveness against infection, and uses the data from all test results. When calculating VE_{SP},

we removed asymptomatic people, and in calculating $VE_{SPH,}$ only data for hospitalized cases remained. All three used the same model and covariates, only with data split by severity.

While VE_P and VE_H are used as parameters in the simulation, the TND VE estimates derived from the simulated data represent VE_{SP} and VE_{SPH}. VE_{SPH} is what is observed from cohort or TND studies looking at VE against hospitalization, since this estimate only looks at hospitalized cases.

The relationship between these values can be found in the following equations:

$$VE_{SP} = 1 - (1 - VE_S)(1 - VE_P)$$
$$VE_{SPH} = 1 - (1 - VE_S)(1 - VE_P)(1 - VE_H)$$

In this simulation, we want to evaluate the possible bias of VE estimates under different settings in order to determine how VE changes by disease severity and under-reporting. In this paper, bias is defined as the estimated value minus the true value. Table 2 below lists the five scenarios tested by the simulation. The first is the base scenario, where the probability of testing is the same, no matter vaccination status or disease severity. These probabilities are listed, alongside VE_S, VE_P, VE_H, and VE_{SP} and VE_{SPH}.

The second scenario is where the probability of testing is the same between vaccinated and unvaccinated persons, but different by disease severity. The third scenario further changes things by making the probability of testing among the unvaccinated half of that in the vaccinated group, where VE is still differential by symptom severity. The fourth scenario represents the potential influence of self-testing, where all probabilities of testing are much lower than that of the previous scenarios, as a way to represent those who self-tested or used rapid-testing at home and did not report the results.

Additionally, the relationship between vaccinated and unvaccinated testing probabilities is not just half. Finally, the last setting looks at when the vaccine only protects against infection, so that VE_S, VE_{SP}, and VE_{SPH} would all be the same.

Table 2. Table listing tested scenarios and VE parameters

Test Reporting		**VE by severity**	**True VE values**
Same Probability of Testing by Disease Severity and Vaccination Status			
$P_{Infected,Vaccinated} = 0.7$	$P_{Infected, Unvaccinated} = 0.7$	$VE_S \rightarrow 0.7$	$VE_S \rightarrow 0.7$
$P_{Mildly\ Symptomatic, Vaccinated} = 0.7$	$P_{Mildly\ Symptomatic, Unvaccinated} = 0.7$	$VE_P \rightarrow 0.4$	$VE_{SP} \rightarrow 0.82$
$P_{Hospitalized, Vaccinated} = 0.7$	$P_{Hospitalized, Unvaccinated} = 0.7$	$VE_H \rightarrow 0.2$	$VE_{SPH} \rightarrow 0.856$
Probability of Testing Differential by Disease Severity, Same by Vaccination Status			
$P_{Infected,Vaccinated} = 0.2$	$P_{Infected, Unvaccinated} = 0.2$	$VE_S \rightarrow 0.7$	$VE_S \rightarrow 0.7$
$P_{Mildly\ Symptomatic, Vaccinated} = 0.7$	$P_{Mildly\ Symptomatic, Unvaccinated} = 0.7$	$VE_P \rightarrow 0.4$	$VE_{SP} \rightarrow 0.82$
$P_{Hospitalized, Vaccinated} = 0.9$	$P_{Hospitalized, Unvaccinated} = 0.9$	$VE_H \rightarrow 0.2$	$VE_{SPH} \rightarrow 0.856$
Probability of Testing Differential by Disease Severity and Vaccination Status			

$P_{Infected,Vaccinated} = 0.2$	$P_{Infected, Unvaccinated} = 0.1$	$VE_S \rightarrow 0.7$	$VE_S \rightarrow 0.7$
$P_{Mildly\ Symptomatic, Vaccinated} = 0.7$	$P_{Mildly\ Symptomatic, Unvaccinated} = 0.35$	$VE_P \rightarrow 0.4$	$VE_{SP} \rightarrow 0.82$
$P_{Hospitalized, Vaccinated} = 0.9$	$P_{Hospitalized, Unvaccinated} = 0.45$	$VE_H \rightarrow 0.2$	$VE_{SPH} \rightarrow 0.856$
Self Testing (More Extreme Under-Reporting, Unvaccinated Testing Probability Not Half of Vaccinated)			
$P_{Infected,Vaccinated} = 0.1$	$P_{Infected, Unvaccinated} = 0.025$	$VE_S \rightarrow 0.7$	$VE_S \rightarrow 0.7$
$P_{Mildly\ Symptomatic, Vaccinated} = 0.5$	$P_{Mildly\ Symptomatic, Unvaccinated} = 0.2$	$VE_P \rightarrow 0.4$	$VE_{SP} \rightarrow 0.82$
$P_{Hospitalized, Vaccinated} = 0.7$	$P_{Hospitalized, Unvaccinated} = 0.3$	$VE_H \rightarrow 0.2$	$VE_{SPH} \rightarrow 0.856$
Vaccine Effectiveness Same Across Symptom Severity			
$P_{Infected,Vaccinated} = 0.2$	$P_{Infected, Unvaccinated} = 0.1$	$VE_S \rightarrow 0.7$	$VE_S \rightarrow 0.7$
$P_{Mildly\ Symptomatic, Vaccinated} = 0.7$	$P_{Mildly\ Symptomatic, Unvaccinated} = 0.35$	$VE_P \rightarrow 0$	$VE_{SP} \rightarrow 0.7$
$P_{Hospitalized, Vaccinated} = 0.9$	$P_{Hospitalized, Unvaccinated} = 0.45$	$VE_H \rightarrow 0$	$VE_{SPH} \rightarrow 0.7$

The number of runs for each scenario was chosen to be 500, estimated using the same method as Morris et al. and some trial and error (Morris, 2019). Bias was the key performance measure of interest in this study, so we used the corresponding equation. After some preliminary runs, we found that standard error was less than 0.01, meaning that the variance was less than 0.001. Given the following

equation (1) and a desire to have a Monte Carlo SE of bias lower than 0.005, we obtain 400 as the number of simulation runs.

$$Monte\ Carlo\ SE(Bias) = \sqrt{Var(\hat{\theta})/n_{sim}} \quad (1)$$

When running preliminary results, a simulation run of 400 times on occasion had a MCSE of over 0.005, so the number of runs was increased to 500.

Results

Table 3 shows the results of running the simulation 500 times for each scenario with a population size N of 10,000. It reports the estimated parameter (VE_S, VE_{SP}, and VE_{SPH}), the expected or true value, the average of all 500 runs, standard error of the estimate, mean bias, Monte Carlo SE of the bias, and root mean squared error.

Table 3. Results with population size (N) = 10,000 and number of simulation runs = 500

Estimated Parameter	**True Value**	**Average Estimate (of 500 runs)**	**Standard Error of the Estimate Across**	**Mean Bias**	**Monte Carlo Standard Error of**	**Root Mean Squared Error**

			Simulations		**Bias**	
Same Probability of Testing by Disease Severity and Vaccination Status						
VE_S	0.7	0.699	0.00994	-0.001	0.000447	0.00998
VE_{SP}	0.82	0.820	0.00871	-0.0002	0.000389	0.00870
VE_{SPH}	0.856	0.845	0.0297	-0.0109	0.00141	0.0316
Probability of Testing Differential by Disease Severity, Same by Vaccination Status						
VE_S	0.7	0.778	0.00926	0.0775	0.00350	0.0781
VE_{SP}	0.82	0.820	0.00861	0.000493	0.000386	0.00861
VE_{SPH}	0.856	0.845	0.0255	-0.0107	0.00124	0.0276
Probability of Testing Differential by Disease Severity and Vaccination Status						
VE_S	0.7	0.769	0.0117	0.0689	0.00313	0.0699
VE_{SP}	0.82	0.813	0.0108	-0.00722	0.000582	0.0130
VE_{SPH}	0.856	0.837	0.0321	-0.0194	0.00168	0.0375
Self Testing (More Extreme Under-Reporting, Unvaccinated Testing Probability Not Half of Vaccinated)						

VE_S	0.7	0.786	0.0155	0.0864	0.00393	0.0878
VE_{SP}	0.82	0.811	0.0152	-0.00883	0.000786	0.0176
VE_{SPH}	0.856	0.835	0.0380	-0.0210	0.00194	0.0434
Vaccine Effectiveness Same Across Symptom Severity						
VE_S	0.7	0.691	0.0149	-0.00931	0.000785	0.0175
VE_{SP}	0.7	0.692	0.0163	-0.00841	0.000821	0.0183
VE_{SPH}	0.7	0.686	0.0487	-0.0136	0.00226	0.0505

For all scenarios and parameter estimates, Monte Carlo Standard Errors of the Bias were lower than 0.005. Most average estimates were close to the true value of the parameter, save a few. For instance, VE_S in the second, third, and fourth scenarios (0.778, 0.769, and 0.786, respectively, compared to the true value of 0.7). Standard error of the estimate was generally larger for VE_{SPH} than when estimating VE_{SP} and VE_S (around 0.03 for VE_{SPH} and 0.01 for VE_{SP} and VE_S). This was the same pattern seen in root mean square error in the base scenario and VE same by severity scenario (around 0.03 or 0.05 for VE_{SPH}, and 0.01 for VE_{SP} and VE_S). However, the root mean square error for the other three scenarios had higher values for VE_S and VE_{SPH} than VE_{SP}.

Overall, mean bias seemed to be more negative than positive, indicating more underestimation than overestimation. While the first setting had little bias, the second, third, and forth settings all had major

overestimation when looking at VE_S (0.778, 0.769, and 0.786 were estimated, compared to the true value of 0.7). Some milder underestimation was seen in the VE_{SPH} estimates for those settings (0.845, 0.837, and 0.835 compared to the true value of 0.856). The last scenario mildly underestimated all VE estimates.

Next, plots were made to further examine the relationship between testing probability and VE estimates for VE_S, VE_{SP}, and VE_{SPH}. The first set of plots are shown below in Figure 1.

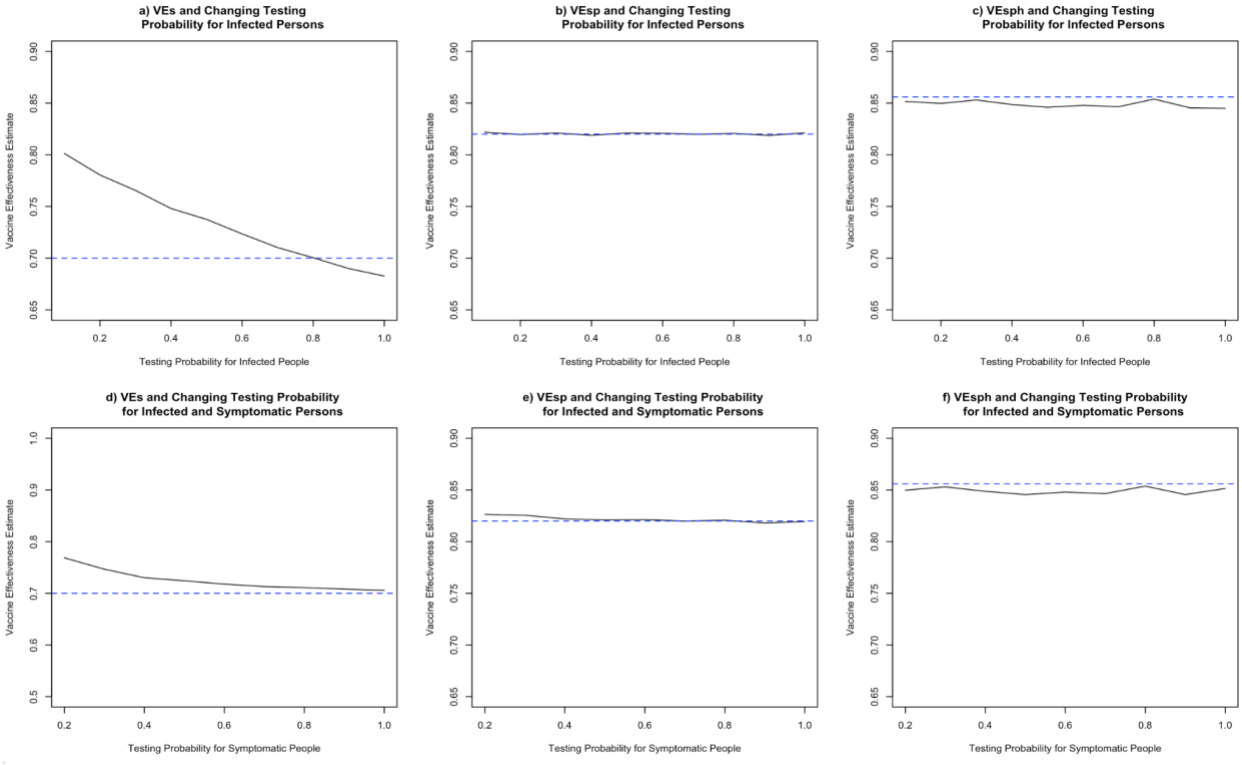

Figure 1. VE_S, VE_{SP}, and VE_{SPH} Estimates for Varying Testing Probability. VE_P and VE_H are 0.4 and 0.2, respectively. The true VE_S is 0.7, shown as a dotted blue horizontal line on VE_S. The true VE_{SP} is 0.82, and the true VE_{SPH} value is 0.856, both plotted on their respective graphs. Plot a) shows VE_S as testing probability for infected people increases from 0.1 to 1 as testing probability for symptomatic people and hospitalized

people remain at 0.8 and 0.9, respectively. Plot b) follows the same setting but shows VE_{SP} instead of VE_S. Plot c) is the same setting as the previous two, with VE_{SPH} estimates. Plot d) shows VE_S as testing probability for symptomatic people increases from 0.2 to 1 and testing probability for infected people increases from 0.1 to 0.9, always 0.1 lower than symptomatic people. The testing probability for hospitalized people is constant at 0.9. Plot e) is the same setting, but shows VE_{SP}. Plot f) likewise, is the same setting but shows VE_{SPH}.

In this figure, we can see that the general patterns are the same no matter if testing probability for symptomatic people is included. VE_{SP} and VE_{SPH} both show flat lines that approximate the true VE values (Figure 1b, 1c, 1e, and 1f). Meanwhile when looking at the VE_S plots, we can see that as the testing probability increases, the estimate approaches the true value. There is a good amount of overestimation when testing probabilities for infected (and symptomatic) people are low. In Figure 1a), the VE_S estimate starts to be underestimated as testing probability for infected people becomes greater than 0.9.

Figure 2 looks at a similar setting as Figures 1d, 1e, and 1f), but varies VE_P and VE_H values.

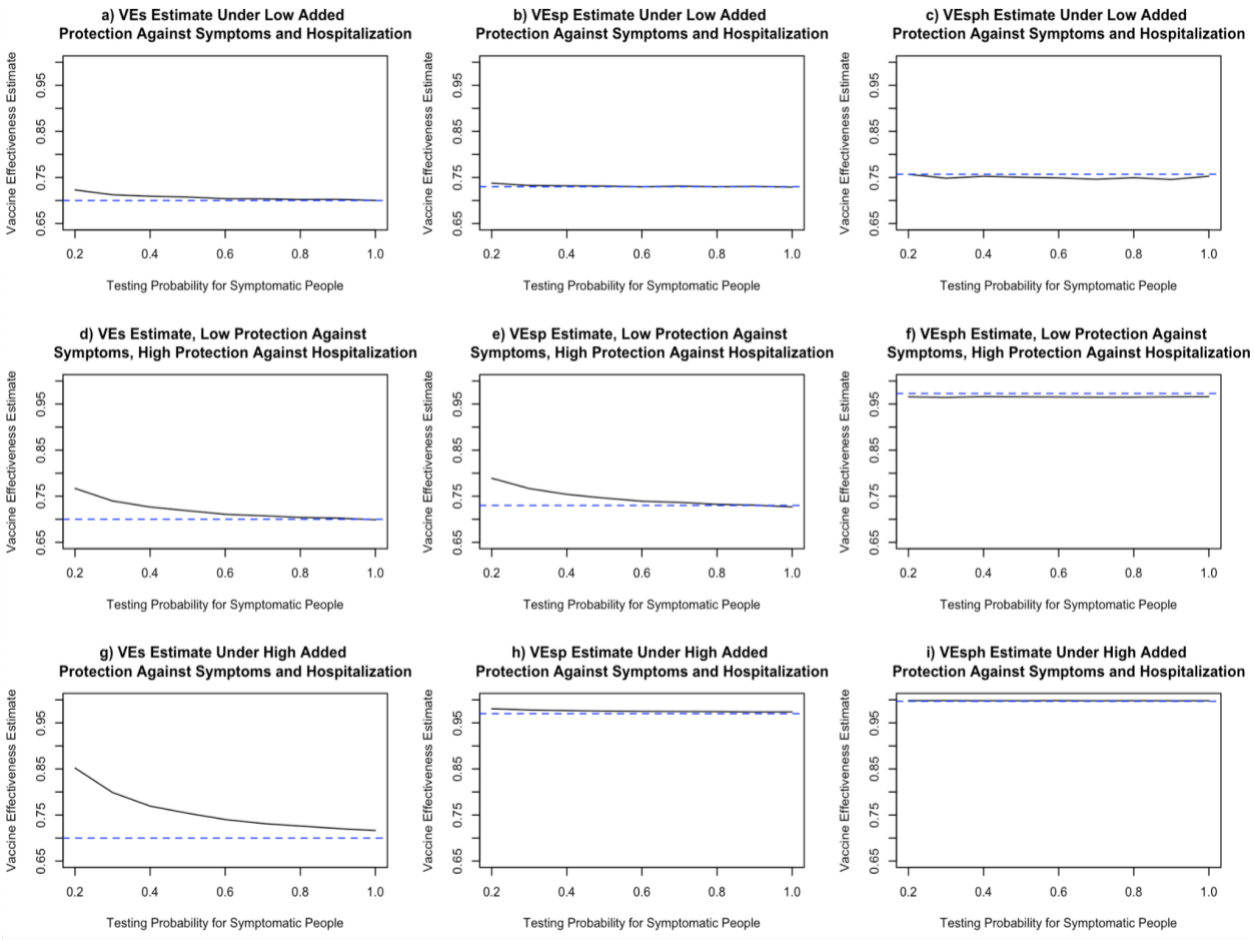

Figure 2. VE_S, VE_{SP}, and VE_{SPH} Estimates for Varying Testing Probability and Changing VE_P and VE_H. In all the plots, testing probability for hospitalized people is constant at 0.9. Testing probability for symptomatic people increases from 0.2 to 1, and testing probability for infected persons is always 0.1 lower, starting from 0.1 and ending at 0.9. Plot a) shows VE_S estimates under changing symptomatic and infected testing probabilities when VE_P and VE_H = 0.1. The true VE_S = 0.7 is the blue dotted horizontal line. Plot b) is the same, but for VE_{SP} estimates. The true VE_{SP} = 0.73. Plot c) looks at VE_{SPH}, and the true VE_{SPH} value is shown by the blue dotted line at 0.757. Plot d) is VE_S as symptomatic and infected testing probabilities change and VE_P = 0.1 and VE_H = 0.9. VE_S = 0.7. Plot e) is the same setting with VE_{SP}, where the true VE_{SP} value = 0.73. Plot f) then looks at VE_{SPH} and its true value, 0.973. Plot g) shows the effect of

changing symptomatic and infected testing probabilities on VE_S. True VE_S = 0.7. Plot h) shows the same for VE_{SP}. True VE_{SP} here = 0.97. Plot i) is the same for VE_{SPH}, and true VE_{SPH} = 0.997.

General patterns in this set of plots were similar to those of Figure 1. All VE_{SPH} estimates were constant over changing testing probabilities for symptomatic and infected people, with a flat line that closely matched the true VE_{SPH} values (Figures 2c, 2f, and 2i). When VE_P and VE_H were both set to be similarly low (at 0.1) or high (0.9), the plots were similarly flat and closely followed the true VE_{SP} values (Figures 2b and 2h). Figures 2a, 2d, 2e, and 2g all had similar patterns. When testing probabilities for symptomatic people and infected people were low, overestimation of VE_S (Figures 2a, 2d, and 2g) and VE_{SP} (Figure 2e) were high. The overestimation lessened and the estimate approached the true value as those testing probabilities approached 1.

Discussion

The simulation model ran five different settings to simulate bias in the test negative design when estimating vaccine effectiveness for COVID-19. The first setting was when the probability of testing was the same by disease severity and vaccination status. VE_S, VE_{SP}, and VE_{SPH} estimates were all close to the true value, as expected. Scenarios two, three, and four all had testing probabilities that were differential in some way by disease severity. These scenarios also had major overestimations of VE_S. It is likely due to an overrepresentation of hospitalized people when calculating effectiveness. That is, the symptomatic and infected populations have low testing probabilities, so a fewer percentage of those populations would show up in the test result data. In comparison, the testing population for

hospitalized people is much higher, and so more of that population (at least, more than their representative sample in the entire test-seeking population) would be found in the test result data. Furthermore, the vaccine had an additional protective effect ($VE_H = 0.2$) against hospitalization, which meant less people would be hospitalized and instead be infected or symptomatic – both groups with lower testing probabilities.

The final scenario had only VE_S as a non-zero amount, and this resulted in overall underestimation of the true VE values. This scenario, like the previous three, had differential testing probabilities by disease severity. However, since VE_P and VE_H were both zero, the impact of this was different than in the other settings. While testing probabilities for hospitalized persons were still the highest of the three, and still caused an over-representation of those cases in the testing data, the impact was lessened due to the lack of an additional protective effect provided by the vaccine (VE_H). This may have been overshadowed by a general underestimation that may be due to the size of this population and how much it accurately represents the total, unobserved, population. In general, the results obtained by test results may not be entirely representative of the test-seeking population.

The relationship between testing probability, VE_H, and all VE estimates were also further examined in the plots. Figure 1 focused on how changing testing probability and an over-representation of hospitalized people in the sample population can cause major overestimation when testing probabilities for the other groups are low, as shown in plots a) and d). Figure 2 also looked at changing testing probabilities for symptomatic people and infected people, but added changes in added protection to progression to symptoms and progression to severe disease by varying levels of VE_P and VE_H. These results, especially those of the first row (Plots 2a, 2b, and 2c), further emphasized the

findings of the final scenario previously tested. When VE_P and VE_H were very small, the impact of differential testing probabilities by disease severity is lessened greatly, and can be essentially negated. We also saw that when there is a big difference between VE_P and VE_H, the pattern previously seen in VE_S estimates showed up (Plot 2e): overestimation of VE_{SP} occurred when testing probabilities for symptomatic people and infected people were low compared to that of hospitalized people. This seems to indicate that differential testing probabilities by disease severity can cause bias in VE estimates, and that these differences are accentuated by large additional vaccine protection against different levels of disease severity.

There are a few limitations to this simulation study. In practice, some sources of bias may be present simultaneously, and the simulated population does not accurately reflect the complexity and nuance of a real population. Additionally, this simulation censors after the first SARS-CoV-2 infection, which does not allow us to see how bias can occur when multiple SARS-CoV-2 infections are possible but some are not observed. Furthermore, we have a few strong assumptions in this simulation. One of which is that the test for COVID is 100% sensitive and 100% specific. How could the addition of more false negatives bias VE? However, this simulation study is ultimately a good basis in which to understand general patterns between under-representation of case populations and the impact of differences in disease severity on VE bias in the test negative design.

Overall, this study indicates bias occurs when introducing different testing probabilities depending on disease severity, which is an effect that is made stronger by increased vaccine effectiveness against symptomatic disease or severe outcomes. VE against infection is most heavily affected, and very much overestimated. On the other hand, VE against symptomatic COVID-19 and against severe outcomes

like hospitalization are slightly underestimated. Going forward, more complexity could be added to the simulation, including the impact of other covariates on top of age. More importantly, a good next step would be to see how differences in testing probability by disease severity can be accounted for analytically when using the TND.

References

Alonso, P. L., Sacarlal, J., Aponte, J. J., Leach, A., Macete, E., Milman, J., Mandomando, I., Spiessens, B., Guinovart, C., Espasa, M., Bassat, Q., Aide, P., Ofori-Anyinam, O., Navia, M. M., Corachan, S., Ceuppens, M., Dubois, M. C., Demoitié, M. A., Dubovsky, F., ... Cohen, J. (2004). Efficacy of the RTS,S/AS02A vaccine against Plasmodium falciparum infection and disease in young African children: Randomised controlled trial. *Lancet*, *364*(9443), 1411–1420. https://doi.org/10.1016/S0140-6736(04)17223-1

Baden, L. R., el Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., Diemert, D., Spector, S. A., Rouphael, N., Creech, C. B., McGettigan, J., Khetan, S., Segall, N., Solis, J., Brosz, A., Fierro, C., Schwartz, H., Neuzil, K., Corey, L., ... Zaks, T. (2021). Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *New England Journal of Medicine*, *384*(5), 403–416. https://doi.org/10.1056/NEJMOA2035389/

Edwards, K., & Orenstein, W. (2023). *COVID-19: Vaccines - UpToDate*. https://www.uptodate.com/contents/covid-19-vaccines

FDA. (2020.). *Emergency Use Authorization for Vaccines Explained*. Retrieved March 21, 2023, from https://www.fda.gov/vaccines-blood-biologics/vaccines/emergency-use-authorization-vaccines-explained

Jackson, M. L., & Nelson, J. C. (2013). The test-negative design for estimating influenza vaccine effectiveness. *Vaccine*, *31*(17), 2165–2168. https://doi.org/10.1016/J.VACCINE.2013.02.053

Jain, V. K., Rivera, L., Zaman, K., Espos, R. A., Sirivichayakul, C., Quiambao, B. P., Rivera-Medina, D. M., Kerdpanich, P., Ceyhan, M., Dinleyici, E. C., Cravioto, A., Yunus, M., Chanthavanich, P., Limkittikul, K., Kurugol, Z., Alhan, E., Caplanusi, A., Durviaux, S., Boutet, P., … Innis, B. L. (2013). Vaccine for Prevention of Mild and Moderate-to-Severe Influenza in Children. *New England Journal of Medicine*, *369*(26), 2481–2491. https://doi.org/10.1056/NEJMOA1215817/

Kulkarni, P. S., Desai, S., Tewari, T., Kawade, A., Goyal, N., Garg, B. S., Kumar, D., Kanungo, S., Kamat, V., Kang, G., Bavdekar, A., Babji, S., Juvekar, S., Manna, B., Dutta, S., Angurana, R., Dewan, D., Dharmadhikari, A., Zade, J. K., … Raj, C. v. (2017). A randomized Phase III clinical trial to assess the efficacy of a bovine-human reassortant pentavalent rotavirus vaccine in Indian infants. *Vaccine*, *35*(45), 6228–6237. https://doi.org/10.1016/J.VACCINE.2017.09.014

Lau, H., Khosrawipour, T., Kocbach, P., Ichii, H., Bania, J., & Khosrawipour, V. (2021). Evaluating the massive underreporting and undertesting of COVID-19 cases in multiple global epicenters. *Pulmonology*, *27*(2), 110–115. https://doi.org/10.1016/J.PULMOE.2020.05.015

Lipsitch, M., & Dean, N. E. (2020). Understanding COVID-19 vaccine efficacy. *Science*, *370*(6518), 763–765. https://doi.org/10.1126/science.abe5938

Lopez Bernal, J., Andrews, N., Gower, C., Gallagher, E., Simmons, R., Thelwall, S., … Ramsay, M. (2021). Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *New England Journal of Medicine*, *385*(7), 585–594.

https://doi.org/10.1056/NEJMOA2108891/SUPPL_FILE/NEJMOA2108891_DISCLOSURES.PDF

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102. https://doi.org/10.1002/SIM.8086

Page, J., Hinshaw, D., & McKay, B. (2021). *In Hunt for Covid-19 Origin, Patient Zero Points to Second Wuhan Market - WSJ*. https://www.wsj.com/articles/in-hunt-for-covid-19-origin-patient-zero-points-to-second-wuhan-market-11614335404

Pereira, I. (2020). *US administers 1st doses of Pfizer coronavirus vaccine - ABC News*. https://abcnews.go.com/US/us-administer-1st-doses-pfizer-coronavirus-vaccine/story?id=74703018

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J. L., Pérez Marc, G., Moreira, E. D., Zerbini, C., Bailey, R., Swanson, K. A., Roychoudhury, S., Koury, K., Li, P., Kalina, W. v., Cooper, D., Frenck, R. W., Hammitt, L. L., … Gruber, W. C. (2020). Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *New England Journal of Medicine*, *383*(27), 2603–2615. https://doi.org/10.1056/NEJMOA2034577

Polinski, J. M., Weckstein, A. R., Batech, M., Kabelac, C., Kamath, T., Harvey, R., Jain, S., Rassen, J. A., Khan, N., & Schneeweiss, S. (2022). Durability of the Single-Dose Ad26.COV2.S Vaccine in the Prevention of COVID-19 Infections and Hospitalizations in the US Before and During the

Delta Variant Surge. *JAMA Network Open*, *5*(3), e222959–e222959. https://doi.org/10.1001/JAMANETWORKOPEN.2022.2959

Ranzani, O. T., Hitchings, M. D. T., Dorion, M., D'Agostini, T. L., De Paula, R. C., De Paula, O. F. P., ... Croda, J. (2021). Effectiveness of the CoronaVac vaccine in older adults during a gamma variant associated epidemic of covid-19 in Brazil: test negative case-control study. *BMJ*, *374*, 2015. https://doi.org/10.1136/BMJ.N2015

Sadoff, J., Gray, G., Vandebosch, A., Cárdenas, V., Shukarev, G., Grinsztejn, B., Goepfert, P. A., Truyers, C., Fennema, H., Spiessens, B., Offergeld, K., Scheper, G., Taylor, K. L., Robb, M. L., Treanor, J., Barouch, D. H., Stoddard, J., Ryser, M. F., Marovich, M. A., ... Douoguih, M. (2021). Safety and Efficacy of Single-Dose Ad26.COV2.S Vaccine against Covid-19. *New England Journal of Medicine*, *384*(23), 2187–2201. DOI: 10.1056/NEJMoa2101544

Schwartz, L. M., Halloran, M. E., Rowhani-Rahbar, A., Neuzil, K. M., & Victor, J. C. (2017). Rotavirus vaccine effectiveness in low-income settings: An evaluation of the test-negative design. *Vaccine*, *35*(1), 184. https://doi.org/10.1016/J.VACCINE.2016.10.077

Shircliff EJ, Rosenberg ES, Collens LM, et al. *Notes from the Field:* School-Based and Laboratory-Based Reporting of Positive COVID-19 Test Results Among School-Aged Children — New York, September 11, 2021–April 29, 2022. MMWR Morb Mortal Wkly Rep 2022;71:1029–1031. DOI: http://dx.doi.org/10.15585/mmwr.mm7132a2.

Tang, P., Hasan, M. R., Chemaitelly, H., Yassine, H. M., Benslimane, F. M., al Khatib, H. A., AlMukdad, S., Coyle, P., Ayoub, H. H., al Kanaani, Z., al Kuwari, E., Jeremijenko, A., Kaleeckal, A. H., Latif, A. N., Shaik, R. M., Abdul Rahim, H. F., Nasrallah, G. K., al Kuwari, M. G., al Romaihi, H. E., ... Abu-Raddad, L. J. (2021). BNT162b2 and mRNA-1273 COVID-19 vaccine effectiveness against the SARS-CoV-2 Delta variant in Qatar. *Nature Medicine 2021 27:12*, *27*(12), 2136–2143. https://doi.org/10.1038/s41591-021-01583-4

Trach, D. D., Clemens, J. D., Ke, N. T., Thuy, H. T., Son, N. D., Canh, D. G., Hang, P. v.d., & Rao, M. R. (1997). Field trial of a locally produced, killed, oral cholera vaccine in Vietnam. *Lancet*, *349*(9047), 231–235. https://doi.org/10.1016/S0140-6736(96)06107-7

Tsang, N. N. Y., So, H. C., Cowling, B. J., Leung, G. M., & Ip, D. K. M. (2022). Effectiveness of BNT162b2 and CoronaVac COVID-19 vaccination against asymptomatic and symptomatic infection of SARS-CoV-2 omicron BA.2 in Hong Kong: a prospective cohort study. *The Lancet Infectious Diseases*, *23*(4), 421–434. https://doi.org/10.1016/s1473-3099(22)00732-0

Vandenbroucke, J. P., Brickley, E. B., Vandenbroucke-Grauls, C. M. J. E., & Pearce, N. (2020). A test-negative design with additional population controls can be used to rapidly study causes of the SARS-COV-2 epidemic. *Epidemiology*, 836–843. https://doi.org/10.1097/EDE.0000000000001251

WHO. (2020). Laboratory testing for 2019 novel coronavirus (2019-nCoV) in suspected human cases. Retrieved November 25, 2022, from https://www.who.int/publications/i/item/10665-331501

WHO. (2021). *Vaccine efficacy, effectiveness and protection*. World Health Organization. https://www.who.int/news-room/feature-stories/detail/vaccine-efficacy-effectiveness-and-protection