Assessing Vaccine Effects on Infectiousness in COVID-19 Vaccine Trials

John Hocter

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Science

University of Washington

2021

Reading Committee:
Holly Janes, Chair
Elizabeth Brown

Program Authorized to Offer Degree:
Biostatistics

University of Washington

Abstract

Assessing Vaccine Effects on Infectiousness in COVID-19 Vaccine Trials

John Hocter

Chair of the Supervisory Committee:
Holly Janes

Vaccines have played an important role in the public health response to the COVID-19 pandemic. Clinical trials for several candidate vaccines demonstrated high efficacy at preventing severe / symptomatic disease. These vaccines have helped lower the burden of disease in the countries where they have been widely distributed (e.g., United States). The next generation of COVID-19 vaccine trials are designed to evaluate other vaccine effects (e.g., efficacy against infection). Understanding and quantifying how vaccines affect infectiousness is important for making sound public health policy, and analyzing such effects in a clinical trial requires selecting an appropriate endpoint and vaccine efficacy parameter. Many analyses of vaccine effects on infectiousness condition on infection (e.g., comparing infected vaccinated people to infected unvaccinated people), but conditional analyses have shortcomings. Most glaringly, they can lack power when efficacy against infection is high and may suffer from post-randomization selection bias. An alternative approach is to explore endpoints and vaccine efficacy parameters that do not condition on infection (i.e., unconditional analyses).

In this thesis, we define and discuss multiple methods for unconditional analyses of vaccine effects on infectiousness. As these methods commonly rely on proxies of infectiousness rather than directly observed secondary transmission, we also discuss several infectiousness proxies derived from measures of viral shedding. Finally, we use simulations to demonstrate the operating characteristics of a number of these methods and infectiousness proxies in a clinical trial whose design is based on an ongoing COVID-19 vaccine trial.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

DEDICATION

To my fiancée, Grace

Chapter 1

INTRODUCTION

Since its discovery in late 2019, the disease COVID-19, caused by the virus SARS-CoV-2, has spread around the world, killing more than 3.5M people as of June 2021. As a result of the rapid spread of the disease, high case-fatality and hospitalization rates, and lack of effective therapeutics, countries around the world enacted policies to slow the spread of the disease. COVID-19 morbidity varies substantially by age, with infection fatality rates (IFRs) ranging from 0.05% for 17-49 year-olds to 9% for people 65+[1]. Asymptomatic infection further complicates studies of COVID-19 – an estimated 30-40% of total infections with SARS-CoV-2 are asymptomatic, though again estimates vary quite widely as rates of asymptomatic infection may also depend on age and the presence of comorbidities [2, 3, 4, 5, 6]. While current research indicates asymptomatic individuals are slightly less infectious than symptomatic individuals [7, 8, 9], asymptomatic infection still occurs with moderate frequency.

Given the high burden of mortality, pressure placed on hospital systems around the globe, and resources required to track asymptomatic infection, the first generation of COVID-19 vaccine trials focused on demonstrating vaccine efficacy at preventing severe or symptomatic COVID-19. Several candidate vaccines demonstrated high efficacy at preventing such disease [10, 11, 12, 13]. Following approval from drug regulators, these vaccines were distributed around the world. Initial data from observational studies indicated that the vaccines may also be highly effective at preventing all infection [14, 15, 16]. However, the extent to which these vaccines affect secondary transmission is still unknown. Understanding the effect of vaccination on secondary transmission from is vital for public health policy (e.g., should vaccinated people be required to wear masks) and for encouraging vaccine uptake. Answering

this outstanding question is a priority in the next generation of COVID-19 vaccine trials.

Demonstrating a vaccine effect on secondary transmission requires selecting an appropriate vaccine efficacy parameter for analysis, as a wide range of types of vaccine effects have been parameterized and can be estimated in clinical trials. We go into detail on defining some of these vaccine effects in Chapter 4. Vaccine effects on secondary transmission are frequently assessed via the vaccine efficacy for infectiousness (VE_I) parameter. Importantly, VE_I assesses the vaccine effect *conditional on infection* (i.e., comparing the infectiousness of infected vaccinated people to infected unvaccinated people). As a result, VE_I may understate vaccine effects on secondary transmission in settings with high vaccine efficacy against acquisition (VE_S) where few vaccinated people are infected and thus most vaccinated people never have an opportunity to transmit the disease onward.

In order to better understand vaccine effects on transmission, the next generation of COVID-19 trials will assess VE_I by tracking close contacts of infected participants to identify secondary transmission events. However, such an approach is highly resource-intensive and, if the vaccine involved is even moderately effective at preventing acquisition, may lack power as a result of a low number of primary infection endpoints among vaccinated trial participants. Additionally, in the SARS-CoV-2 setting (unlike, say, the HIV setting), it can be difficult to identify the direction of transmission with certainty (i.e., it can be difficult to tell who infected who), further lowering the number of confirmed secondary transmission events from trial participants. In vaccine trials for other pathogens, researchers have instead (or additionally) assessed vaccine effects on infectiousness proxies (e.g., viral shedding). Human and animal models for many viral pathogens (e.g., SARS-CoV, influenza, HIV-1, HSV2, CMV) have shown strong correlations between viral shedding and infectiousness [17, 18, 19, 20, 21], which is often referred to as the dose-response model. Epidemiological and math modeling work focused on SARS-CoV-2 and transmission have also demonstrated a correlation between viral load (VL) and infectiousness[22, 17].

When using VL as a proxy for infectiousness, we frequently rely on summary measures of viral shedding related to the nature of infectiousness. A number of post-infection measures

of viral shedding that are functions of VL have been used in previous clinical trials, including peak VL, area under VL curve (AUC), and duration of viral shedding. We go further into detail on the definitions and rationale for each of these summary measures in Chapter 2. Choosing an appropriate infectiousness proxy endpoint is primarily a question of scientific interest (e.g., which endpoint correlates with the best understanding of the relationship between VL and infectiousness), though this decision may also involve statistical concerns (e.g., which endpoint has lower variability).

After settling on a proxy for infectiousness, we must also choose an appropriate statistical method for analyzing the vaccine effect on the proxy. We go into detail on these statistical methods in Chapters 5 through 7. As with VE_I, many methods for analyzing vaccine effects on infectiousness proxies condition on infection. However, post-randomization selection bias can affect such inference, and so causal inference methods are frequently employed. Several possibilities exist for unconditional analyses (e.g., that do not condition on infection), which can avoid issues with post-randomization selection bias, such as the chop-lump test [23], Lachenbruch two-part test [24], and other two-part tests [25]. These tests can be well-powered to detect vaccine effects on either acquisition or on a post-infection endpoint. However, they do not offer an estimate of the effect size (or corresponding confidence intervals).

1.1 Objective

In this paper we explore what we call the "transmission potential" endpoint, an associated vaccine efficacy parameter, and methods for estimation / inference based on the product method from Follmann et al [26]. Though we discuss it in the context of VL summary measures, this approach is generally applicable to proxies of infectiousness.

Chapter 2

VL SUMMARY MEASURES

SARS-CoV-2 viral RNA can be isolated from a variety of samples from infected patients (e.g., saliva, nasopharyngeal, fecal). These swabs are analyzed via quantitative PCR (qPCR), which (by way of a standardized sample) can be used to estimate the viral load (VL) (e.g., number of copies of viral RNA per mL of sample). We may represent VL on either the absolute or the log scale. Since SARS-SoV-2 is primarily transmitted through respiratory droplets, we most commonly use VL as measured in nasopharyngeal swabs. For chronic infections where viral load stabilizes after a period of acute infection (e.g., HIV), it is possible to utilize viral load as a proxy for infectiousness (e.g., VL set point for HIV)[27]. In the context of short-term infections where VL is highly variable over the course of infection, like with SARS-CoV-2, we instead tend to utilize a summary measure of VL over the course of infection. In this chapter, we define a number of VL summary measures that can be used and provide rationale for each. Where relevant, we also highlight previous uses of each summary measure in studies of infectiousness.

We consider these summary measures in the context of a randomized clinical trial, where some proportion of participants is assigned to receive the vaccine and the remainder is assigned not to receive the vaccine (e.g., receives placebo, delayed vaccination). Let Z be treatment assignment such that $Z = 0$ for unvaccinated participants and $Z = 1$ for vaccinated participants. Let $V_{zi}(s)$ be the viral load for subject i in group z at time since infection s.

2.1 Peak VL

Perhaps the most commonly used summary measure for analyzing VL in acute infections is the peak VL, which is the highest VL over the course of infection. The likelihood of

transmission has been shown to be highly correlated with the viral dose at exposure for a range of pathogens either directly or through epidemiological models (i.e., the more virus a susceptible person is exposed to, the more likely they are to be infected). Though viral dose at exposure depends on several factors (e.g., proximity during exposure, length of exposure), peak VL corresponds with the point at which onward transmission is most likely, controlling for these other factors. As a result, a higher peak VL generally corresponds with greater infectiousness. Such relationships have been studied in CMV [19, 18], HHV6 [19], and HIV[28]. Additionally, a dose-response relationship has been found in research in the transmission of SARS-CoV[29]. Vaccine effects on peak viral load have been assessed in clinical trials on a range of pathogens in both humans and animals, including HIV [30], simian HIV [31], and ebola [32].

Additionally, math modeling efforts analyzing SARS-CoV-2 transmission events have found an association between the periods of highest VL and the periods of highest transmission [22], and demonstrated that a vaccine which reduces peak VL would lead to a substantial reduction in total transmission [33]. Other work has indicated that the duration of infectiousness in SARS-CoV-2 infections is fairly short and corresponding with the time around peak VL [34]. Additionally, real-world data collected while vaccines were initially rolled out indicated at least one of the approved SARS-CoV-2 vaccines might lead to lower VL in infected patients[35].

Though general and SARS-CoV-2 specific evidence support peak VL as an appropriate proxy for infectiousness, it is not without its weaknesses. In particular, it fails to capture information about duration of periods of high VL, which can drastically impact the amount of transmission. For example, a person who returns from peak VL to no shedding very quickly is less likely to infect others than someone who continues to shed at 90% of their peak VL for several days. An analysis that focuses only on peak VL could lead to misleading conclusions about infectiousness. Futhermore, practical realities of clinical trials make it difficult to reliably capture peak VL, especially with pathogens such as SARS-CoV-2 where VL is believed to peak before symptom onset. Parametric approaches can be used to estimate

peak VL if it is not captured in the study period, but such model-based estimates can be difficult to validate.

Let peak VL be V_{zi}^P such that:

$$V_{zi}^P = \sup_{0 \leq s < \infty} \{V_{zi}(s)\}$$

2.2 Duration of shedding

Another commonly used VL summary measure is duration of shedding. Duration of shedding has a clear correlation with infectiousness – in order to infect others one must be shedding some amount of live pathogen, so the end of shedding corresponds with the end of any possibility of infectiousness. As with peak VL, duration of shedding has been used for a range of other infectious diseases, including influenza, [36, 37], HSV-2 [38], and MERS [34, 39, 40]. Vaccine effects on duration of shedding have also been assessed for several human and animal pathogens, including influenza [37], polio[41], porcine reproductive and respiratory syndrome virus (PRRSv) [42], and *E. coli* [43].

Because of the association between shedding and the possibility of infectiousness, duration of shedding has been well studied in SARS-CoV-2 infections[44, 34, 45, 39]. Duration of shedding was particularly important for determining appropriate public health policy during the SARS-CoV-2 pandemic, as understanding the duration of shedding was critical for setting isolation guidelines for infected people to minimize onward transmission. Understanding the vaccine effect on duration of shedding is critical for updating public policy to remain relevant in a post-vaccination world. For example, if a vaccinated person were to be infected, how long they should isolate to avoid infecting others. Estimating duration of shedding might require less frequent sampling than peak VL, though less frequent sampling also increases the impact of interval censoring on the estimation of duration of shedding.

Duration of shedding also has substantial limitations as a proxy for infectiousness. First and foremost, the qPCR tests that are used to detect viral shedding for SARS-CoV-2 infections detect viral RNA instead of live virus. Someone who is shedding viral RNA but not live

virus is not infectious, and thus looking at the duration of shedding estimated by qPCR may not be informative. Furthermore, even if tests were able to distinguish between shedding live virus and viral RNA, long periods of low levels of viral shedding may not correspond with infectiousness (e.g., if infected people need to be shedding at or above some level in order to be infectious). To combat this, some researchers have considered duration of shedding above a threshold V, which ideally would correspond with the VL threshold above which onward transmission is more likely.

Let duration of shedding be D_{zi} such that:

$$D_{zi} = \inf\{s : V_{zi}(r) = 0 \forall r > s\}$$

and let duration of shedding above viral threshold V be $D_{zi}(V)$ such that:

$$D_{zi}(V) = \inf\{s : V_{zi}(r) < V \forall r > s\}$$

2.3 Area under VL curve (AUC)

Yet another possible summary measure of VL is the area under the viral load curve (AUC). Heuristically, AUC measures the total amount of virus shed over the duration of infectiousness. This combines information about both the duration and magnitude of shedding, and as a result AUC combines some of the strengths of the previous two measures. Specifically, AUC is most heavily influenced by the periods of high viral shedding where infectiousness is the highest. However, unlike peak VL, AUC also captures information about the duration of the periods of high VL. Like duration of shedding, AUC is influenced by periods of low-level shedding that may not correspond with infectiousness, but these periods have less influence on the summary measure. AUC has been used to study viral dynamics in a variety of viral infections, including RSV[46], CMV[47], and HIV[48]. However, in the SARS-CoV-2 setting, AUC has high variability [49], and requires frequent surveillance sampling to be captured accurately.

Let AUC be AUC_{zi} such that:

$$AUC_{zi} = \int_0^\infty V_{zi}(s)ds$$

2.4 Truncated / threshold AUC

AUC can also be modified to limit some of the downsides related to capturing periods of low-shedding towards the end of infection. By truncating AUC at a point in time (e.g., AUC in first week of infection) and / or calculating AUC above some VL threshold V, we can capture information about the magnitude and duration of the periods of highest VL without capturing information about the less-infectious periods of lower shedding. This can be especially beneficial when such periods actually correspond with shedding viral RNA instead of live virus. However, such a summary measure is dependent on the threshold and / or the truncation time, which can be difficult to specify in advance without substantial knowledge about the relationship between VL and infectiousness.

Let r be the truncation time and V be the threshold VL. Then let AUC above VL threshold V truncated at time r be $AUC_{zi}(r, V)$ such that:

$$AUC_{zi}(r, V) = \int_0^r (V_{zi}(s) - V)ds$$

2.5 Transmission Potential

For each of the above summary measures, many analysis methods condition on infection (i.e., comparing VL summary measures between vaccinated and unvaccinated people who are infected). However, as mentioned previously, these conditional methods have shortcomings that can be particularly glaring in high efficacy against acquisition settings. For example, conditional analysis would not include the vaccine effect on viral shedding as a result of preventing infection, and thus may understate / misrepresent the overall vaccine effects. An unconditional endpoint would maximize power to detect vaccine efficacy in high efficacy settings since it includes both infected and uninfected participants, and would not be subject to post-randomization selection bias. Such an endpoint has been used in the study of HIV vaccines [26].

In Sections 3.2.1 and 3.2.1 we provide detailed definitions of the transmission potential endpoint for a generic infectiousness proxy, for which we could use any of the above VL

summaries.

2.6 Time to VL above threshold

When assessing vaccine effects related to time to events, the event of interest may be something other than infection (e.g., symptomatic / severe disease in first generation COVID-19 vaccine trials). These events are frequently related to disease progression or infectiousness. One such event which has been used in HIV vaccine research [50] is time to viral load crossing a VL threshold V. In HIV, where viral load is closely tied to disease progression and infectiousness, it is obvious why such an endpoint is of interest. Such an endpoint could also be of interest for SARS-CoV-2 patients for a variety of reasons. Suppose it is discovered that SARS-CoV-2 infected people are highly infectious when their VL exceeds some threshold V, and not very infectious otherwise. Or, suppose that VL exceeding V is an important clinical milestone. Then the vaccine effect on VL crossing V would clearly be an important indicator of its effect on infectiousness or disease progression. As with transmission potential, this endpoint could be analyzed unconditionally, and thus would not be subject to post-randomization selection bias.

There are some clear downsides to using such an endpoint. Specifically, analyses will depend on the threshold chosen, which, as in 2.4, may be difficult to specify in advance. Additionally, such an endpoint does not capture any information about VL after crossing the threshold (e.g., duration, overall magnitude). Finally, depending on the threshold, analysis of such an endpoint may have low power in trials powered for analysis of any infection. We carefully define time to VL above threshold in 3.2.2.

Chapter 3

DATA STRUCTURE AND ENDPOINT DEFINITIONS

3.1 Data Structure

As before, we consider data generated in the context of a randomized clinical trial, where some proportion of participants is assigned to receive the vaccine and the remainder is assigned not to receive the vaccine (e.g., receives placebo, delayed vaccination). Let Z be treatment assignment such that $Z = 0$ for unvaccinated participants and $Z = 1$ for vaccinated participants. Let there be n_z subjects in each treatment group. Let T_{zi} be the time to the event of interest. In much of this paper, the event of interest is infection. However, we will use different events of interest (e.g., severe / symptomatic disease, VL crossing threshold V) where noted. In this formulation, the timescale for T_{zi} can be either study time or calendar time. Assume that T_{zi} are independent, identically distributed within each group defined by Z and that the two groups are independent of one another. Let τ be the landmark time of interest.

Let $f_z(t) = lim_{\Delta t \to 0} \frac{P(t \leq T_{zi} < t+\Delta t)}{\Delta t}$ be the pdf of T_{zi} and $S_z(t) = P(T_{zi} > t) = \int_t^\infty f_z(t)$ be the survival function of T_{zi}. Let $h_z(t) = lim_{\Delta t \to 0} \frac{P(t \leq T_{zi} < t+\Delta t | T_{zi} \geq t)}{\Delta t} = \frac{f_z(t)}{S_z(t)}$ be the hazard function of T_{zi}. When noted, we assume a proportional hazards model ($h_z(t) = h_0(t)e^{\beta z}$). We also denote the cumulative hazard function for T_{zi} as $H_z(t) = \int_0^t h_z(s)ds$ and the cumulative incidence function for T_{zi} as $F_z(t) = \int_0^t f_z(s)ds$

Let $V_{zi}(s, T_{zi})$ be the viral load for subject i in group z at time since infection s. We assume that $V_{zi}(s, T_{zi})$ are independent, identically distributed within each group defined by Z. When noted, we will also assume that $V_{zi}(s, T_{zi})$ is independent of T_{zi} (i.e., viral load trajectories depend only on time since infection, not time of infection). When making this assumption we will simplify $V_{zi}(s, T_{zi})$ to $V_{zi}(s)$

3.2 Endpoint definitions

3.2.1 Transmission potential

We will now define the endpoint we refer to as the "transmission potential: endpoint. Heuristically, transmission potential represents the idea that uninfected individuals cannot transmit infection onto others, while infected individuals have some level of infectiousness that can be represented via a proxy. Our aim is to define instantaneous and cumulative versions of vaccine efficacy parameters for transmission potential (analogous to instantaneous and cumulative versions of vaccine efficacy for susceptibility, which we define in 4.2). In order to do that, we define instantaneous and cumulative versions of the transmission potential endpoint. The cumulative version of transmission potential represents the level of infectiousness a person has reached at any time before threshold time τ, while the instantaneous version of transmission potential aims to characterize the instantaneous level of infectiousness.

Cumulative transmission potential

Let X_{zi} be a post-infection proxy of infectiousness (e.g., one of the summary measures from 2). We assume that X_{zi} is independent of T_{zi} (i.e., infectiousness is independent of time of infection). We construct the cumulative version of the transmission potential endpoint so that, once a person is infected, their infectiousness proxy becomes constant. Let cumulative transmission potential observed at landmark time τ be $X_{zi}^{C*}(\tau)$ such that:

$$X_{zi}^{C*}(\tau) = \begin{cases} X_{zi} & T_{zi} \leq \tau \\ 0 & T_{zi} > \tau \end{cases} \tag{3.1}$$

In other words, the cumulative transmission potential endpoint is a piecewise constant function with a change point at the time of infection.

Instantaneous transmission potential

Defining instantaneous transmission potential requires a little more care to account for the nature of infectiousness. Let $X_{zi}(s, T_{zi})$ be the value of some post-infection proxy of infectiousness at time since infection s among an individual infected at time T_{zi}. Here we make the assumption that $X_{zi}(s, T_{zi})$ is independent of T_{zi} and we can simplify $X_{zi}(s, T_{zi})$ to $X_{zi}(s)$. We then define instantaneous transmission potential at time t for a person infected at time T_{zi} to be $X_{zi}^{I*}(t, T_{zi})$ such that:

$$X_{zi}^{I*}(t, T_{zi}) = \begin{cases} X_{zi}(t - T_{zi}) & T_{zi} \leq t \\ 0 & T_{zi} > t \end{cases} \tag{3.2}$$

Since T_{zi} is fixed for individual i from group z, we can also write $X_{zi}^{I*}(t, T_{zi})$ as $X_{zi}^{I*}(t)$. The instantaneous transmission potential endpoint lends itself more naturally to time-varying proxies for infectiousness (e.g., VL) rather than summary measures (e.g., VL AUC).

3.2.2 Time to VL above threshold

As mentioned in 2.6, in some cases the event of interest is not infection, but rather VL crossing some threshold V. Let $T_{zi}(V)$ represent the time until VL exceeds V, which we define to be:

$$T_{zi}(V) = T_{zi} + inf\{s : sup_{0 \leq r \leq s} V_{zi}(r) \geq V\} \tag{3.3}$$

Let $R_{zi}(V) = inf\{s : sup_{0 \leq r \leq s} V_{zi}(r) \geq V\}$ for future notation.

We will use the typical time-to-event functions (see 3.1) for $T_{zi}(V)$, with the addition that each function is now a function of both t and V, instead of only t. Let $S_z(t, V)$ be the survival function, $h_z(t, V)$ be the hazard, and $f_z(t, V)$ be the pdf for $T_{zi}(V)$. Also let $H_z(t, V)$ be the cumulative hazard function and $F_z(t, V)$ be the cumulative incidence function. When indicated, we assume a proportional hazards model (i.e., $h_z(t, V) = h_0(t, V)e^{\beta(V)z}$).

We have already assumed that T_{zi} is i.i.d. within groups. When we also assume that the distribution of $V_{zi}(s)$ is i.i.d. within groups, it is clear that $T_{zi}(V)$ is i.i.d. within groups.

We note that, in order for VL to cross some threshold V, a person must first become infected. So we can re-write the cumulative incidence function as:

$$\begin{aligned} F_z(t,V) &= P(T_{zi}(V) \leq t) = P(T_{zi} + R_{zi}(V) \leq t) \\ &= P(T_{zi} \leq t \cap R_{zi}(V) \leq t - T_{zi}) \\ &= P(T_{zi} \leq t) P(R_{zi}(V) \leq t - T_{zi} | T_{zi} \leq t) \\ &= F_z(t) P(R_{zi}(V) \leq t - T_{zi} | T_{zi} \leq t) \end{aligned} \tag{3.4}$$

Chapter 4

DEFINING VACCINE EFFICACY PARAMETERS

4.1 Introduction to vaccine effects

Vaccines may affect recipients in a number of ways, from preventing infection to lowering the severity of disease to shortening the duration of infection. When considering such effects, we may be interested in vaccine effects *conditional on infection* (e.g., does a vaccine prevent people from developing severe disease, given they have been infected) or unconditional on infection. So that we can assess these types of vaccine effects in a systematic way and conduct proper statistical inference, we must clearly define and label different vaccine effects. We start this chapter by introducing and defining a number of commonly used vaccine effect parameters, largely drawing on definitions from Halloran et al [51]. We utilize the data structure and definitions from Chapter 3. In general, these vaccine effects take the form:

$$VE = 1 - RR \tag{4.1}$$

where RR is some ratio of outcomes between vaccinated and unvaccinated groups. We will then define vaccine efficacy parameters for the transmission potential endpoints from Sections 3.2.1 and 3.2.1 and the time to VL above threshold V from section 3.2.2. These vaccine efficacy parameters will also take the form shown in Equation 4.1.

4.2 Vaccine Efficacy for Acquisition

We often want to assess whether a vaccine prevents people from being infected with a pathogen. This vaccine effect has been referred to be several names, including vaccine efficacy for acquisition, vaccine efficacy for susceptibility, and vaccine efficacy for infection. We opt for the former and denote it as VE_S.

As described in Halloran et al [51], VE_S can be parameterized via cumulative incidence or instantaneous hazard. These distinct parameterizations are commonly referred to as cumulative and instantaneous VE_S, respectively. We denote them as $VE_S^C(\tau)$ and $VE_S^I(t)$, where cumulative VE_S is evaluated at landmark time τ and instantaneous VE_S is evaluated at time t. For this parameter, our event of interest is infection (i.e., T_{zi} represents time to infection). Then the vaccine efficacy parameters are defined as:

$$VE_S^C(\tau) = 1 - \frac{F_1(\tau)}{F_0(\tau)} = 1 - \frac{1 - S_1(\tau)}{1 - S_0(\tau)}$$

$$VE_S^I(t) = 1 - \frac{h_1(t)}{h_0(t)}$$

Under the proportional hazards assumption, these reduce to:

$$VE_S^C(\tau) = 1 - \frac{1 - S_0(\tau)^{e^\beta}}{1 - S_0(\tau)}$$

$$VE_S^I(t) = 1 - e^\beta = VE_S^I$$

We briefly mention estimation techniques for $VE_S^C(\tau)$ and $VE_S^I(t)$ in 6.1.

4.3 Vaccine Efficacy for Disease

In other cases, especially for pathogens such as SARS-CoV-2 where asymptomatic infection is common, it may be of greater scientific interest to assess whether a vaccine prevents people from developing disease rather than only looking at whether the vaccine prevents infection. In fact, assessing such a vaccine effect was the primary goal of the first generation of COVID-19 vaccine trials. This type of vaccine effect is commonly called vaccine effect on disease, and is denoted as VE_D. As with VE_S, we can define both instantaneous and cumulative VE_D, which we denote as $VE_D^C(\tau)$ and $VE_D^I(t)$. In this case, our event of interest is development of severe / symptomatic disease (i.e., T_{zi} is the time to severe / symptomatic disease). Then we can define the two forms of VE_D to be:

$$VE_D^C(\tau) = 1 - \frac{F_1(\tau)}{F_0(\tau)} = 1 - \frac{1 - S_1(\tau)}{1 - S_0(\tau)}$$

$$VE_D^I(t) = 1 - \frac{h_1(t)}{h_0(t)}$$

Under the proportional hazards assumption, these reduce to:

$$VE_D^C(\tau) = 1 - \frac{1 - S_0(\tau)^{e^\beta}}{1 - S_0(\tau)}$$

$$VE_D^I(t) = 1 - e^\beta = VE_D^I$$

We briefly mention estimation techniques for $VE_D^C(\tau)$ and $VE_D^I(t)$ in 6.1.

4.4 Vaccine Efficacy for Infectiousness

We may also be interested in whether vaccines limit secondary transmission after infection. This type of vaccine effect is commonly called vaccine efficacy for infectiousness (VE_I), and it is conditional on infection. Let p_z be the transmission probability from a single exposure event between an infected person from group Z and a susceptible person. Then we can define VE_I in terms of this transmission probability:

$$VE_I = 1 - \frac{p_1}{p_0}$$

In practice, p_z is difficult to assess, as gathering information about individual exposure events is resource-intensive and exposure events are frequently heterogeneous in terms of size of exposure. For example, when considering SARS-CoV-2 exposure, an hour sitting indoors next to an infected person is a much larger exposure event than fifteen minutes standing outdoors more than 6 feet away from an infected person. So, it is common to define VE_I in terms of the number of infections per susceptible contact (susceptible person who meets exposure criteria), also called the secondary attack rate (SAR). Let SAR_z be the secondary attack rate in group z (i.e., the probability that a susceptible contact exposed to an infected person from group z is infected). Then we can define VE_I in terms of the secondary attack rate:

$$VE_I = 1 - \frac{SAR_1}{SAR_0}$$

4.5 Vaccine Efficacy for Transmission Potential

We now define both cumulative and instantaneous vaccine efficacy for transmission potential (VE_{TP}). Both take the form of ratios as shown in Equation 4.1.

4.5.1 Cumulative vaccine efficacy for transmission potential

Definition

Let cumulative vaccine efficacy for transmission potential evaluated at landmark time τ be $VE^C_{TP}(\tau)$ such that:

$$VE^C_{TP}(\tau) = 1 - \frac{E[X^{C*}_{zi}(\tau)|Z=1]}{E[X^{C*}_{zi}(\tau)|Z=0]} \tag{4.2}$$

For later use, we denote

$$\Delta^C_{TP}(\tau) = 1 - VE^C_{TP}(\tau)$$

and

$$\Phi^C_{TP}(\tau) = \log(\Delta^C_{TP}(\tau))$$

Simplification

With appropriate simplifying assumptions, it is possible to simplify $VE^C_{TP}(\tau)$ substantially. Even before making such assumptions, we can simplify $E[X^{C*}_{zi}(t)|Z=z]$ via the law of total probability.

$$E[X^{C*}_{zi}(\tau)|Z=z] = E_t[E[X^{C*}_{zi}(\tau)|Z=z, T_{zi}=t]]$$

$$= \int_0^\infty E[X^{C*}_{zi}(\tau)|Z=z, T_{zi}=t]P(T_{zi}=t|Z=z)dt$$

$$= \int_0^\infty E[X^{C*}_{zi}(\tau)|Z=z, T_{zi}=t]f_z(t)dt$$

$$= \int_0^\tau E[X^{C*}_{zi}(\tau)|Z=z, T_{zi}=t]f_z(t)dt + \int_\tau^\infty E[X^{C*}_{zi}(\tau)|Z=z, T_{zi}=t]f_z(t)dt$$

By equation 3.1, this becomes

$$= \int_0^\tau E[X_{zi}|Z=z, T_{zi}=t] f_z(t)dt + 0$$
$$= \int_0^\tau E[X_{zi}|Z=z, T_{zi}=t] \frac{f_z(t)}{F_z(\tau)} F_z(\tau)dt$$
$$= \int_0^\tau E[X_{zi}|Z=z, T_{zi}=t] f_z(t|T_{zi} \leq \tau) F_z(\tau)dt$$
$$= E[X_{zi}|Z=z, T_{zi} \leq \tau] F_z(\tau)$$

We can thus write $VE_{TP}^C(\tau)$ as:

$$VE_{TP}^C(\tau) = 1 - \frac{E[X_{zi}^{C*}(\tau)|Z=1]}{E[X_{zi}^{C*}(\tau)|Z=0]} = 1 - \frac{F_1(\tau)}{F_0(\tau)} \frac{E[X_{zi}|Z=1, T_{zi} \leq \tau]}{E[X_{zi}|Z=0, T_{zi} \leq \tau]} \tag{4.3}$$

Also note that

$$\Delta_{TP}^C(\tau) = \frac{F_1(\tau)}{F_0(\tau)} \frac{E[X_{zi}|Z=1, T_{zi} \leq \tau]}{E[X_{zi}|Z=0, T_{zi} \leq \tau]}$$

And

$$\Phi_{TP}^C(\tau) = \log(\frac{F_1(\tau)}{F_0(\tau)}) + \log(E[X_{zi}|Z=1, T_{zi} \leq \tau]) - \log(E[X_{zi}|Z=0, T_{zi} \leq \tau])$$

Expression in terms of other vaccine effects

Returning to $VE_{TP}^C(\tau)$, we can see that:

$$VE_{TP}^C(\tau) = 1 - \frac{F_1(\tau)}{F_0(\tau)} \frac{E[X_{zi}|Z=1, T_{zi} \leq \tau]}{E[X_{zi}|Z=0, T_{zi} \leq \tau]}$$
$$= 1 - (1 - VE_S^C(\tau))(1 - \frac{E[X_{zi}|Z=0, T_{zi} \leq \tau] - E[X_{zi}|Z=1, T_{zi} \leq \tau]}{E[X_{zi}|Z=0, T_{zi} \leq \tau]}) \tag{4.4}$$

If we denote the vaccine-induced reduction in mean X at time τ to be

$$\delta_X^C(\tau) = E[X_{zi}|Z=0, T_{zi} \leq \tau] - E[X_{zi}|Z=1, T_{zi} \leq \tau]$$

Then equation 4.4 becomes:

$$VE_{TP}^C(\tau) = 1 - (1 - VE_S^C(\tau))(1 - \frac{\delta_X^C(\tau)}{E[X_{zi}|Z=0, T_{zi} \leq \tau]}) \tag{4.5}$$

where it becomes clear that $VE^C_{TP}(\tau)$ depends on both $VE^C_S(\tau)$ and the proportional reduction in the mean infectiousness proxy, conditional on infection at / before τ.

4.5.2 Instantaneous vaccine efficacy for transmission potential

Definition

We denote instantaneous vaccine efficacy for transmission potential evaluated at time t as $VE^I_{TP}(t)$. Let $VE^I_{TP}(t)$ be:

$$VE^I_{TP}(t) = 1 - \frac{E[X^{I*}_{zi}(t)|Z=1]}{E[X^{I*}_{zi}(t)|Z=0]} \tag{4.6}$$

For later use, we denote

$$\Delta^I_{TP}(t) = 1 - VE^I_{TP}(t)$$

and

$$\Phi^I_{TP}(t) = \log(\Delta^I_{TP}(t))$$

Simplification

With appropriate simplifying assumptions, it is possible to simplify $VE^I_{TP}(t)$ substantially. Even before making such assumptions, we can simplify $E[X^{I*}_{zi}(t)|Z=z]$ via the law of total probability.

$$E[X^{I*}_{zi}(t)|Z=z] = E_r[E[X^{I*}_{zi}(t)|Z=z, T_{zi}=r]]$$
$$= \int_0^\infty E[X^{I*}_{zi}(t)|Z=z, T_{zi}=r]P(T_{zi}=r|Z=z)dr$$
$$= \int_0^\infty E[X^{I*}_{zi}(t)|Z=z, T_{zi}=r]f_z(r)dr$$
$$= \int_0^t E[X^{T*}_{zi}(t)|Z=z, T_{zi}=r]f_z(r)dr + \int_t^\infty E[X^{T*}_{zi}(t)|Z=z, T_{zi}=r]f_z(r)dr$$

By equation 3.2, this becomes

$$= \int_0^t E[X_{zi}(t-T_{zi})|Z=z, T_{zi}=r]f_z(r)dr + 0 \tag{4.7}$$

Which we can further simplify to:

$$= \int_0^t E[X_{zi}(t - T_{zi})|Z = z, T_{zi} = r] \frac{f_z(r)}{F_z(t)} F_z(t) dr$$

$$= \int_0^t E[X_{zi}(t - T_{zi})|Z = z, T_{zi} = r] f_z(r|T_{zi} \leq t) F_z(t) dr$$

$$= F_z(t) E[X_{zi}(t - T_{zi})|Z = z, T_{zi} \leq t]$$

Thus we could write $VE^I_{TP}(t)$:

$$VE^I_{TP}(t) = 1 - \frac{F_1(t)}{F_0(t)} \frac{E[X_{zi}(t - T_{zi})|Z = 1, T_{zi} \leq t]}{E[X_{zi}(t - T_{zi})|Z = 0, T_{zi} \leq t]}$$

We make the commonly used simplifying assumptions of proportional hazards (i.e., $h_z(t) = e^\beta h_0(t)$), and low incidence (i.e., $h_z(t) \approx f_z(t)$). Under these assumptions, $\frac{F_1(t)}{F_0(t)} \approx e^\beta$ We can then write this expression of $VE^I_{TP}(t)$ as:

$$VE^I_{TP}(t) = 1 - e^\beta \frac{E[X_{zi}(t - T_{zi})|Z = 1, T_{zi} \leq t]}{E[X_{zi}(t - T_{zi})|Z = 0, T_{zi} \leq t]} \tag{4.8}$$

4.6 Vaccine Efficacy for Crossing VL Threshold V

If we are interested in determining whether a vaccine has an effect on whether vaccinated people cross VL threshold V, we use crossing VL threshold V as the event of interest (see 2.6. We can define a vaccine efficacy parameter to describe the vaccines effect on crossing over such a threshold, which we call vaccine efficacy for viral load. We denote this parameter as $VE_{VL}(V)$. As with VE_S, VE_D, and VE_{TP}, we can define both cumulative and instantaneous versions of this parameter, which we denote ($VE^C_{VL}(\tau, V)$ and $VE^I_{VL}(t, V)$).

$$VE^C_{VL}(\tau, V) = 1 - \frac{F_1(\tau, V)}{F_0(\tau, V)} \tag{4.9}$$

$$VE^I_{VL}(t, V) = 1 - \frac{h_1(t, V)}{h_0(t, V)} \tag{4.10}$$

Note that, using 3.4, we can also write 4.9 as:

$$VE^C_{VL}(\tau, V) = 1 - \frac{F_1(\tau)}{F_1(\tau)} \frac{P(R_{zi}(V) \leq \tau - T_{zi} | T_{zi} \leq \tau, Z = 1)}{P(R_{zi}(V) \leq \tau - T_{zi} | T_{zi} \leq \tau, Z = 0)}$$
$$= 1 - (1 - VE^C_S(\tau)) \frac{P(R_{zi}(V) \leq \tau - T_{zi} | T_{zi} \leq \tau, Z = 1)}{P(R_{zi}(V) \leq \tau - T_{zi} | T_{zi} \leq \tau, Z = 0)} \quad (4.11)$$

where it becomes clear that $VE^C_{VL}(\tau, V)$ depends on both $VE^C_S(\tau)$ and the proportional reduction in the probability of people infected at / before time τ having VL cross VL threshold V at / before time τ.

Chapter 5

STATISTICAL ESTIMATION OF PROXIES OF INFECTIOUSNESS

In the real world, we are unable to directly observe the VL trajectory to calculate the summary measures from Chapter 2. Rather, we observe sampled viral load at a range of times post-infection, and must estimate the summary measures from our data. We frequently rely on empirical estimation techniques, which are straightforward but may ultimately estimate a slightly different VL parameter (e.g., see section 5.2.1). Alternatively, we may utilize model-based estimation strategies, in which VL samples are used to estimate a continuous VL curve with a parametric model. Summary measures are extracted from the estimated VL curve. Since our focus is primarily characterizing the operating characteristics of inference / estimation methods related to the vaccine efficacy parameters from Chapter 4, we will focus more on defining / discussing empirical estimation of VL summaries, and will touch only briefly on model-based estimation techniques, especially those used in our simulations (see Chapter 8).

5.1 Trial data

We have the same type of data for both empirical and model-based estimation. As a reminder, we denote the viral load for participant i in group z at time since infection s as $V_{zi}(s)$. We observe $V^*_{zi}(s^*_{zij})$, where $V^*_{zi}(s^*_{zij})$ is the sample VL from the sample taken at times $s^*_{zij}, j = 1, ..., m_{zi}$. Note that in COVID-19 vaccine trials VL sampling typically starts as a result of an initiating event, which may be symptom onset (as in the first generation trials) or a positive test initiated by a set diagnostic testing schedule (as in the second generation of trials). We thus refer to s^*_{zij} is time since first positive VL sample. We let d_{zi} represent the time from

infection to first positive VL sample, and so $s^*_{zij} + d_{zi}$ represents time since infection s_{zij}.

We note that in trials where participants are tested routinely to monitor for asymptomatic infection, we may collect and test VL samples before infection, in which case we would have negative s^*_{zij}. Since we would observe no viral shedding before infection, $V^*_{zi}(s^*_{zij}) = 0 \forall s^*_{zij} < 0$. Unless noted, we assume here that $s_{zij} > 0$.

Additionally, though Additionally, we assume that participants follow a fixed VL sampling schedule after diagnosis (e.g., sampled daily for 2 weeks), and we have thus the same number of samples for each participant. Thus we say $m_{zi} = m$ for all participants.

Finally, it is known that VL measurements from qPCR tests have some amount of measurement error. Combining all the above, we can decompose $V^*_{zi}(s^*_{zij})$ into:

$$V^*_{zi}(s^*_{zij}) = V_{zi}(s^*_{zij} + d_{zij}) + \epsilon_{zi}(s^*_{zij})$$

where $\epsilon_{zi}(s^*_{zij})$ is sampling error with $E[\epsilon_{zi}(s^*_{zij})] = 0$.

5.2 Empirical estimation

Each of the summary measures discussed in Chapter 2 can be estimated empirically. For empirical estimation, we base our estimates on $V^*_{zi}(s^*_{zij})$ and $s^*_{zij}, j = 1, ..., m$, and do not attempt to account for sampling error.

In order for conclusions based on empirical estimators to apply to the parameters they represent, we frequently need to assume that the vaccine does not affect time from infection to the start of sampling. In trials with fixed diagnostic testing schedules (e.g., all participants tested every other day), it is clear that this assumption is justified. However, in other trial designs (e.g., where testing is spurred by symptom onset), the vaccine might affect time from infection to the start of sampling and could lead to incorrect conclusions about the vaccine effect on the parameter of interest. For example, if a vaccine delayed symptom onset without affecting VL, infections would take longer to identify in the vaccinated group, and we would observe shorter duration of shedding, lower peak VLs, and lower AUCs among vaccinated participants, despite the vaccine actually having no effect on these parameters. The simula-

tions in Chapters 8 and 9 are based on COVID-19 vaccine trials with fixed diagnostic testing schedules.

5.2.1 Observed peak VL

The empirical estimator of peak VL is the observed peak VL, which we denote as V_{zi}^{P*} and define to be:

$$V_{zi}^{P*} = \max_{j=1,\ldots,m} \{V_{zi}^{*}(s_{zij}^{*})\}$$

It is well known that observed maxima are biased estimators of actual maxima, and that bias decreases as m increases. Frequent sampling, especially during the early stages of infection when VL is likely to peak, is necessary to extend conclusions based on V_{zi}^{P*} to V_{zi}^{P}.

5.2.2 Observed duration of shedding

The empirical estimator of duration of shedding is observed duration of shedding, which we denote as D_{zi}^{*} and define to be:

$$D_{zi}^{*} = \max\{s_{zij}^{*} : V_{zi}^{*}(s_{zij}^{*}) > 0\}$$

The assumption of no vaccine effect on time from infection to the start of sampling is crucial when using observed duration of shedding as an estimator of duration of shedding.

An additional complication with observed duration of shedding is that we are likely to encounter censored data. In COVID-19 vaccine trials, VL samples may be collected from infected participants until they are no longer shedding observable levels of viral RNA or until a fixed time after the first positive test (e.g., 14 days after first positive test). Additionally, some infected participants may be lost to follow-up. In any setting where infected participants are not followed until shedding is observed to have stopped, we have right-censored data. Furthermore, we do not know the exact time at which shedding stopped between samples, so we also have interval-censored data, though we may opt to ignore

interval censoring by using observed days of shedding or some similar measure. Methods for censored data are frequently used to compare duration of shedding in viral infections, but the estimation / inference methods in Chapters 6 and 7 would need to be updated to allow for censored data.

5.2.3 Observed AUC

The empirical estimator of AUC is observed AUC, which we denote AUC^*_{zi} and define to be:

$$AUC^*_{zi} = \sum_{j=1}^{m-1} \frac{V^*_{zi}(s^*_{zi(j+1)}) + V^*_{zi}(s^*_{zij})}{2}(s^*_{zi(j+1)} - s^*_{zij})$$

As with duration of shedding, the assumption of no vaccine effect on time from infection to the start of sampling is important when using observed AUC to draw conclusions about vaccine effects on VL. We may also have to deal with censored data as we do for observed duration of shedding.

5.2.4 Observed truncated / threshold AUC

We may also define the empirical estimator of the generalized AUC to allow for truncation and thresholds as in 2.4. We do not know the exact time of infection, so we define the truncation time r in terms of s^*_{zij} (i.e., AUC in first week after first positive test). Then our generalized AUC is:

$$AUC^*_{zi}(r, V) = \sum_{j:s^*_{zij} \leq r} \left(\frac{V^*_{zi}(s^*_{zi(j+1)}) + V^*_{zi}(s^*_{zij})}{2} - V \right) (s^*_{zi(j+1)} - s^*_{zij})$$

As with duration of shedding, the assumption of no vaccine effect on time from infection to the start of sampling is important when using observed AUC to draw conclusions about vaccine effects on VL.

5.3 Model-based estimation

As mentioned above, one may prefer model-based estimators for the summary measures in Chapter 2 to empirical estimators for a number of reasons. Model-based estimators may improve the precision of our estimates and can account for the bias described above (e.g., empirical estimates of peak are biased with finite sample size), allowing us to draw conclusions about the vaccine effect on the actual summary measure instead of the observed summary measure. Model-based estimators can also help us account for the delay in identifying infections, especially in trials where diagnostic testing is spurred by an initiating event (e.g., symptom onset). In these trials, we can use model-based estimators to remove (or relax) an assumption that vaccines have no effect on time from infection to diagnosis. However, model-based estimators can lead to invalid inference (e.g., inflated Type-1 error) and / or bias if the model is mis-specified, and as such may be difficult to employ for viruses where VL trajectories are irregular or are not completely understood.

A number of parametric models have been employed to estimate the VL curve from longitudinal VL samples. For example, Kissler et al used a parametric model to estimate continuous VL curves for individuals sampled during the 2020 NBA season restart[49]. Their model assumed SARS-CoV-2 viral trajectories included a proliferation phase, in which VL grows exponentially to a peak, followed by a clearance phase in which VL declines exponentially to 0. They used VL samples to estimate the length of proliferation, length of clearance, and peak VL, which they turned into a "triangular" log-VL curve. They used these curves to estimate duration of shedding and peak log-VL, and they could also be used to estimate AUC.

Again, the primary focus of this paper is the operating characteristics of inference / estimation methods for vaccine efficacy parameters, so we do not go into further detail on model-based estimation methods.

Chapter 6

STATISTICAL ESTIMATION OF VACCINE EFFICACY PARAMETERS

In this chapter, we describe estimation and inference methods for the vaccine efficacy parameters from Chapter 4. Rather than observing time to event T_{zi} for all individuals, some will be censored at time C_{zi} either due to loss-to-follow-up or administrative censoring. Then, rather than observing T_{zi}, we observe $(\tilde{T}_{zi}, \delta_{zi})$ where $\tilde{T}_{zi} = \min(T_{zi}, C_{zi})$ and $\delta_{zi} = I(\tilde{T}_{zi} = T_{zi})$. In Section 6.2, the event of interest is infection, while in Section 6.3, the event of interest is crossing VL threshold V.

6.1 Estimating common vaccine efficacy parameters

The methods for estimating commonly used vaccine efficacy parameters such as VE_S, VE_D, and VE_I are well-known, and we do not comment on them at length here. These methods include Cox proportional hazards models to estimate $VE_S^I(t)$ and $VE_D^I(t)$ and nonparametric estimates of the cumulative incidence ratio (via Nelson-Aalen estimates of cumulative hazard) to estimate $VE_S^C(\tau)$ and $VE_D^C(\tau)$. We use the latter method to conduct inference on $VE_S^C(\tau)$ in simulations in Chapters 8-9.

6.2 Estimating vaccine efficacy for transmission potential

6.2.1 Estimating cumulative vaccine efficacy for transmission potential

Let X_{zi} be the infectiousness proxy we choose for our cumulative transmission potential endpoint (Section 3.1). For example, X_{zi} could be any of the empirical estimators of the VL summary measures we defined in Chapter 5. Suppose we observe X_{zi} for all infected individuals (i.e., no censoring) and that X_{zi} is constant immediately at the time of infection.

Let $I_z(\tau) = \sum_i^{n_z} I[\tilde{T}_{zi} \leq \tau \cap \delta_{zi} = 1]$ be the number of infections in group Z at or before landmark time τ.

As a reminder, in Equation 4.3, we expressed $VE_{TP}^C(\tau)$ in terms of $\frac{F_1(\tau)}{F_0(\tau)}$, $E[X|Z = 1, T_{zi} \leq \tau]$, and $E[X|Z = 0, T_{zi} \leq \tau]$. We return to the notation:

$$\Phi_{TP}^C(t) = \log(1 - VE_{TP}^C(t))$$

We can thus write:

$$\Phi_{TP}^C(\tau) = \log(\frac{F_1(\tau)}{F_0(\tau)}) + \log(E[X_{zi}|Z = 1, T_{zi} \leq \tau]) - \log(E[X_{zi}|Z = 0, T_{zi} \leq \tau])$$

We can use plug-in estimators for each of the three parts above to estimate $\Phi_{TP}^C(\tau)$ (and thus $VE_{TP}^C(\tau)$).

We can estimate $F_z(\tau)$ by way of the Nelson-Aalen estimator of cumulative hazard:

$$\hat{F}_z(\tau) = 1 - \exp(-\hat{H}_z(\tau))$$

where

$$\hat{H}_z(\tau) = \sum_{t \leq \tau} \frac{d_z(t)}{r_z(t)}$$

where $d_z(t)$ is the number of events in group z at time t and $r_z(t)$ is the number at risk in group z at time t. Under standard assumptions (e.g., non-informative censoring) that are satisfied in our setting, this estimator is consistent and asymptotically normal[52]. Its variance can be estimated via

$$\hat{var}(\hat{F}_z(\tau)) = \exp(-2\hat{H}_z(\tau)) * \sum_{t \leq \tau} \frac{d_z(t)}{r_z(t)^2}$$

Using this estimator for $F_z(\tau)$ for $Z = 0, 1$, we can estimate $\frac{F_1(\tau)}{F_0(\tau)}$ using $\frac{\hat{F}_1(\tau)}{\hat{F}_0(\tau)}$ (and thus $\log(\frac{F_1(\tau)}{F_0(\tau)})$ using $\log(\frac{\hat{F}_1(\tau)}{\hat{F}_0(\tau)})$). Since we have assumed T_{zi} is independent between the two groups, $\hat{F}_1(\tau)$ and $\hat{F}_0(\tau)$ are independent. We can thus use the delta method to estimate the

variance of $\log(\frac{\hat{F}_1(\tau)}{\hat{F}_0(\tau)})$ to be:

$$\hat{var}(\log(\frac{\hat{F}_1(\tau)}{\hat{F}_0(\tau)})) = \frac{\hat{var}(\hat{F}_1(\tau))}{\hat{F}_1(\tau)^2} + \frac{\hat{var}(\hat{F}_0(\tau))}{\hat{F}_0(\tau)^2}$$

We can also estimate $E[X_{zi}|Z = z, T_{zi} \leq \tau]$ using a plug-in estimator – the sample mean of X_{zi} among observed infections at or before τ, which we will label as $\bar{X}_z(\tau)$.

$$\bar{X}_z(\tau) = \frac{\sum_i X_{zi} I[\tilde{T}_{zi} \leq \tau \cap \delta_{zi} = 1]}{I_z(\tau)} \tag{6.1}$$

Let $s_z(\tau)$ denote the standard deviation of X_{zi} among observed infections at or before τ. By the Central Limit Theorem $\bar{X}_1(\tau)$ and $\bar{X}_0(\tau)$ are asymptotically normal, and so are $\log(\bar{X}_1(\tau))$ and $\log(\bar{X}_0(\tau))$. Via the delta method, we estimate the variance of $\log(\bar{X}_z(\tau))$ with:

$$\hat{var}(\log(\bar{X}_z(\tau))) = \frac{s_z(\tau)^2}{\bar{X}_z(\tau)^2 I_z(\tau)}$$

We can combine these plug-in estimators to estimate $\Phi^C_{TP}(\tau)$

$$\hat{\Phi}^C_{TP}(\tau) = \log(\frac{\hat{F}_1(\tau)}{\hat{F}_0(\tau)}) + \log(\bar{X}_1(\tau)) - \log(\bar{X}_0(\tau))$$

With the simplifying assumption that X_{zi} is independent of T_{zi} (e.g., peak viral load is independent of infection time), and again using the assumption that the two groups are independent, we can estimate the variance of $\hat{\Phi}^C_{TP}(\tau)$ to be

$$\hat{var}(\hat{\Phi}^C_{TP}(\tau)) = \hat{var}(\log(\frac{\hat{F}_1(\tau)}{\hat{F}_0(\tau)})) + \hat{var}(\log(\bar{X}_1(\tau))) + \hat{var}(\log(\bar{X}_0(\tau)))$$

This estimator is asymptotically normal. In large-sample settings we can create Wald confidence intervals for and conduct inference on $\Phi^C_{TP}(\tau)$, then exponentiate to get valid confidence intervals / inference for $\hat{VE}^C_{TP}(\tau)$

6.2.2 Estimating instantaneous vaccine efficacy for transmission potential

Possible estimation approaches for $VE^I_{TP}(t)$

In Section 4.5.2, with several simplifying assumptions (low, constant incidence; proportional hazards), we were able to express $VE^I_{TP}(t)$ in terms of $E[X_{zi}(t - T_{zi})|Z = z, T_{zi} \leq t]$ and β. We could estimate $VE^I_{TP}(t)$ using plug-in estimators for the pieces of Equation 4.8 as we did for $VE^C_{TP}(\tau)$. This approach has the benefit of requiring fewer simplifying assumptions. However, since many people infected at / before t would no longer be infectious by time t (e.g., they are no longer shedding virus), the distribution of $X_{zi}(t - T_{zi})$ among people infected at / before t would be zero-inflated. This could lead to low power / precision.

An alternative that could limit the effect of zero-inflation would be to use a 2-part model approach. Several different 2-part models could be used; we describe one in which we separate $X_{zi}(t - T_{zi})$ among people infected at / before t into those who are no longer infectious and those who are still infectious. Let S_{zi} represent the duration of infectiousness after infection (note that this differs from the duration of shedding from 2.2), which we define to be:

$$S_{zi} = \inf\{s : s \geq 0 \cap X_{zi}(r) = 0 \forall r \geq s\}$$

We can then re-write the instantaneous transmission potential endpoint shown in Definition 3.2 such that:

$$X^{I*}_{zi}(t) = \begin{cases} X_{zi}(t - T_{zi}) & T_{zi} \leq t \cap S_{zi} \geq t - T_{zi} \\ 0 & T_{zi} > t \cup S_{zi} < t - T_{zi} \end{cases} \tag{6.2}$$

Using this definition, we see that $X_{zi}(t - T_{zi}) = 0$ if $S_{zi} < (t - T_{zi})$, so we we can rewrite 4.8 as:

$$VE^I_{TP}(t) = 1 - e^{\beta} \frac{E[X_{zi}(t - T_{zi})|Z = 1, T_{zi} \leq t, S_{zi} \geq (t - T_{zi})]}{E[X_{zi}(t - T_{zi})|Z = 0, T_{zi} \leq t, S_{zi} \geq (t - T_{zi})]} \frac{P(S_{zi} \geq (t - T_{zi})|Z = 1, T_{zi} \leq t)}{P(S_{zi} \geq (t - T_{zi})|Z = 0, T_{zi} \leq t)} \tag{6.3}$$

We could then estimate $VE^I_{TP}(t)$ using plug-in estimators for each of the component parts of Equation 6.3. This approach would limit the effect of zero-inflation, but introduces complexity into the estimation method as it requires estimating additional parameters ($P(S_{zi} \geq (t - T_{zi})|Z = z, T_{zi} \leq t)$), which could lower power / precision.

Finally, we could make the strong simplifying assumptions of constant incidence and no vaccine effect on the distribution of duration of infectiousness (i.e., $P(S_{zi} = s|Z = 1) = P(S_{zi} = s|Z = 0) = P(S_{zi} = s) \forall s$). Then $\frac{P(S_{zi} \geq (t-T_{zi})|Z=1, T_{zi} \leq t)}{P(S_{zi} \geq (t-T_{zi})|Z=0, T_{zi} \leq t)} = 1$. We can then re-write Equation 6.3 as:

$$VE^I_{TP}(t) = 1 - e^{\beta} \frac{E[X_{zi}(t - T_{zi})|Z = 1, T_{zi} \leq t, S_{zi} \geq (t - T_{zi})]}{E[X_{zi}(t - T_{zi})|Z = 0, T_{zi} \leq t, S_{zi} \geq (t - T_{zi})]} \tag{6.4}$$

This approach could yield improved power by limiting the zero-inflation and removing the need to estimate an additional component. However, the strong assumptions required to yield this form may not be valid.

For simplicity, we will only describe the estimation approach based on Equation 6.4 under these strong assumptions. Estimation approaches for the forms in Equations 4.8 and 6.3 would be very similar.

Two-step estimation approach under strong assumptions

Let $X_{zi}(s)$ be the instantaneous infectiousness proxy at time since infection s we choose for our instantaneous transmission potential endpoint (Section 3.2). We observe $X_{zi}(t - Tzi)$ over the duration of infectiousness for all infected individuals (i.e., no censoring). For infected patients, we also observe an indicator of continued infectiousness at time t $\gamma_{zi}(t) = I[\exists r \geq t \text{ s.t. } X_{zi}(r - T_{zi}) > 0]$. Let $I_z(t) = \sum_i^{n_z} I[\tilde{T}_{zi} \leq t \cap \delta_{zi} = 1 \cap \gamma_{zi} = 1]$ be the number of currently infectious participants in group z at time t.

With some minor tweaks to account for duration of infectiousness, the product method developed by Follmann and Huang [26] can be used to estimate $VE^I_{TP}(t)$ under the strong assumptions noted above. We return to the notation:

$$\Phi^I_{TP}(t) = \log(1 - VE^I_{TP}(t))$$

Now, we have that under these assumptions:

$$\Phi^I_{TP}(t) = \beta + \log(E[X_{zi}(t - T_{zi})|Z = 1, T_{zi} = t, S_{zi} \geq (t - T_{zi})]) - \log(E[X_{zi}(t - T_{zi})|Z = 0, T_{zi} = t, S_{zi} \geq (t - T_{zi})]) \quad (6.5)$$

We can estimate β using $\hat{\beta}$ from a Cox proportional hazards model. We will also use the resulting estimate of the variance of $\hat{\beta}$, $\hat{var}(\hat{\beta})$.

The plug-in estimator $\bar{X}_Z(t)$ could be used for $E[X_{zi}(t - T_{zi})|Z = z, T_{zi} = t, S_{zi} \geq (t - T_{zi})]$, where in this setting we restrict the sample average to infected participants who are still infectious at time t:

$$\bar{X}_z(t) = \frac{\sum_i^{n_z} X_{zi}(t - \tilde{T}_{zi}) I[\tilde{T}_{zi} \leq t \cap \delta_{zi} = 1 \cap \gamma_{zi} = 1]}{I_z(t)} \quad (6.6)$$

Let $s_z(t)$ denote the standard deviation of $X_{zi}(t - T_{zi})$ among infectious individuals at time t. By the Central Limit Theorem $\bar{X}_0(t)$ and $\bar{X}_1(t)$ are asymptotically normal, and so are $\log(\bar{X}_1(t))$ and $\log(\bar{X}_0(t))$. Via the delta method, we estimate the variance of $\log(\bar{X}_z(t))$ with:

$$\hat{var}(\log(\bar{X}_z(t))) = \frac{s_z(t)^2}{\bar{X}_z(t)^2 I_z(t)}$$

We can put these estimators together to estimate $\Phi^I_{TP}(t)$

$$\hat{\Phi}^I_{TP}(t) = \hat{\beta} + \log(\bar{X}_1(t)) - \log(\bar{X}_0(t))$$

With the simplifying assumption that $X_{zi}(t - T_{zi})$ is independent of T_{zi} (e.g., viral load is independent of infection time), and again using the assumption that the two groups are independent, we can estimate the variance of $\hat{\Phi}^I_{TP}(t)$ to be

$$\hat{var}(\hat{\Phi}^I_{TP}(t)) = \hat{var}(\hat{\beta}) + \hat{var}(\log(\bar{X}_1(t))) + \hat{var}(\log(\bar{X}_0(t)))$$

This estimator is asymptotically normal. In large-sample settings we can create Wald confidence intervals for and conduct inference on $\Phi^I_{TP}(t)$, then exponentiate to get valid confidence intervals / inference for $\hat{VE^I_{TP}}(t)$

However, given the relatively short duration of infectiousness in a SARS-CoV-2 infections compared to the length of the trial, many (even most) individuals infected before time t would no longer be infectious at time t. Thus the number of individuals contributing to estimating $E[X_{zi}(t - T_{zi})|Z = 1, T_{zi} \leq t, S_{zi} \geq (t - T_{zi})]$ would be fairly small relative to the total number of infections. With additional assumptions about the type of vaccine effect (e.g., constant over the duration of infectiousness), it might be possible to use the ratio of cumulative infectiousness proxies measures‘ instead of instantaneous infectiousness proxies.

Using this method to estimate $VE^I_{TP}(\tau)$ requires more assumptions about the structure of the data and the type of vaccine effect than using $VE^C_{TP}(\tau)$. As a result, we will focus our simulations in Chapters 8 and 9 on the cumulative transmission potential endpoint.

6.3 Estimating vaccine efficacy for crossing VL threshold V

6.3.1 Methods for single threshold V

As a reminder, for $VE_{VL}(V)$, the endpoint of interest is crossing a VL threshold V, which occurs at time $T_{zi}(V)$. In a trial setting, again, some individuals will be censored at time C_{zi}, either due to loss-to-follow-up or administrative censoring. Specifically, patients who are infected but do not cross VL threshold V are censored at the end of post-infection follow-up. Rather than observing $T_{zi}(V)$, we observe $(\tilde{T}_{zi}(V), \delta_{zi}(V))$ where $\tilde{T}_{zi}(V)$ =min$(T_{zi}(V), C_{zi})$ and $\delta_{zi}(V)) = I(\tilde{T}_{zi}(V) = T_{zi}(V))$.

When estimating $VE^C_{VL}(\tau, V)$ or $VE^I_{VL}(t, V)$, we may specify a specific threshold based on scientific knowledge. For example, if it were discovered that a person is only infectious when VL exceeds V^*_{thresh}, we might specifically estimate and conduct inference on $VE^C_{VL}(\tau, V^*_{thresh})$

or $VE^I_{VL}(t, V^*_{thresh})$. For a fixed threshold, we would use the same estimation / inference methods as we do for VE_S and VE_D (see Section 6.1). Again, we do not discuss these commonly used methods in detail.

6.3.2 Simultaneous estimation / inference for unknown V

In the absence of such clear scientific knowledge, however, it may be difficult to decide upon a threshold *a priori*. We may instead choose to specify multiple discrete thresholds or a continuous range of thresholds and identify any vaccine effects across these threshold options simultaneously. Rather than creating individual confidence intervals / running multiple hypothesis tests, Gilbert and Sun [50] developed methods to simultaneously estimate and conduct inference on vaccine effects on time to crossing a VL threshold. Their method allows for valid simultaneous inference / confidence intervals either across a continuous range of VL thresholds (e.g., $V \in [V_{low}, V_{high}]$) or a discrete set of thresholds (e.g., $V \in \{V_k, k = 1, ..., m\}$). We will discuss here the key steps / conclusion in this simultaneous inference / confidence interval method. We refer readers to the original paper for greater detail.

We start with a number of definitions of functions common to time-to-event endpoints. Let $N_{zi}(t, V) = I[\tilde{T}_{zi}(V \leq t \cap \delta_{zi}(V) = 1]$ and $N_z(t, V) = \sum_i^{n_z} N_{zi}(t, V)$. Let $R_{zi}(t, V) = I[\tilde{T}_{zi}(V \geq t]$ and $R_z(t, V) = \sum_i^{n_z} R_{zi}(t, V)$. Additionally, let $M_{zi}(t, V) = N_{zi}(t, V) - \int_0^t R_{zi}(s, V) dH_z(s, V) ds$ and $r_z(t, V) = P[\tilde{T}_{zi}(V) \geq t]$. Finally, let $\rho_z = \lim_{n\to\infty} \frac{n_z}{n}$.

We will use the Kaplan-Meier estimator of $S_z(t, V)$ and Nelson-Aalen estimators of $h_z(t, V)$ and $H_z(t, V)$, which we denote as $\hat{S}_z(t, V)$, $\hat{h}_z(t, V)$, and $\hat{H}_z(t, V)$, respectively. We also use the Kaplan-Meier estimator of $F_z(t, V)$, $\hat{F}_z(t, V) = 1 - \hat{S}_z(t, V)$. Finally, we use $d\hat{M}_{zi}(s, V) = N_{zi}(t, V) - \int_0^t R_{zi}(s, V)\hat{h}_{zi}(s, V) ds$. The following is a martingale representation of $\hat{F}_z(t, V)$:

$$\sqrt{n_z}(F_z(\tau, V) - \hat{F}_z(\tau, V)) = S_z(\tau, V) \int_0^\tau \frac{n_z^{-\frac{1}{2}} \sum_{i=1}^{n_z} dM_{zi}(s, V)}{r_z(s, V)} + o_p(1) \qquad (6.7)$$

which converges in distribution to a mean-zero random variable with variance

$$\sigma_z^2(\tau, V) = S_z^2(\tau, V) \int_0^\tau \frac{dH_z(s, V)}{r_z(s, V)}$$

This variance can be consistently estimated with:

$$\hat{\sigma}_z^2(\tau, V) = n_z \hat{S}_z^2(\tau, V) \int_0^\tau \frac{d\hat{H}_z(s, V)}{R_z(s, V)}$$

Let $U(\tau, V) = \sqrt{n}\left(\frac{\hat{F}_1(\tau,V)}{\hat{F}_0(\tau,V)} - \frac{F_1(\tau,V)}{F_0(\tau,V)}\right)$. We can write:

$$\begin{aligned} U(\tau, V) &= \sqrt{n}\left(\frac{\hat{F}_1(\tau, V)}{\hat{F}_0(\tau, V)} - \frac{F_1(\tau, V)}{F_0(\tau, V)}\right) \\ &= \sqrt{n}\left(\frac{1}{F_0(\tau, V)}\left(\hat{F}_1(\tau, V) - F_1(\tau, V)\right) - \frac{F_1(\tau, V)}{(F_0(\tau, V))^2}\left(\hat{F}_0(\tau, V) - F_0(\tau, V)\right)\right) + o_p(1) \end{aligned} \tag{6.8}$$

By the central limit theorem, $U(\tau, V)$ converges in distribution to a mean-zero normal random variable with variance

$$\sigma^2(\tau, V) = \rho_1^{-1}(F_0(\tau, V))^2 \sigma_1^2(\tau, V) + \rho_0^{-1}(F_1(\tau, V))^2 (F_0(\tau, V))^{-4} \sigma_0^2(\tau, V)$$

This variance can be consistently estimated with:

$$\hat{\sigma}^2(\tau, V) = \frac{n}{n_1}(\hat{F}_0(\tau, V))^2 \hat{\sigma}_1^2(\tau, V) + \frac{n}{n_0}(\hat{F}_1(\tau, V))^2 (\hat{F}_0(\tau, V))^{-4} \hat{\sigma}_0^2(\tau, V)$$

We could use this variance to generate confidence intervals for $VE_{VL}(\tau, V) = 1 - \frac{F_1(\tau,V)}{F_0(\tau,V}$ for a single V. However, to make simultaneous confidence intervals across a range of possible values for V, additional steps are needed.

From 6.7 and 6.8, we can see that

$$\begin{aligned} U(\tau, V) = &\sqrt{\frac{n}{n_1}}\left(\frac{1}{F_0(\tau, V)} S_1(\tau, V) \int_0^\tau \frac{n_1^{-\frac{1}{2}} \sum_{i=1}^{n_1} dM_{1i}(s, V)}{r_1(s, V)}\right) \\ &- \sqrt{\frac{n}{n_0}}\left(\frac{F_1(\tau, V)}{(F_0(\tau, V))^2} S_0(\tau, V) \int_0^\tau \frac{n_0^{-\frac{1}{2}} \sum_{i=1}^{n_0} dM_{0i}(s, V)}{r_0(s, V)}\right) + o_p(1) \end{aligned} \tag{6.9}$$

Gilbert and Sun showed that 6.9 converges weakly to a mean-zero Gaussian process for $V \in [V_{low}, V_{high}]$ (as well as for $V \in \{V_k, k = 1, ..., m\}$). Let Z_{zi}, $i = 1, ..., n_z$ be i.i.d. standard normal random variables.

Then we can define $U^*(\tau, V)$:

$$U^*(\tau, V) = \sqrt{\frac{n}{n_1}} \left(\frac{1}{\hat{F}_0(\tau, V)} \hat{S}_1(\tau, V) \int_0^\tau \frac{n_1^{\frac{1}{2}} \sum_{i=1}^{n_1} Z_{1i} d\hat{M}_{1i}(s, V)}{R_1(s, V)} \right) - \sqrt{\frac{n}{n_0}} \left(\frac{\hat{F}_1(\tau, V)}{(\hat{F}_0(\tau, V))^2} \hat{S}_0(\tau, V) \int_0^\tau \frac{n_0^{\frac{1}{2}} \sum_{i=1}^{n_0} Z_{0i} d\hat{M}_{0i}(s, V)}{R_0(s, V)} \right) \quad (6.10)$$

Gilbert and Sun showed that, conditional on the observed data, $U^*(\tau, V)$ converges weakly to the same process as $U(\tau, V)$. In order to generate simultaneous confidence bands for $VE_V^C(\tau, V)$ for $V \in [V_{low}, V_{high}]$, we use the quantity $c_{\alpha/2}$, which is the asymptotic $1 - \alpha$ quantile of

$$sup_{V_{low} \leq V \leq V_{high}} \left| \frac{U(\tau, V)}{\hat{\sigma}(\tau, V)} \right|$$

We can consistently estimate $c_{\alpha/2}$ by repeatedly generating copies of i.i.d. standard normal random variables Z_{zi}, $i = 1, ..., n_z$ then calculating $U_b^*(\tau, V)$ for each repetition $b, b = 1, ..., B$. We then extract the $1 - \alpha$ quantile of

$$sup_{V_{low} \leq V \leq V_{high}} \left| \frac{U_b^*(\tau, V)}{\hat{\sigma}(\tau, V)} \right|, b = 1, .., B$$

which we denote as $\hat{c}_{\alpha/2}$. We can then generate uniform $100(1 - \alpha)\%$ confidence bands for $VE_{VL}(\tau, V), V \in [V_{low}, V_{high}]$ with:

$$\hat{VE}_{VL}(\tau, V) \pm n^{-\frac{1}{2}} \hat{c}_{\alpha/2} \hat{\sigma}(\tau, V)$$

.

Chapter 7

COVARIATE-ADJUSTED VACCINE EFFICACY FOR TRANSMISSION POTENTIAL

In vaccine trials, it is possible (and in fact, likely) that some covariates are associated with the event and / or outcome of interest. This includes covariates impacting the outcome that are equally distributed between treatment groups (i.e., precision variables) or covariates associated with the outcome that differ in distribution between treatment groups (i.e., confounding variables). We would like to be able to control for both of these classes of covariates in the context of VE_{TP}, as is done with other vaccine effects. For example, in a multi-site vaccine trial, placebo-group incidence rates frequently vary substantially by site, and so vaccine efficacy parameters are adjusted to account for site (e.g., including site as a covariate in a Cox proportional hazards model). In the context of transmission potential, covariates could be associated with differences in both incidence and infectiousness proxy values, and so we need to extend our VE_{TP} parameter to allow us to adjust for both. In this chapter, we define cumulative vaccine efficacy for transmission potential adjusted for covariates W, which we denote $(VE^{C}_{TP,W}(\tau)$. We also show how this parameter can be estimated when W is discrete. In Chapters 8 and 9, we will show that this estimation method yields valid inference.

7.1 Definition

Suppose $W_{zi} = W_{zi1}, ..., W_{zik}$ is a set of k variables that are associated with infection and / or the infectiousness proxy included in the transmission potential endpoint, conditional on infection. Let $F_{W|Z}(W_{zi})$ denote the distribution function of W conditional on treatment assignment Z. We assume that W_{zi} is independent of Z (as in a randomized trial), and

thus $F_{W|Z}(W_{zi}) = F_W(W_{zi})$. This distribution of W can be specified based on data from an external source or a known population distribution or the distribution of W in the study. Let W-adjusted $VE^C_{TP}(\tau)$ be $VE^C_{TP,W}(\tau)$, which we define to be:

$$VE^C_{TP,W}(\tau) = 1 - \frac{\int_W E[X^*_{zi}(\tau)|Z=1, W_{zi}=w]dF_W(w)}{\int_W E[X^*_{zi}(\tau)|Z=0, W_{zi}=w]dF_W(w)}$$

By the same approach as Section 4.5.1, it is clear that $VE^C_{TP,W}(\tau)$ can be re-written as:

$$VE^C_{TP,W}(\tau) = 1 - \frac{\int_W E[X_{zi}|Z=1, W_{zi}=w, T_{zi}\leq\tau]F_1(\tau|W_{zi}=w)dF_W(w)}{\int_W E[X_{zi}|Z=0, W_{zi}=w, T_{zi}\leq\tau]F_0(\tau|W_{zi}=w)dF_W(w)} \tag{7.1}$$

As with $VE^C_{TP}(\tau)$, we denote

$$\Delta^C_{TP,W}(\tau) = 1 - VE^C_{TP,W}(\tau)$$

and

$$\Phi^C_{TP,W}(\tau) = \log(\Delta^C_{TP,W}(\tau))$$

7.2 Estimating $VE^C_{TP,W}(\tau)$ with discrete W

$VE^C_{TP,W}(\tau)$ could be estimated in a number of different ways – we focus on the case where W is discrete and leave remaining cases for future work.

A discrete W, combined with Z, divides the population into distinct strata; we assume that observations are independent, identically distributed within each stratum defined by W, Z. If W is discrete, we can write Equation 7.1 as:

$$VE^C_{TP,W}(\tau) = 1 - \frac{\sum_W E[X_{zi}|Z=1, W_{zi}=w, T_{zi}\leq\tau]F_1(\tau|W_{zi}=w)P(W_{zi}=w)}{\sum_W E[X_{zi}|Z=0, W_{zi}=w, T_{zi}\leq\tau]F_0(\tau|W_{zi}=w)P(W_{zi}=w)} \tag{7.2}$$

We can use use plug-in estimators within each stratum to estimate the contribution of each stratum to the sum in the numerator and denominator of Equation 7.2. We can then combine these estimates to get an estimate of $VE^C_{TP,W}(\tau)$ (by way of an estimate of

$\Phi^C_{TP,W}(\tau)$). Estimating $VE^C_{TP,W}(\tau)$ in the case where at least element of W is continuous (Equation 7.1) is more complicated, and we leave that for future work.

The distribution of W may be estimated from study data or provided as a constant. For example, if W was birth sex, we could specify a 50-50 male-female split when estimating $VE^C_{TP,W}(\tau)$ to estimate a sex-pooled $VE^C_{TP,W}(\tau)$. However, if W represents some other characteristic with an unknown distribution in the population, we can estimate $F_W(W)$ from the sample. We will provide estimators and associated variance for both options.

Let $\xi_{Z,W}(\tau) = E[X_{zi}|Z = z, W_{zi} = w, T_{zi} \leq \tau]F_z(\tau|W_{zi} = w)P(W_{zi} = w)$ and let $\sum_W \xi_{Z,W}(\tau) = \xi_Z(\tau)$. Then:

$$VE^C_{TP,W}(\tau) = 1 - \frac{\sum_W \xi_{1,W}(\tau)}{\sum_W \xi_{0,W}(\tau)} = 1 - \frac{\xi_1(\tau)}{\xi_0(\tau)}$$

and thus:

$$\Phi^C_{TP,W}(\tau) = \log(\xi_1(\tau)) - \log(\xi_0(\tau))$$

We can estimate $E[X_{zi}|Z = z, W_{zi} = w, T_{zi} \leq \tau]$ using $\bar{X}_{Z,W}(\tau)$, the sample mean of X_{zi} among observed infections at / before τ in the stratum defined by $(W_{zi} = w, Z = z)$. Let $I_{Z,W}(\tau) = \sum_i^{n_z} I[\tilde{T}_{zi} \leq \tau \cap \delta_{zi} = 1 \cap W_{zi} = w]$ be the number of infections in the stratum defined by $(W_{zi} = w, Z = z)$ at or before time τ. Then

$$\bar{X}_{Z,W}(\tau) = \frac{\sum_i X_{zi} I[\tilde{T}_{zi} \leq \tau \cap \delta_{zi} = 1 \cap W_{zi} = w]}{I_{Z,W}(\tau)}$$

$\bar{X}_{Z,W}(\tau)$ has estimated variance $\frac{S^2_{Z,W}}{I_{Z,W}}$, where $S_{Z,W}$ is the sample standard deviation of X_{zi} among individuals infected at or before time τ in the stratum defined by $(W_{zi} = w, Z = z)$.

Let $d_{z,w}(t)$ be the number of events and $rz, w(t)$ be the number at risk in the in the stratum defined by $(W_{zi} = w, Z = z)$ at time t. We can then estimate $F_Z(\tau|W_{zi} = w)$ by way of stratified Nelson-Aalen estimators of cumulative hazard.

$$\hat{F}_Z(\tau|W_{zi} = w) = 1 - \exp(-\hat{H}_z(\tau|W_{zi} = w))$$

where

$$\hat{H}_z(\tau|W_{zi}=w)=\sum_{t\leq\tau}\frac{d_{z,w}(t)}{r_{z,w}(t)}$$

This estimator is consistent and asymptotically normal[52]. Its variance can be estimated via

$$\hat{var}(\hat{F}_Z(\tau|W_{zi}=w))=\exp(-2\hat{H}_z(\tau|W_{zi}=w))*\sum_{t\leq\tau}\frac{d_{z,w}(t)}{r_{z,w}(t)^2}$$

When the distribution of W is provided (so $P(W_{zi}=w)$ is a known constant), we can estimate $\xi_{Z,W}(\tau)$ with

$$\hat{\xi}_{Z,W}(\tau)=\bar{X}_{Z,W}(\tau)\hat{F}_Z(\tau|W_{zi}=w)P(W_{zi}=w)$$

which, via the delta method, has estimated variance:

$$\hat{var}(\hat{\xi}_{Z,W}(\tau))=P(W_{zi}=w)^2\left(\bar{X}^2_{Z,W}\hat{var}(\hat{F}_Z(\tau|W_{zi}=w))+\hat{F}_Z(\tau|W_{zi}=w)^2\frac{S^2_{Z,W}}{I_{Z,W}}\right)$$

We can add up the strata-specific estimators to estimate $\xi_Z(\tau)$ with $\hat{\xi}_Z(\tau)$

$$\hat{\xi}_Z(\tau)=\sum_W\hat{\xi}_{Z,W}(\tau)$$

Since the observations from different strata are independent, we can estimate the variance of $\hat{\xi}_Z(\tau)$ to be:

$$\hat{var}(\hat{\xi}_Z(\tau))=\sum_W\hat{var}(\hat{\xi}_{Z,W}(\tau))$$

We can then estimate $\Phi^C_{TP,W}(\tau)$ with $\hat{\Phi}^C_{TP,W}(\tau)$

$$\hat{\Phi}^C_{TP,W}(\tau)=\log(\hat{\xi}_1(\tau))-\log(\hat{\xi}_0(\tau)) \tag{7.3}$$

which has variance (via the delta method):

$$\hat{var}(\hat{\Phi}^C_{TP,W}(\tau)) = \frac{\hat{var}(\hat{\xi}_1(\tau))}{(\hat{\xi}_1(\tau))^2} + \frac{\hat{var}(\hat{\xi}_0(\tau))}{(\hat{\xi}_0(\tau))^2} \tag{7.4}$$

In large-sample settings we can use 7.3 and 7.4 to create Wald confidence intervals for and conduct inference on $\Phi^C_{TP,W}(\tau)$, then exponentiate to get valid confidence intervals / inference for $\hat{VE}^C_{TP,W}(\tau)$.

If the distribution of W is not specified (so we must estimate $P(W_{zi} = w)$), we can employ a very similar approach. Let $p_W(w) = P(W_{zi} = w)$. We can estimate $p_W(w)$ with the standard binomial estimator:

$$\hat{p}_W(w) = \frac{\sum_i I[W_{zi} = w]}{n}$$

which has estimated variance

$$\hat{var}(\hat{p}_W(w)) = \frac{\hat{p}_W(w)(1 - \hat{p}_W(w))}{n}$$

We can then estimate $\xi_{Z,W}(\tau)$ with $\hat{\xi}_{Z,W}(\tau)$ where:

$$\hat{\xi}_{Z,W}(\tau) = \bar{X}_{Z,W}(\tau)\hat{F}_Z(\tau|W_{zi} = w)\hat{p}_W(w)$$

which, via the delta method, has estimated variance:

$$\begin{aligned}\hat{var}(\hat{\xi}_{Z,W}(\tau)) = \bar{X}^2_{Z,W}\hat{p}_W(w)^2\hat{var}(\hat{F}_Z(\tau|W_{zi} = w)) + \\ \hat{F}_Z(\tau|W_{zi} = w)^2\hat{p}_W(w)^2\frac{S^2_{Z,W}}{I_{Z,W}} + \bar{X}^2_{Z,W}\hat{F}_Z(\tau|W_{zi} = w)^2\hat{var}(\hat{p}_W(w))\end{aligned} \tag{7.5}$$

We can then estimate $\xi_Z(\tau)$ with $\hat{\xi}_Z(\tau)$ where:

$$\hat{\xi}_Z(\tau) = \sum_W \hat{\xi}_{Z,W}(\tau)$$

Again, since our observations are independent between strata, we can estimate the variance of $\hat{\xi}_Z(\tau)$ to be:

$$\hat{var}(\hat{\xi}_Z(\tau)) = \sum_W \hat{var}(\hat{\xi}_{Z,W}(\tau))$$

We can then estimate $\Phi^C_{TP,W}(\tau)$ with $\hat{\Phi}^C_{TP,W}(\tau)$

$$\hat{\Phi}^C_{TP,W}(\tau) = \log(\hat{\xi}_1(\tau)) - \log(\hat{\xi}_0(\tau)) \tag{7.6}$$

which has variance (via the delta method):

$$\hat{var}(\hat{\Phi}^C_{TP,W}(\tau)) = \frac{\hat{var}(\hat{\xi}_1(\tau))}{(\hat{\xi}_1(\tau))^2} + \frac{\hat{var}(\hat{\xi}_0(\tau))}{(\hat{\xi}_0(\tau))^2} \tag{7.7}$$

In large-sample settings we can use 7.6 and 7.7 to create Wald confidence intervals for and conduct inference on $\Phi^C_{TP,W}(\tau)$, then exponentiate to get valid confidence intervals / inference for $\hat{VE}^C_{TP,W}(\tau)$.

Chapter 8

SIMULATION METHODOLOGY

We utilized a series of simulations to assess the operating characteristics of the above methods for estimating and conducting inference on $VE^C_{TP}(\tau)$ and $VE^C_{TP,W}(\tau)$. Finally, we explored the simultaneous confidence interval / inference methods for $VE^C_{VL}(\tau, V)$ across a range of V. These simulations are based on the initial design of an ongoing trial to assess the efficacy of SARS-CoV-2 vaccines in college students. The primary objectives of the trial are to demonstrate vaccine efficacy against acquisition (VE_S) and vaccine effect on peak viral load conditional on infection; vaccine effects on other post-infection outcomes are secondary endpoints. In reality, the trial protocol has been updated, though the changes are not reflected in the simulations. Additionally, though we refer to vaccine and placebo group throughout this section, the placebo group is actually deferred vaccination group, since the vaccine used in the study has already been demonstrated to be effective.

8.1 Trial Design

We suppose a total trial enrollment of 12,000, randomly assigned in a 1:1 ratio between the vaccine and placebo groups. Participants are followed for 16 weeks after enrollment (at which time all placebo group participants are to be vaccinated), so we consider participants who have not been infected by 16 weeks to be censored. To account for the fact that some participants who enroll may actually have been previously infected with SARS-CoV-2, we simulate that 10% of enrolled participants are seropositive at enrollment, based on data from previous trials. Since participants who were seropositive are identified only after enrollment, such participants are excluded from the primary analysis but included in the total trial enrollment. Furthermore, we suppose that 2% of enrolled participants do not receive their

second dose. Those participants are also excluded from the analysis (i.e., the analysis is per-protocol).

8.2 SARS-CoV-2 acquisition

For the remaining participants, we simulate infection times (T_{zi}) using the exponential distribution. We define our endpoints in terms of time since enrollment and use threshold time $\tau = 16$ weeks. We assume a 4% 16-week incidence in the placebo group, and in the vaccine we use a 4% * (1 - $VE_S^C(\tau)$) incidence, with $VE_S^C(\tau)$ varied. We also simulate exponential censoring times (C_{zi}) based on 5% loss-to-follow-up over the length of the study. Remaining participants are censored at 16 weeks of follow-up. From T_{zi} and C_{zi}, we generate survival data ($\tilde{T}_{zi}, \delta_{zi}$). In this trial design, all infections that occur after the administration of the second vaccine dose count towards the primary analysis (i.e., we only count events occurring after 4 weeks of follow-up).

In the simulations to demonstrate the techniques for estimating $VE_{TP,W}^C(\tau)$ from Chapter 7, we used a similar design and simulation approach. These simulations additionally assumed enrollment occurred at two sites where baseline incidence could differ between the sites. Additionally, these simulations assumed 50:50 male-to-female enrollment overall and allowed for baseline incidence to differ by sex. We also allowed for imbalances in sex and enrollment site by treatment assignment.

8.3 VL trajectories

We assumed participants were tested for SARS-CoV-2 infection daily, allowing for infections to be captured before symptom onset. For participants infected with SARS-CoV-2 during study follow-up, we simulated viral load trajectories using an approach adapted from the model-based approach employed by Kissler et al in the analysis of data from the 2020 NBA restart [49] (discussed in 5.3).

Specifically, for each infected person we simulated a peak Ct value, as well as length of proliferation and clearance stages. We used the posterior distributions of each of the three

parameters from Kissler et al's model to simulate data. We simulated duration of proliferation stage fro a gamma distribution with shape parameter 2.3 and inverse scale parameter 0.7 and duration of clearance stage from a gamma distribution with shape parameter 2.4 and inverse scale parameter 0.4. We assumed that the vaccine had no effect on the distribution of the duration of proliferation or on the distribution of the duration of clearance (e.g., scale, shape parameters remain the same).

We simulated peak log-VLs from followed a normal distribution (with a mean of 5.5 and sd of 1.8 in the placebo group and truncated at the qPCR limit of detection of 40Ct). We parameterized the vaccine effect on viral shedding as a reduction in peak log-VL, which we denote as Δ peak log-VL. For example, Δ peak log-VL = 1 indicates that the mean peak log-VL in the vaccine group is one unit lower than in the placebo group, after truncation. We assumed no vaccine effect on the variability of peak log-VL. In the trials to evaluate $VE^{C}_{TP,W}(\tau)$, we allowed mean placebo-group peak log-VL to differ by sex (i.e., sex could be a precision variable or confounder).

After simulating VL trajectories using this "triangle model", we simulated sampling from the trajectories. We assumed diagnosis occurred within 1 day of infection (given daily diagnostic testing schedule), and sampling occurred every 24 hours thereafter for 2 weeks. We simulated measurement error consistent with typical values for qPCR when generating our samples. Measurement error was assumed to be additive with a mean of 0 and a standard deviation of 0.2 Ct units (corresponding to <0.1 $\log_{10}$ units on the log-VL scales). This level of measurement error falls within operating guidelines for qPCR [53], and this amount of error (or less) has been used in other simulation studies [54]. The level of measurement error has some effect on the absolute results of the simulation studies, but does not affect the relative results / trends shown in Chapter 9.

We then calculated observed peak log-VL and observed log-VL AUC using the equations from Sections 5.2.1 and 5.2.3, respectively. We simulated trials using both observed peak log-VL and observed log-VL AUC as the proxy for infectiousness for our transmission potential endpoint. We also used sampled VL trajectories to calculate $T_{zi}(V)$ for $V \in \{4, 5, 6\}$ so that

we could evaluate the simultaneous confidence interval technique for $VE^C_{VL}(\tau, V)$ across a range of V.

8.4 Inference

After using the above methods to generate simulated data, we estimated $VE^C_S(\tau)$, $VE^C_{TP,W}(\tau)$, $VE^C_S(\tau)$, and $VE_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$ using the techniques from Chapters 6 and 7. We constructed 95% confidence intervals for each of the above parameters in addition to simultaneous 95% confidence interval bands for $VE_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$. Finally, we conducted one-sided, $\alpha = 0.025$ hypothesis tests against a null hypothesis of $VE \leq 0$ for each of $VE^C_S(\tau)$, $VE^C_{TP,W}(\tau)$, $VE^C_S(\tau)$, and $VE_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$. We also used one-sided, $\alpha = 0.025$, two-sample t-tests to compare both peak log-VL and log-VL AUC between vaccine and placebo groups so that we could compare the performance of these conditional methods to our unconditional approach. We simulated 1,000 trials under each set of parameters to get reliable estimates of power and 95% confidence interval coverage probability for each test / confidence interval calculated.

Chapter 9

SIMULATION RESULTS

We present selected results from the simulations described in Chapter 8. Additional results are contained in Appendix A. Unless otherwise noted, all power results discussed are the result of one-sided, $\alpha = 0.025$ tests against a null hypothesis of $VE \leq 0$ and all confidence intervals displayed are 95% confidence intervals.

9.1 Validation and performance of estimation methods for $VE^C_{TP}(\tau)$

9.1.1 Validation

Before assessing the power of inference methods for $VE^C_{TP}(\tau)$, we validated that the estimation approach generated unbiased estimates, standard errors, and appropriate confidence interval coverage for $VE^C_{TP}(\tau)$.

Unbiased estimation

We expect $\hat{\Phi}^C_{TP}(\tau)$ to be an unbiased estimator of $\Phi^C_{TP}(\tau)$. Thus, the geometric mean of our estimates $\hat{VE}^C_{TP}(\tau)$ should equal to $VE^C_{TP}(\tau)$. Figure 9.1 shows the geometric mean of our estimates of $\hat{VE}^C_{TP}(\tau)$ from the 1,000 simulated trials across a range of $VE^C_{TP}(\tau)$ values using both peak log-VL and log-VL AUC as the transmission proxy. The dotted line corresponds with unbiased estimation. We see that the estimator provides unbiased estimates across this range of $VE^C_{TP}(\tau)$ values for both choices of transmission proxy.

Valid variance estimates

In order for inference to be valid, we also need $\hat{var}(\hat{\Phi}^C_{TP}(\tau))$ to consistently estimate $var(\hat{\Phi}^C_{TP}(\tau))$. Figure 9.2 shows the mean $\hat{var}(\hat{\Phi}^C_{TP}(\tau))$ from the 1,000 simulated trials across a range of

trial settings using both peak log-VL and log-VL AUC as the transmission proxy compared to the empirical variance of $\hat{\Phi}^C_{TP}(\tau)$. The dotted line represents valid estimation of the variance of the estimator. We see that $\hat{var}(\hat{\Phi}^C_{TP}(\tau))$ generally provides accurate estimates of the variance of the estimator.

Confidence interval coverage probability

Since our estimation technique provides unbiased estimates of $VE^C_{TP}(\tau)$ and our estimates of $var(\hat{\Phi}^C_{TP}(\tau))$ are valid, we expect that 95% confidence intervals will have approximately 95% coverage probability. Figure 9.3 demonstrates that this is the case – on average, the confidence intervals generated captured the true value of $VE^C_{TP}(\tau)$ 95% of the time, with variability consistent with that expected from simulations of 1,000 repetitions and no apparent trends in the coverage probability with $VE^C_{TP}(\tau)$.

9.1.2 Performance

9.1.3 Comparison to conditional analysis

One of the primary rationales for the unconditional analysis using VE_{TP} allows is resilience in high VE_S settings when compared to conditional analyses. Figure 9.4 shows the power to detect a vaccine effect on peak observed log-VL, conditional on infection, across a range of Δ peak log-VL and VE_S values (using a one-sided two-sample t test). As expected, we see that power decreases as VE_S increases. Figure 9.5 shows the power to detect $VE^C_{TP}(\tau)$ using peak log-VL as the transmission proxy. Figure 9.6 overlays the power curves from Figures 9.4 and 9.5. We see that the unconditional analysis has less power to detect a vaccine effect when VE_S is low, but in mild-to-moderate VE_S scenarios (e.g., $VE_S \geq 30\%$), the unconditional analysis provides far higher power. Thus we can see that $VE^C_{TP}(\tau)$ provides an attractive alternative to analyses of post-infection proxies of infectiousness when vaccine efficacy against infection is expected to be high. The same trends hold if log-VL AUC is used as the proxy of infectiousness (see Figure A.1 in Appendix).

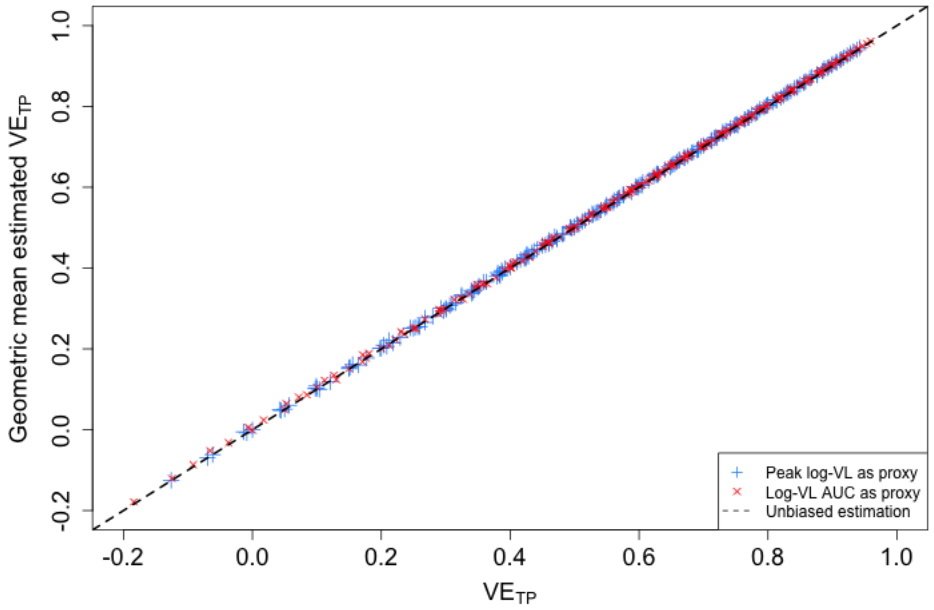

Figure 9.1: Geometric mean of estimates of $\hat{VE}^C_{TP}(\tau)$ compared to $VE^C_{TP}(\tau)$ based on 1,000 simulated trials.

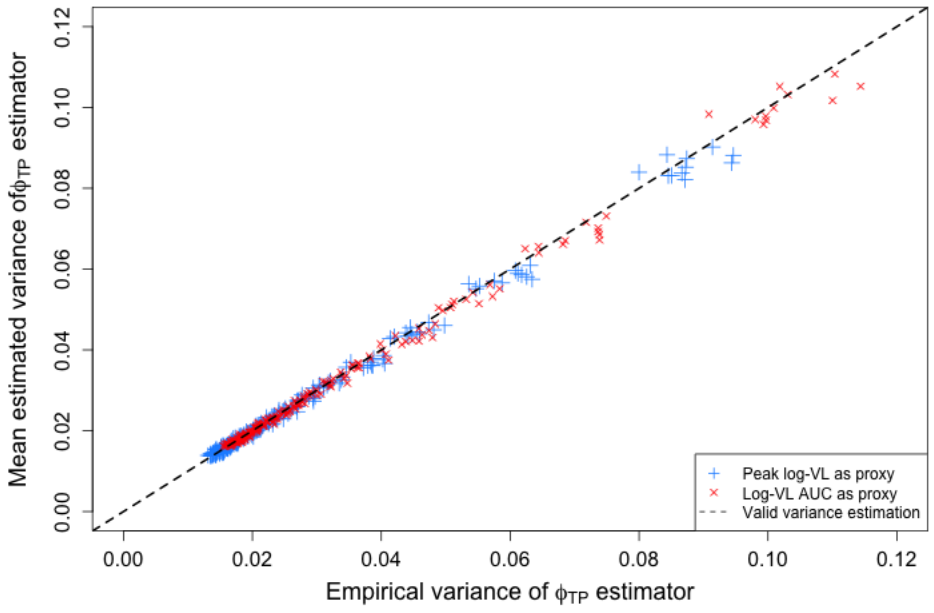

Figure 9.2: Average $\hat{var}(\hat{\Phi}^C_{TP}(\tau))$ compared to empirical variance of $\hat{\Phi}^C_{TP}(\tau)$ based on 1,000 simulated trials.

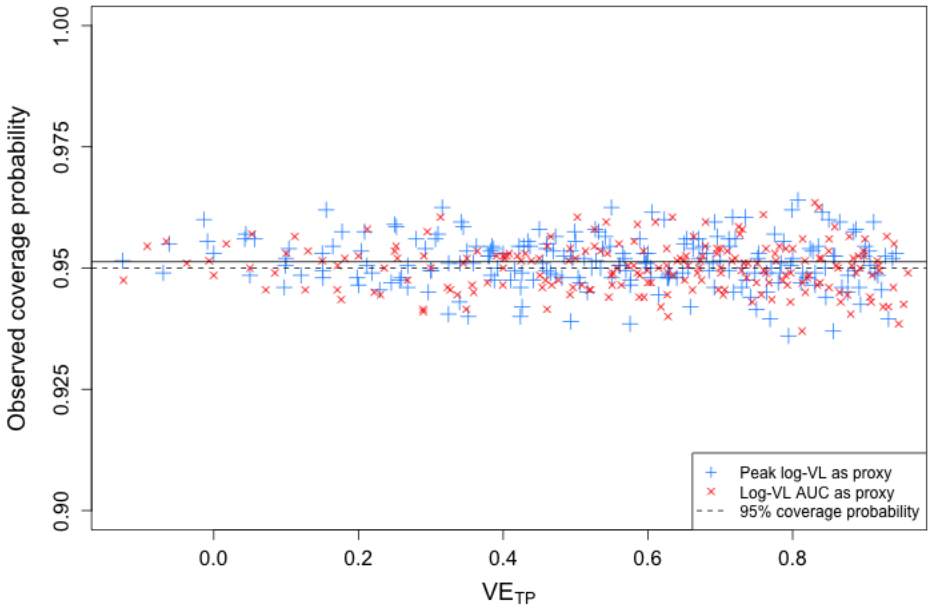

Figure 9.3: Observed coverage probability of 95% confidence intervals for $VE_{TP}^{C}(\tau)$ based on 1,000 simulated trials.

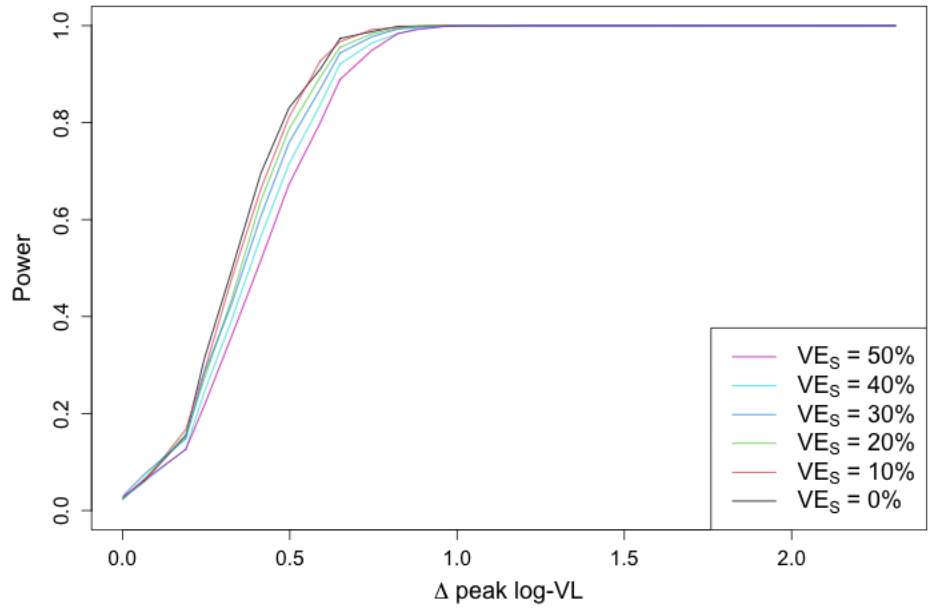

Figure 9.4: Power to detect vaccine effect on peak observed log-VL, conditional on infection, across a range of VE_S values

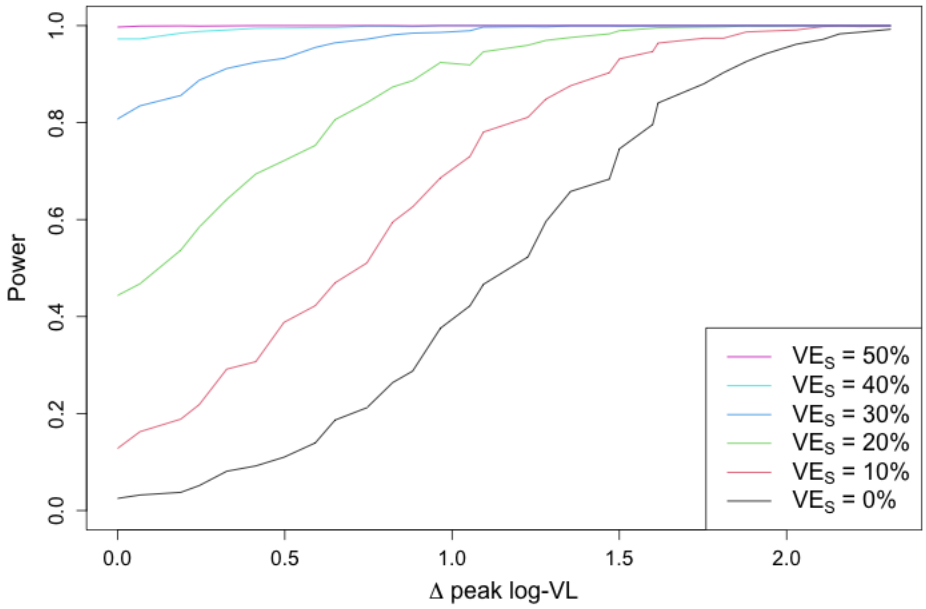

Figure 9.5: Power to detect $VE^C_{TP}(\tau)$ across a range of VE_S values

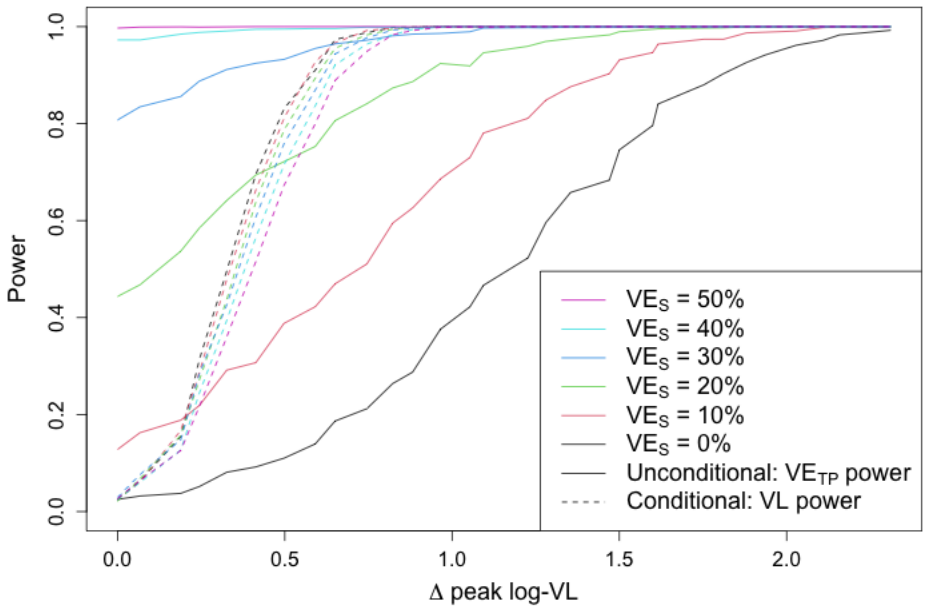

Figure 9.6: Power to detect vaccine effect on peak observed log-VL, conditional on infection, and $VE^C_{TP}(\tau)$ across a range of VE_S values.

9.1.4 Comparison of peak log-VL and VL AUC

Selecting an appropriate VL summary measure to use as a proxy for infectiousness is primarily a question of scientific interest (e.g., which is most likely to accurately represent the relationship between viral shedding and infectiousness), and statistical concerns characteristics are generally a secondary concern. However, we feel it appropriate to note the difference in power between analysis conducted using the two. Given the higher variability of log-VL AUC, we expected lower power, even when standardizing to compare simulations with the same actual $VE^C_{TP}(\tau)$. Figure 9.7 shows the power to detect $VE^C_{TP}(\tau)$ across a range of VE_{TP} values using both summary measures as proxies. We see, as expected, that using log-VL AUC as the infectiousness proxy leads to slightly lower power.

9.2 Validation of estimation methods for $VE^C_{TP,W}(\tau)$

Our simulations with covariates affecting incidence, VE_S, peak log-VL, and Δ peak log-VL were primarily focused on demonstrating that the methods described in Chapter 7 led to valid inference.

9.2.1 Unbiased estimation

We expect $\hat{\Phi}^C_{TP,W}(\tau)$ to be an unbiased estimator of $\Phi^C_{TP,W}(\tau)$. Thus, the geometric mean of our estimates $\hat{VE}^C_{TP,W}(\tau)$ should equal to $VE^C_{TP,W}(\tau)$. Figure 9.8 shows the geometric mean $\hat{VE}^C_{TP,W}(\tau)$ from the 1,000 simulated trials across a range of $VE^C_{TP,W}(\tau)$ values using both peak log-VL and log-VL AUC as the transmission proxy. The dotted line corresponds with unbiased estimation. We see that the estimator provides unbiased estimates across this range of $VE^C_{TP,W}(\tau)$ values for both choices of transmission proxy.

9.2.2 Valid variance estimates

In order for inference to be valid, we also need $\hat{var}(\hat{\Phi}^C_{TP,W}(\tau))$ to consistently estimate $var(\hat{\Phi}^C_{TP,W}(\tau))$. Figure 9.9 shows the mean $\hat{var}(\hat{\Phi}^C_{TP,W}(\tau))$ from the 1,000 simulated trials

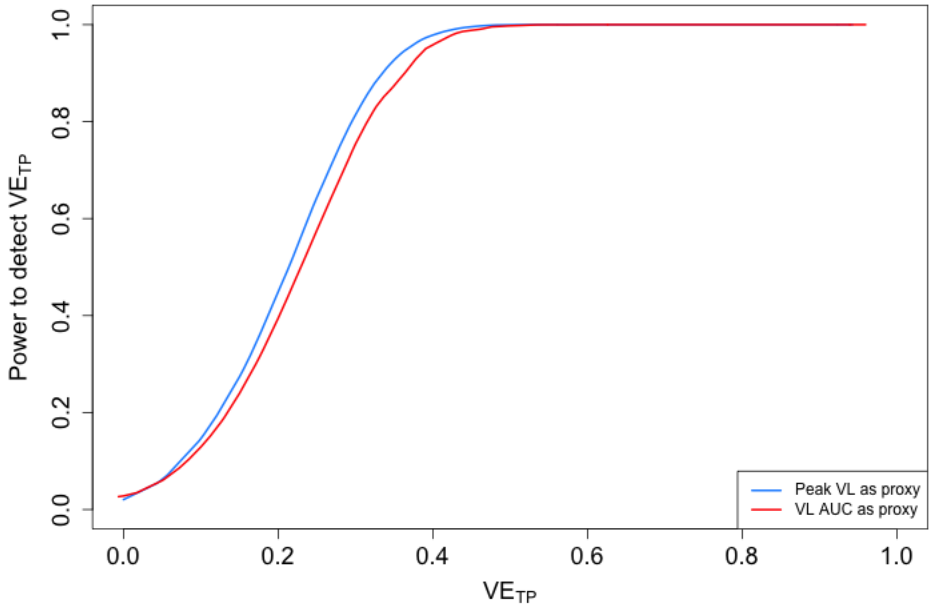

Figure 9.7: Power to detect $VE^C_{TP}(\tau)$ across a range of $VE^C_{TP}(\tau)$ values using peak log-VL and log-VL AUC as infectiousness proxies

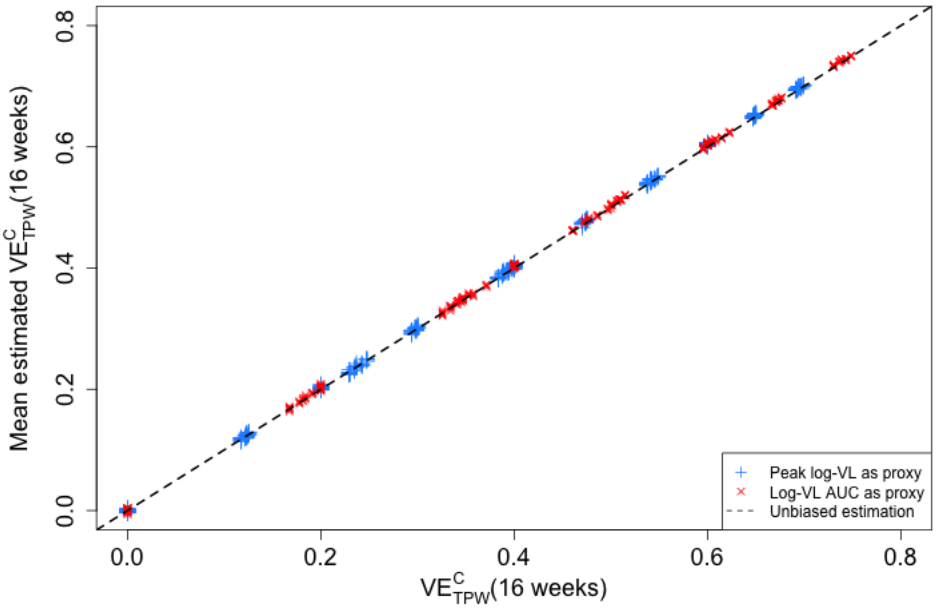

Figure 9.8: Mean $\hat{VE}^C_{TP,W}(\tau)$ compared to $VE^C_{TP,W}(\tau)$ based on 1,000 simulated trials.

across a range of trial settings using both peak log-VL and log-VL AUC as the transmission proxy compared to the empirical variance of $\hat{\Phi}^C_{TP,W}(\tau)$. The dotted line represents valid variance estimation. We see that $\hat{var}(\hat{\Phi}^C_{TP,W}(\tau))$ generally provides accurate estimates of the variance of the estimator.

9.2.3 Confidence interval coverage probability

Since our estimation technique provides unbiased estimates of $VE^C_{TP,W}(\tau)$ and our estimates of $var(\hat{VE}^C_{TP,W}(\tau))$ are valid, we expect that 95% confidence intervals will have approximately 95% coverage probability. Figure 9.10 demonstrates that this is the case – on average, the confidence intervals generated captured the true value of $VE^C_{TP,W}(\tau)$ 95% of the time.

9.3 Validation and performance of estimation methods for $VE_{VL}(\tau, V)$

9.3.1 Validation

Unbiased estimation

We expect $\hat{VE}_{VL}(\tau, V)$ to be an unbiased estimator of $VE_{VL}(\tau, V)$. Figure 9.11 shows the mean $\hat{VE}^C_{TP,W}(\tau)$ from the 1,000 simulated trials across a range of $VE^C_{TP,W}(\tau)$ values using both peak log-VL and log-VL AUC as the transmission proxy. The dotted line corresponds with unbiased estimation. We see that the estimator provides unbiased estimates across this range of $VE_{VL}(\tau, V)$ values.

Valid variance estimates

In order for inference to be valid, we also need $\hat{var}(\hat{VE}_{VL}(\tau, V))$ to consistently estimate $var(\hat{VE}_{VL}(\tau, V))$. Figure 9.12 shows the mean $\hat{var}(\hat{VE}_{VL}(\tau, V))$ from the 1,000 simulated trials across a range of trial settings compared to the empirical variance of $\hat{VE}_{VL}(\tau, V)$. The dotted line represents valid variance estimation. We see that $\hat{var}(\hat{VE}_{VL}(\tau, V))$ generally provides accurate estimates of the variance of the estimator.

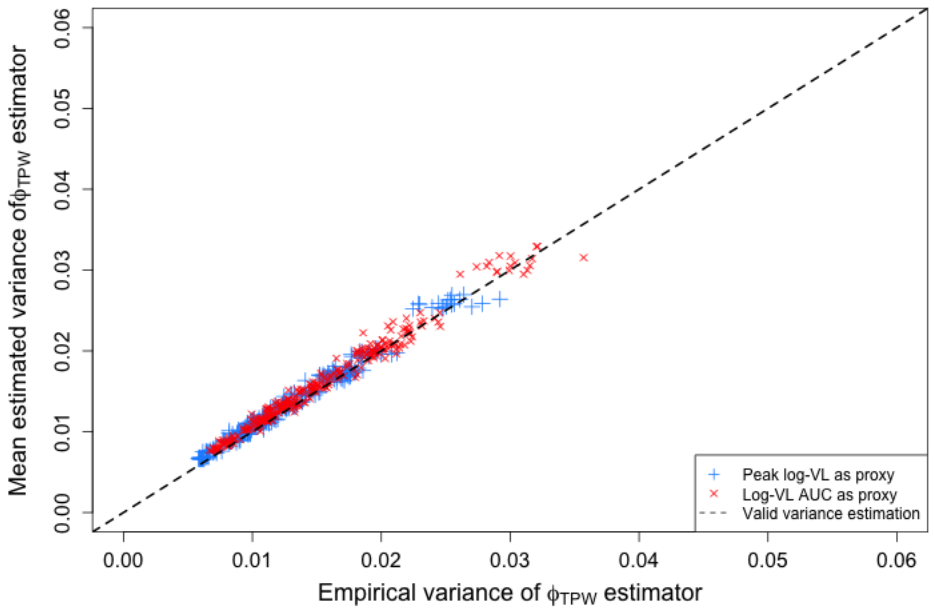

Figure 9.9: Average $\hat{var}(\hat{\Phi}^C_{TP,W}(\tau))$ compared to empirical variance of $\hat{\Phi}^C_{TP,W}(\tau)$ based on 1,000 simulated trials.

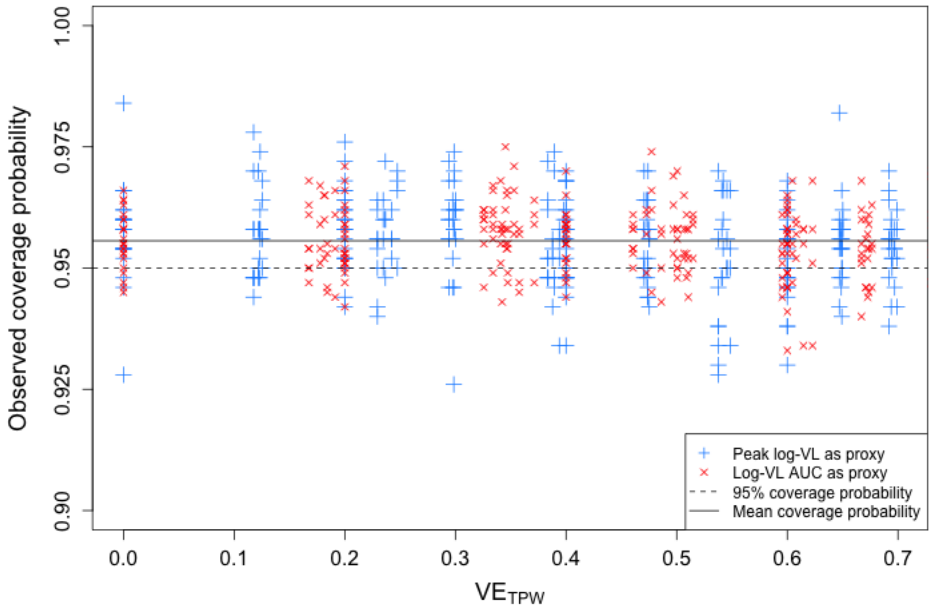

Figure 9.10: Observed coverage probability of 95% confidence intervals for $VE^{C,W}_{TP}(\tau)$ based on 1,000 simulated trials.

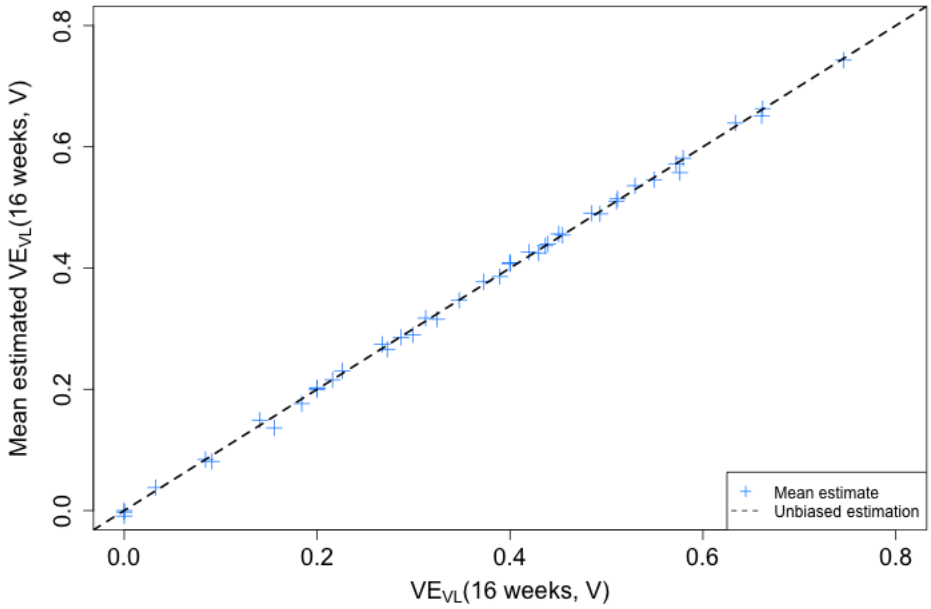

Figure 9.11: Mean $\hat{VE}_{VL}(\tau, V)$ compared to $VE_{VL}(\tau, V)$ based on 1,000 simulated trials.

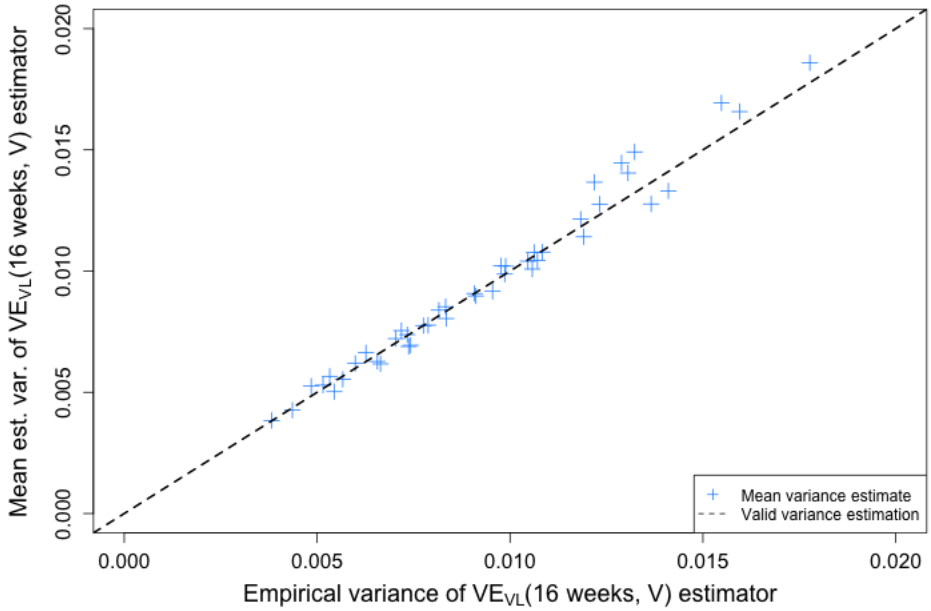

Figure 9.12: Average $\hat{var}(\hat{VE}_{VL}(\tau, V))$ for all $V \in \{4, 5, 6\}$ compared to empirical variance of $\hat{VE}_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$ based on 1,000 simulated trials.

Simultaneous confidence band coverage probability

Since our estimation technique provides unbiased estimates of $VE_{VL}(\tau, V)$ and our estimates of $var(\hat{VE}_{VL}(\tau, V))$ are valid, we expect the simultaneous confidence band method described in 6.3.2 to create valid confidence bands. In the context of our simulations, a success would indicate the confidence band contains $VE_{VL}(\tau, V)$ for all $V \in \{4, 5, 6\}$. Figure 9.13 demonstrates that this is the case – on average, the simultaneous confidence intervals generated captured the true value of $VE_{VL}(\tau, V)$ for all $V \in \{4, 5, 6\}$ 95% of the time.

9.3.2 Performance

We also find it of interest to characterize the performance of the simultaneous confidence method compared to pointwise confidence intervals, and to compare the inference methods to detect any vaccine effect on $VE_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$ to the methods to detect vaccine effects on $VE^{C}_{TP}(\tau)$.

Comparison of simultaneous confidence intervals to pointwise confidence intervals

We expect to sacrifice some precision when generating simultaneous confidence bands instead of calculating pointwise confidence intervals, leading to wider confidence intervals. Figure 9.14 shows the average 95% simultaneous confidence band for $VE_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$ based on 1.000 simulated trials compared to the average 95% pointwise confidence intervals from the same trial when $VE_S = 0.4$ and Δ peak log-VL = 1. As expected, we sacrifice some precision compared to a pointwise confidence interval, but the benefit of not pre-specifying a threshold in the presence of uncertainty may outweigh the loss of precision. Similar figures for other combinations of VE_S and Δ peak log-VL can be found in the Appendix.

Comparison to VE_S

When comparing an analysis detecting $VE_{VL}(\tau, V)$ to an analysis detecting $VE^{C}_{S}(\tau)$, we note that there is a trade-off between the number of events and the overall effect size. We

expect to sacrifice some power in detecting an effect on time to crossing a VL threshold if there is no vaccine effect on VL, as we would trade off events without increasing the effect size. However, if there is a moderate-to-high vaccine effect on peak log-VL, we would expect to gain power as we see a larger effect size that compensates for the lost events. Furthermore, if there were opposing effects (i.e., vaccine increases peak log-VL) we would expect to see lower power in the analysis of VE_{VL}. Figure 9.15 compares the power to detect $VE_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$ and $VE_S^C(\tau)$ across a range of VE_S values and Δ peak log-VL values. As expected, we see a slight reduction in power when there is no vaccine effect on VL, but we gain power as Δ peak log-VL increases. The tradeoff in power would be larger for higher VL threshold (e.g., $V \in \{5, 6, 7\}$) where fewer participants cross the threshold in the placebo group.

Comparison to VE_{TP}

Finally, it is of interest to compare the analysis detecting $VE_{VL}(\tau, V)$ to an analysis detecting $VE_{TP}^C(\tau)$, since both methods are options for unconditional analysis of post-infection infectiousness proxies. Figure 9.16 compares the power to detect $VE_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$ and $VE_{TP}^C(\tau)$ across a range of VE_S values and Δ peak log-VL values. We see that, in the context of the model for viral load and vaccine effect on viral load that we have chosen, the analysis to detect $VE_{VL}(\tau, V)$ has higher power. The power advantage increases with Δ peak log-VL.

Upon closer examination, however, this apparent difference in power is largely a function of how $VE_{VL}(\tau, V)$ and $VE_{TP}(\tau)$ parameterize vaccine effects on VL differently. The two parameters (and thus the power to detect them) are affected differently by values of Δ peak log-VL under our simulation model. As we discussed in Sections 4.5.1 and 4.6, $VE_{TP}(\tau)$ depends on the proportional reduction in the mean peak log-VL (in addition to VE_S). $VE_{VL}(\tau, V)$ on the other hand, depends on the proportional reduction in the probability of people infected at / before time τ having VL cross VL threshold V at / before time τ (again, in addition to VE_S). Table 9.1 shows the values of $VE_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$ and

$VE^C_{TP}(\tau)$ for a range of VE_S and and Δ peak log-VL values in our simulation model. It is clear that $VE_{VL}(\tau, 5)$ and $VE_{VL}(\tau, 6)$ are higher than $VE^C_{TP}(\tau)$ for the same combination of VE_S and Δ peak log-VL, and the gap increases with Δ peak log-VL. Since we expect higher power to detect larger values of VE, the gap in power between the two analyses makes sense. Further simulation studies could assess differences between the two methods under different VL models.

VE_S	Δ peak log-VL	$VE_{VL}(\tau, 4)$	$VE_{VL}(\tau, 5)$	$VE_{VL}(\tau, 6)$	$VE^C_{TP}(\tau)$
0	0	0.00	0.00	0.00	0.00
0	1	0.08	0.18	0.30	0.11
0	2	0.22	0.39	0.58	0.23
0.2	0	0.20	0.20	0.20	0.20
0.2	1	0.27	0.35	0.44	0.29
0.2	2	0.37	0.51	0.66	0.38
0.4	0	0.40	0.40	0.40	0.40
0.4	1	0.45	0.51	0.58	0.47
0.4	2	0.53	0.63	0.75	0.54

Table 9.1: Values of $VE_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$ and $VE^C_{TP}(\tau)$ for a range of VE_S and and Δ peak log-VL values

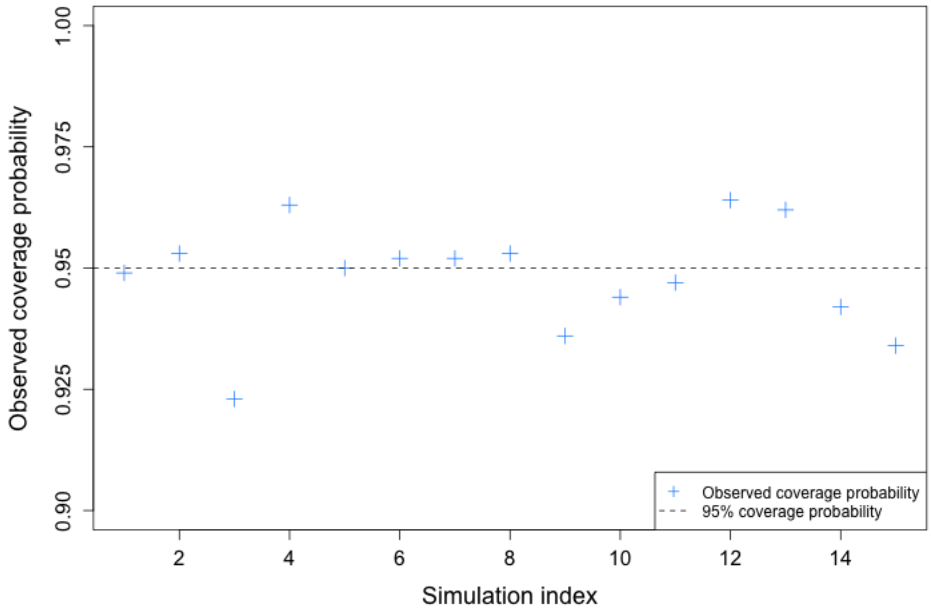

Figure 9.13: Observed coverage probability of 95% simultaneous confidence intervals for $VE_{VL}(\tau, V)$ for all $V \in \{4, 5, 6\}$ based on 1,000 simulated trials.

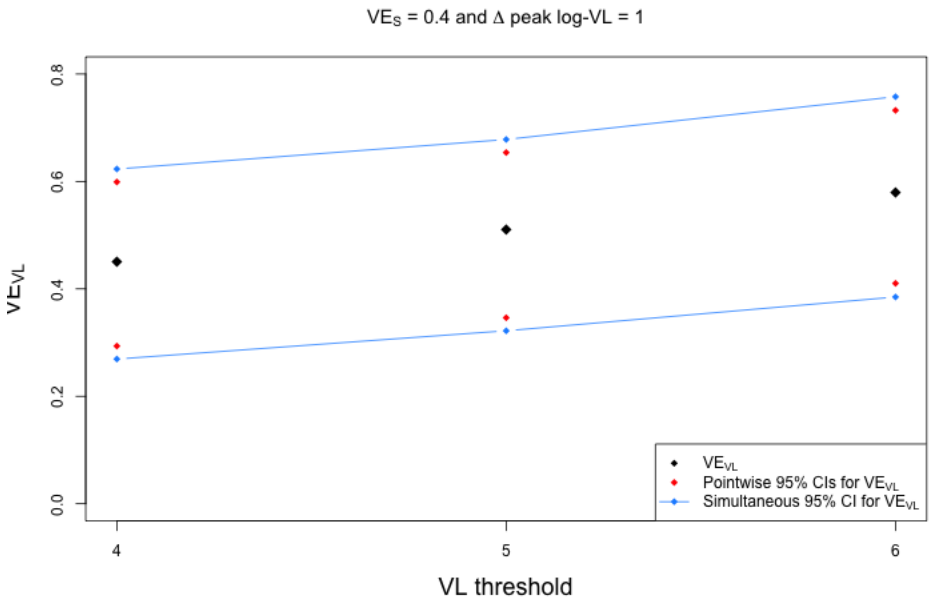

Figure 9.14: Average 95% simultaneous confidence bands (blue lines) compared to average pointwise 95% confidence intervals (red dots) for $VE_{VL}(\tau, V)$ for all $V \in \{4, 5, 6\}$, based on 1,000 simulated trials with $VE_S = 0.4$, Δ peak log-VL = 1

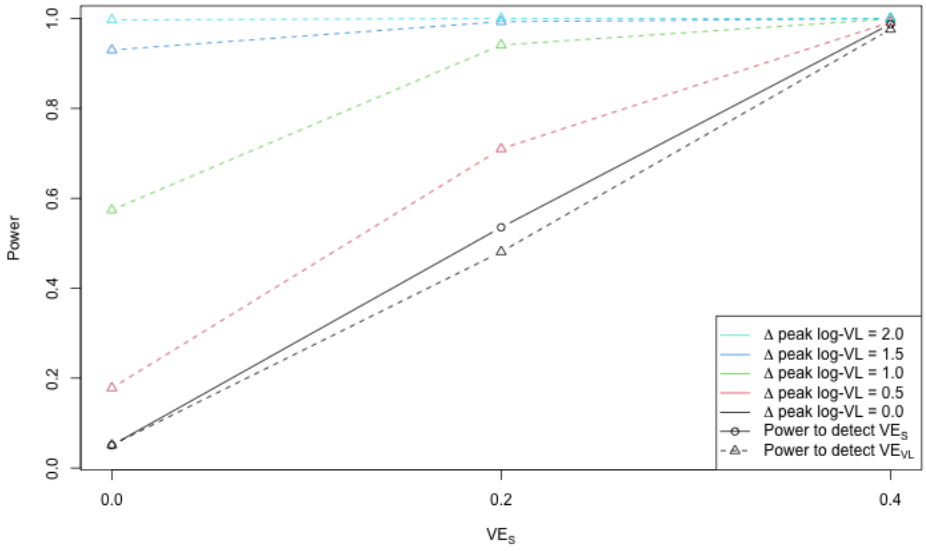

Figure 9.15: Power to detect $VE_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$ and $VE_S^C(\tau)$ across a range of VE_S values and Δ peak log-VL values

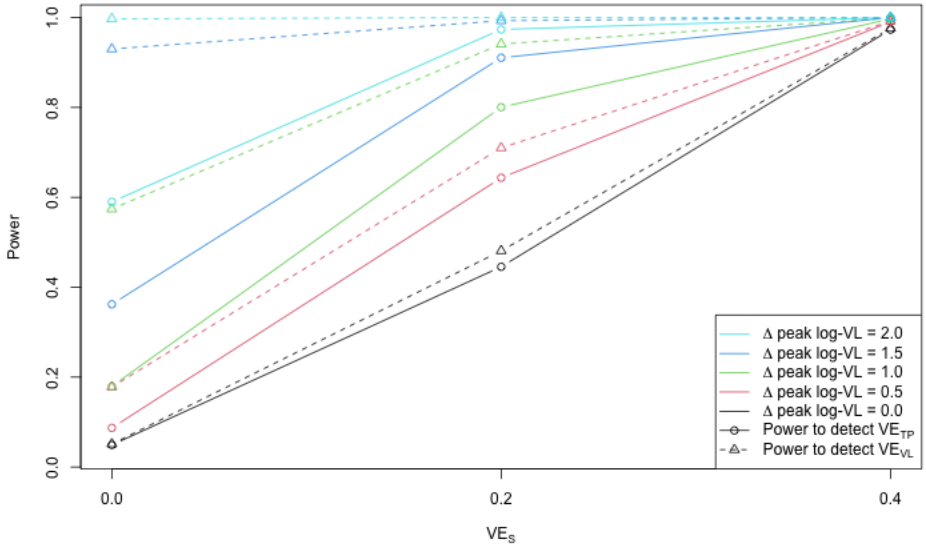

Figure 9.16: Power to detect $VE_{VL}(\tau, V)$ for $V \in \{4, 5, 6\}$ and $VE_{TP}^C(\tau)$ across a range of VE_S values and Δ peak log-VL values

Chapter 10

DISCUSSION

In this thesis, we carefully defined and outlined estimation techniques for different forms of vaccine efficacy for transmission potential (VE_{TP}) and vaccine efficacy for viral load threshold $VE_{VL}(V)$. We also defined and outlined an estimation technique for a discrete covariate-adjusted form of VE_{TP}. Such parameters and strategies are motivated by a desire to understand vaccine effects on infectiousness without conditioning on infection. The transmission potential endpoints and corresponding efficacy parameters rely on proxies of infectiousness; we discussed a number of possible infectiousness proxies derived from viral load and how they are estimated from real-world data. We also conducted simulations studies based on an ongoing COVID-19 vaccine trial design with simulated viral load trajectories based on real-world data to validate the described methods and compare results from different tests.

The simulations showed that, in addition to being valid methods for inference, the techniques for VE_{TP} and $VE_{VL}(V)$ have power to detect vaccine effects on viral load / transmission potential without conditioning on infection, especially in high VE_S settings. When we compare power to detect VE_{TP} to analyses of post-infection endpoints that condition on infection, we see higher power in scenarios with moderate-to-high VE_S. Since real-world data collected as vaccines have rolled out have indicated that VE_S may actually be moderate-to-high, these unconditional endpoints provide an attractive choice as a primary or secondary endpoint in clinical trials.

Importantly, this comparison between the conditional and unconditional analyses here does not account for the issue of post-randomization selection bias. Such methods are beyond the scope of this analysis, but it is likely that accounting for this bias would lower the power of

conditional analysis, making the product method even more preferable. Additionally, VE_{TP} provides fuller context on the total vaccine effect on onward transmission than conditional vaccine effects alone. For example, in scenarios where the vaccine effect on VL conditional on infection is detrimental (e.g., people who would have had low VL end up not getting infected, leaving only high VL infections), conditional VE would show a harmful vaccine effect on VL, whereas VE_{TP} would account for the people who are never infected. This distinction is particularly important for setting appropriate public health policies that apply to vaccinated people. Unconditional VE_{TP} is a useful parameter to inform policy decisions that affect risk of vaccinated people in general transmitting to others (e.g., mask-wearing guidance, social distancing). Conditional VE parameters should also play an important role in setting health policy for vaccinated people who have become infected (e.g., isolation guidelines).

Furthermore, we demonstrated that tests to detect $VE_{VL}(V)$ without pre-specifying a VL threshold V also have decent power in high VE_S settings. The relative power of analyses of VE_{TP} to analyses of $VE_{VL}(V)$ depends on the type of vaccine effect on viral load; choosing one over another is more of a scientific question (e.g., which endpoint is more likely to characterize the true nature of infectiousness, what is the likely form of the vaccine effect on viral load) than a statistical question. That said, further simulation studies could more deeply explore the differences between the methods.

There are a number of downsides to both VE_{TP} and $VE_{VL}(V)$. First, an analysis VE_{TP} relies selecting a proxy of infectiousness, which can be difficult to do for novel diseases where patterns of transmission are not yet fully understood. Different proxies for infectiousness have different trade-offs as discussed in Chapter 2. Generating more reliable / accurate proxies of infectiousness or establishing / verifying surrogates of infectiousness would drastically improve the usefulness of VE_{TP} in public health policy making purposes. In addition to improving proxies for infectiousness, these conditional analysis methods could be applied to actual measures of infectiousness (e.g. X in cumulative transmission endpoint is the number of confirmed secondary transmission events from an infected person), though we would

then face previously discussed issues with directly tracking secondary transmission events. Additionally, in the COVID-19 setting (and likely in the context of other acute respiratory infections to which these approaches could be applied), accurately estimating proxies of infectiousness using empirical estimators requires frequent diagnostic testing and VL sampling post-infection so that the most informative periods of viral shedding can be detected. Such a requirement places a burden on trial participants, which can lead to substantial amounts of missing data (which we do not explore here). Model-based estimators may lower this burden, but can be difficult to validate and may introduce bias if the model is mis-specified.

BIBLIOGRAPHY

[1] Andrew T Levin, William P Hanage, Nana Owusu-Boaitey, Kensington B Cochran, Seamus P Walsh, and Gideon Meyerowitz-Katz. Assessing the age specificity of infection fatality rates for covid-19: systematic review, meta-analysis, and public policy implications. *European journal of epidemiology*, pages 1–16, 2020.

[2] Zhiru Gao, Yinghui Xu, Chao Sun, Xu Wang, Ye Guo, Shi Qiu, and Kewei Ma. A systematic review of asymptomatic infections with covid-19. *Journal of Microbiology, Immunology and Infection*, 54(1):12–16, 2021.

[3] Andreas Kronbichler, Daniela Kresse, Sojung Yoon, Keum Hwa Lee, Maria Effenberger, and Jae Il Shin. Asymptomatic patients as a source of covid-19 infections: A systematic review and meta-analysis. *International journal of infectious diseases*, 98:180–186, 2020.

[4] Daniel P Oran and Eric J Topol. Prevalence of asymptomatic sars-cov-2 infection: a narrative review. *Annals of internal medicine*, 173(5):362–367, 2020.

[5] Daniel P Oran and Eric J Topol. The proportion of sars-cov-2 infections that are asymptomatic: a systematic review. *Annals of internal medicine*, 2021.

[6] Diana Buitrago-Garcia, Dianne Egli-Gany, Michel J Counotte, Stefanie Hossmann, Hira Imeri, Aziz Mert Ipekci, Georgia Salanti, and Nicola Low. Occurrence and transmission potential of asymptomatic and presymptomatic sars-cov-2 infections: A living systematic review and meta-analysis. *PLoS medicine*, 17(9):e1003346, 2020.

[7] Seungjae Lee, Tark Kim, Eunjung Lee, Cheolgu Lee, Hojung Kim, Heejeong Rhee, Se Yoon Park, Hyo-Ju Son, Shinae Yu, Jung Wan Park, et al. Clinical course and molecular viral shedding among asymptomatic and symptomatic patients with sars-cov-2 infection in a community treatment center in the republic of korea. *JAMA internal medicine*, 180(11):1447–1452, 2020.

[8] Rui Zhou, Furong Li, Fengjuan Chen, Huamin Liu, Jiazhen Zheng, Chunliang Lei, and Xianbo Wu. Viral dynamics in asymptomatic patients with covid-19. *International Journal of Infectious Diseases*, 96:288–290, 2020.

[9] Lirong Zou, Feng Ruan, Mingxing Huang, Lijun Liang, Huitao Huang, Zhongsi Hong, Jianxiang Yu, Min Kang, Yingchao Song, Jinyu Xia, et al. Sars-cov-2 viral load in

upper respiratory specimens of infected patients. *New England Journal of Medicine*, 382(12):1177–1179, 2020.

[10] Lindsey R Baden, Hana M El Sahly, Brandon Essink, Karen Kotloff, Sharon Frey, Rick Novak, David Diemert, Stephen A Spector, Nadine Rouphael, C Buddy Creech, et al. Efficacy and safety of the mrna-1273 sars-cov-2 vaccine. *New England Journal of Medicine*, 384(5):403–416, 2021.

[11] Fernando P Polack, Stephen J Thomas, Nicholas Kitchin, Judith Absalon, Alejandra Gurtman, Stephen Lockhart, John L Perez, Gonzalo Pérez Marc, Edson D Moreira, Cristiano Zerbini, et al. Safety and efficacy of the bnt162b2 mrna covid-19 vaccine. *New England Journal of Medicine*, 383(27):2603–2615, 2020.

[12] Jerald Sadoff, Glenda Gray, An Vandebosch, Vicky Cárdenas, Georgi Shukarev, Beatriz Grinsztejn, Paul A Goepfert, Carla Truyers, Hein Fennema, Bart Spiessens, et al. Safety and efficacy of single-dose ad26. cov2. s vaccine against covid-19. *New England Journal of Medicine*, 2021.

[13] Merryn Voysey, Sue Ann Costa Clemens, Shabir A Madhi, Lily Y Weckx, Pedro M Folegatti, Parvinder K Aley, Brian Angus, Vicky L Baillie, Shaun L Barnabas, Qasim E Bhorat, et al. Safety and efficacy of the chadox1 ncov-19 vaccine (azd1222) against sars-cov-2: an interim analysis of four randomised controlled trials in brazil, south africa, and the uk. *The Lancet*, 397(10269):99–111, 2021.

[14] Mark G Thompson, Jefferey L Burgess, Allison Naleway, Harmony Tyner, Sarang K Yoon, Jennifer Meece, Lauren EW Olsho, Alberto Caban-Martinez, Ashley L Fowlkes, Karen Lutrick, et al. Prevention and attenuation of covid-19 by bnt162b2 and mrna-1273 vaccines. *medRxiv*, 2021.

[15] Nick K Jones, Lucy Rivett, Shaun Seaman, Richard J Samworth, Ben Warne, Chris Workman, Mark Ferris, Jo Wright, Natalie Quinnell, Ashley Shaw, et al. Single-dose bnt162b2 vaccine protects against asymptomatic sars-cov-2 infection. *Elife*, 10:e68808, 2021.

[16] Victoria Jane Hall, Sarah Foulkes, Ayoub Saei, Nick Andrews, Blanche Oguti, Andre Charlett, Edgar Wellington, Julia Stowe, Natalie Gillson, Ana Atti, et al. Covid-19 vaccine coverage in health-care workers in england and effectiveness of bnt162b2 mrna vaccine against infection (siren): a prospective, multicentre, cohort study. *The Lancet*, 397(10286):1725–1735, 2021.

[17] Chung-Ming Chu, Vincent CC Cheng, Ivan FN Hung, Kin-Sang Chan, Bone SF Tang, Thomas HF Tsang, Kwok-Hung Chan, and Kwok-Yung Yuen. Viral load distribution in sars outbreak. *Emerging infectious diseases*, 11(12):1882, 2005.

[18] Isabelle Boucoiran, Bryan T Mayer, Elizabeth M Krantz, Arnaud Marchant, Sunil Pati, Suresh Boppana, Anna Wald, Larry Corey, Corey Casper, Joshua T Schiffer, et al. Nonprimary maternal cytomegalovirus infection after viral shedding in infants. *The Pediatric infectious disease journal*, 37(7):627–631, 2018.

[19] Bryan T Mayer, Elizabeth M Krantz, Anna Wald, Lawrence Corey, Corey Casper, Soren Gantt, and Joshua T Schiffer. Estimating the risk of human herpesvirus 6 and cytomegalovirus transmission to ugandan infants from viral shedding in saliva by household contacts. *Viruses*, 12(2):171, 2020.

[20] Ignatius TS Yu, Yuguo Li, Tze Wai Wong, Wilson Tam, Andy T Chan, Joseph HW Lee, Dennis YC Leung, and Tommy Ho. Evidence of airborne transmission of the severe acute respiratory syndrome virus. *New England Journal of Medicine*, 350(17):1731–1739, 2004.

[21] Tim K Tsang, Benjamin J Cowling, Vicky J Fang, Kwok-Hung Chan, Dennis KM Ip, Gabriel M Leung, JS Malik Peiris, and Simon Cauchemez. Influenza a virus shedding and infectivity in households. *The Journal of infectious diseases*, 212(9):1420–1428, 2015.

[22] Ashish Goyal, Daniel B Reeves, E Fabian Cardozo-Ojeda, Joshua T Schiffer, and Bryan T Mayer. Wrong person, place and time: viral load and contact network structure predict sars-cov-2 transmission and super-spreading events. *Medrxiv*, 2020.

[23] Dean Follmann, Michael P Fay, and Michael Proschan. Chop-lump tests for vaccine trials. *Biometrics*, 65(3):885–893, 2009.

[24] Peter A Lachenbruch. Comparisons of two-part models with competitors. *Statistics in medicine*, 20(8):1215–1234, 2001.

[25] Devan V Mehrotra, Xiaoming Li, and Peter B Gilbert. A comparison of eight methods for the dual-endpoint evaluation of efficacy in a proof-of-concept hiv vaccine trial. *Biometrics*, 62(3):893–900, 2006.

[26] Dean Follmann and Chiung-Yu Huang. Incorporating founder virus information in vaccine field trials. *Biometrics*, 71(2):386–396, 2015.

[27] Christophe Fraser, T Déirdre Hollingsworth, Ruth Chapman, Frank de Wolf, and William P Hanage. Variation in hiv-1 set-point viral load: epidemiological analysis and an evolutionary hypothesis. *Proceedings of the National Academy of Sciences*, 104(44):17441–17446, 2007.

[28] James P Hughes, Jared M Baeten, Jairam R Lingappa, Amalia S Magaret, Anna Wald, Guy de Bruyn, James Kiarie, Mubiana Inambao, William Kilembe, Carey Farquhar, et al. Determinants of per-coital-act hiv-1 infectivity among african hiv-1–serodiscordant couples. *Journal of Infectious Diseases*, 205(3):358–365, 2012.

[29] Toru Watanabe, Timothy A Bartrand, Mark H Weir, Tatsuo Omura, and Charles N Haas. Development of a dose-response model for sars coronavirus. *Risk Analysis: An International Journal*, 30(7):1129–1138, 2010.

[30] Igor M Belyakov, Zdenek Hel, Brian Kelsall, Vladimir A Kuznetsov, Jeffrey D Ahlers, Janos Nacsa, David I Watkins, Todd M Allen, Alessandro Sette, John Altman, et al. Mucosal aids vaccine reduces disease and viral load in gut reservoir and blood after mucosal infection of macaques. *Nature medicine*, 7(12):1320–1326, 2001.

[31] Miles P Davenport, Lei Zhang, John W Shiver, Danilo R Casmiro, Ruy M Ribeiro, and Alan S Perelson. Influence of peak viral load on the extent of cd4+ t-cell depletion in simian hiv infection. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 41(3):259–265, 2006.

[32] Masumbuko Claude Kasereka, Austin D Ericson, Andrea L Conroy, Lukaba Tumba, Ombeni Didier Mwesha, and Michael T Hawkes. Prior vaccination with recombinant vesicular stomatitis virus–zaire ebolavirus vaccine is associated with improved survival among patients with ebolavirus infection. *Vaccine*, 38(14):3003–3007, 2020.

[33] David A Swan, Ashish Goyal, Chloe Bracis, Mia Moore, Elizabeth Krantz, Elizabeth R Brown, Fabian Cardozo-Ojeda, Daniel B Reeves, Fei Gao, Peter B Gilbert, et al. Vaccines that prevent sars-cov-2 transmission may prevent or dampen a spring wave of covid-19 cases and deaths in 2021. *medRxiv*, 2020.

[34] Muge Cevik, Matthew Tate, Ollie Lloyd, Alberto Enrico Maraolo, Jenna Schafers, and Antonia Ho. Sars-cov-2, sars-cov, and mers-cov viral load dynamics, duration of viral shedding, and infectiousness: a systematic review and meta-analysis. *The Lancet Microbe*, 2020.

[35] Matan Levine-Tiefenbrun, Idan Yelin, Rachel Katz, Esma Herzel, Ziv Golan, Licita Schreiber, Tamar Wolf, Varda Nadler, Amir Ben-Tov, Jacob Kuint, et al. Decreased sars-cov-2 viral load following vaccination. *medRxiv*, 2021.

[36] Nelson Lee, Paul KS Chan, David SC Hui, Timothy H Rainer, Eric Wong, Kin-Wing Choi, Grace CY Lui, Bonnie CK Wong, Rita YK Wong, Wai-Yip Lam, et al. Viral loads and duration of viral shedding in adult patients hospitalized with influenza. *The Journal of infectious diseases*, 200(4):492–500, 2009.

[37] Surbhi Leekha, Nicole L Zitterkopf, Mark J Espy, Thomas F Smith, Rodney L Thompson, Priya Sampathkumar, et al. Duration of influenza a virus shedding in hospitalized patients and implications for infection control. *Infection Control and Hospital Epidemiology*, 28(9):1071–1076, 2007.

[38] Elizabeth Tronstein, Christine Johnston, Meei-Li Huang, Stacy Selke, Amalia Magaret, Terri Warren, Lawrence Corey, and Anna Wald. Genital shedding of herpes simplex virus among symptomatic and asymptomatic persons with hsv-2 infection. *Jama*, 305(14):1441–1449, 2011.

[39] Jeroen JA van Kampen, David AMC van de Vijver, Pieter LA Fraaij, Bart L Haagmans, Mart M Lamers, Nisreen Okba, Johannes PC van den Akker, Henrik Endeman, Diederik AMPJ Gommers, Jan J Cornelissen, et al. Shedding of infectious virus in hospitalized patients with coronavirus disease-2019 (covid-19): duration and key determinants. *MedRxiv*, 2020.

[40] Bo Zhou, Jianqing She, Yadan Wang, and Xiancang Ma. Duration of viral shedding of discharged patients with severe covid-19. *Clinical Infectious Diseases*, 71(16):2240–2242, 2020.

[41] Majid Laassri, Kathleen Lottenbach, Robert Belshe, Mark Wolff, Margaret Rennels, Stanley Plotkin, and Konstantin Chumakov. Effect of different vaccination schedules on excretion of oral poliovirus vaccine strains. *The Journal of infectious diseases*, 192(12):2092–2098, 2005.

[42] Daniel CL Linhares, Jean Paul Cano, Thomas Wetzell, Joel Nerem, Montserrat Torremorell, and Scott A Dee. Effect of modified-live porcine reproductive and respiratory syndrome virus (prrsv) vaccine on the shedding of wild-type virus from an infected population of growing pigs. *Vaccine*, 30(2):407–413, 2012.

[43] Kevin J Allen, Dragan Rogan, B Brett Finlay, Andrew A Potter, and David J Asper. Vaccination with type iii secreted proteins leads to decreased shedding in calves after experimental infection with escherichia coli o157. *Canadian journal of veterinary research*, 75(2):98–105, 2011.

[44] Jeroen JA van Kampen, David AMC van de Vijver, Pieter LA Fraaij, Bart L Haagmans, Mart M Lamers, Nisreen Okba, Johannes PC van den Akker, Henrik Endeman, Diederik AMPJ Gommers, Jan J Cornelissen, et al. Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (covid-19). *Nature communications*, 12(1):1–6, 2021.

[45] Barry Atkinson and Eskild Petersen. Sars-cov-2 shedding and infectivity. *The Lancet*, 395(10233):1339–1340, 2020.

[46] Monica E Brint, Joshua M Hughes, Aditya Shah, Chelsea R Miller, Lisa G Harrison, Elizabeth A Meals, Jacqueline Blanch, Charlotte R Thompson, Stephania A Cormier, and John P DeVincenzo. Prolonged viral replication and longitudinal viral dynamic differences among respiratory syncytial virus infected infants. *Pediatric research*, 82(5):872–880, 2017.

[47] Peter Schäfer, Werner Tenschert, Liana Cremaschi, Matthias Schröter, Bernhard Zöllner, and Rainer Laufs. Area under the viraemia curve versus absolute viral load: utility for predicting symptomatic cytomegalovirus infections in kidney transplant patients. *Journal of medical virology*, 65(1):85–89, 2001.

[48] Michael J Mugavero, K Rivet Amico, Andrew O Westfall, Heidi M Crane, Anne Zinski, James H Willig, Julia C Dombrowski, Wynne E Norton, James L Raper, Mari M Kitahata, et al. Early retention in hiv care and viral load suppression: implications for a test and treat approach to hiv prevention. *Journal of acquired immune deficiency syndromes (1999)*, 59(1):86, 2012.

[49] Stephen M Kissler, Joseph R Fauver, Christina Mack, Caroline Tai, Kristin Y Shiue, Chaney C Kalinich, Sarah Jednak, Isabel M Ott, Chantal BF Vogels, Jay Wohlgemuth, et al. Viral dynamics of sars-cov-2 infection and the predictive value of repeat testing. *medRxiv*, 2020.

[50] Peter B Gilbert and Yanqing Sun. Failure time analysis of hiv vaccine effects on viral load and antiretroviral therapy initiation. *Biostatistics*, 6(3):374–394, 2005.

[51] M Elizabeth Halloran, Ira M Longini, Claudio J Struchiner, and Ira M Longini. *Design and analysis of vaccine studies*, volume 18. Springer, 2010.

[52] Thomas R Fleming and David P Harrington. *Counting processes and survival analysis*, volume 169. John Wiley & Sons, 2011.

[53] Stephen A Bustin, Vladimir Benes, Jeremy A Garson, Jan Hellemans, Jim Huggett, Mikael Kubista, Reinhold Mueller, Tania Nolan, Michael W Pfaffl, Gregory L Shipley, et al. The miqe guidelines: M inimum i nformation for publication of q uantitative real-time pcr e xperiments, 2009.

[54] Daniel Nettle, Luise Seeker, Dan Nussey, Hannah Froy, and Melissa Bateson. Consequences of measurement error in qpcr telomere data: A simulation study. *PLoS One*, 14(5):e0216118, 2019.

[55] Dvir Aran. Estimating real-world covid-19 vaccine effectiveness in israel using aggregated counts. *medRxiv*, 2021.

[56] William GF Ditcham, Joshua R Lewis, Robert J Dobson, Nining Hartaningsih, Graham E Wilcox, and Moira Desport. Vaccination reduces the viral load and the risk of transmission of jembrana disease virus in bali cattle. *Virology*, 386(2):317–324, 2009.

Appendix A

ADDITIONAL SIMULATION RESULTS

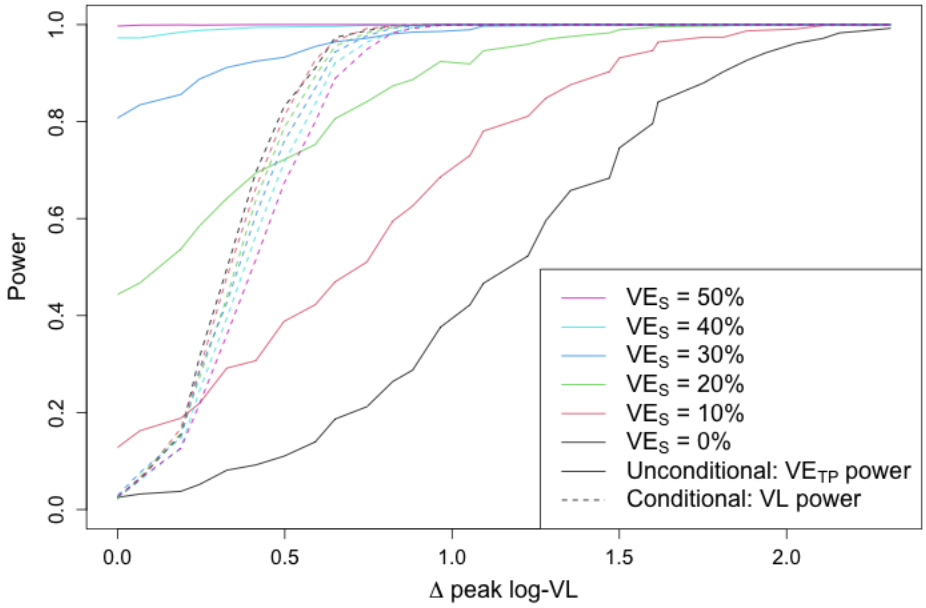

Figure A.1: Power to detect vaccine effect on log-VL AUC, conditional on infection, and $VE^C_{TP}(\tau)$ across a range of VE_S values.

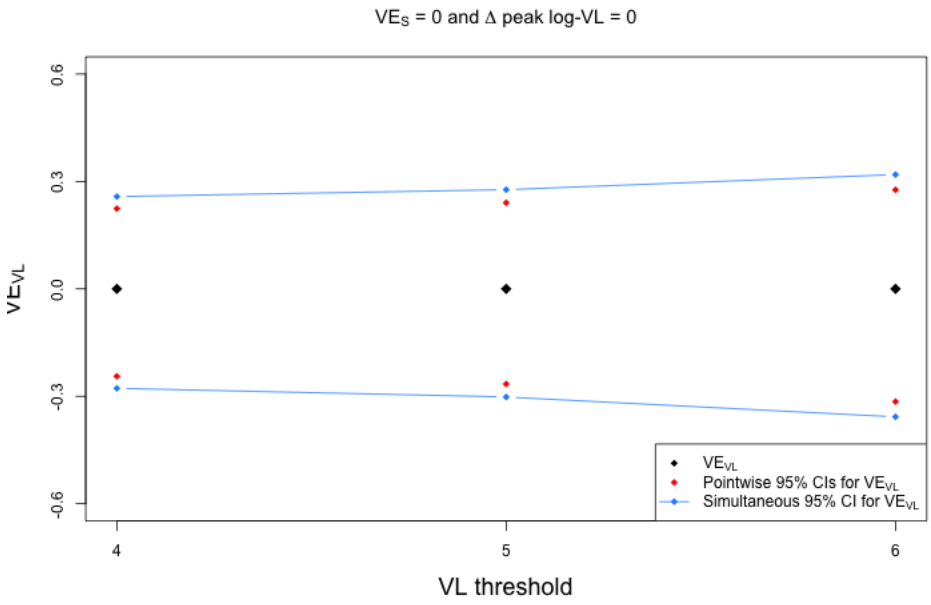

Figure A.2: Average 95% simultaneous confidence bands (blue lines) compared to average pointwise 95% confidence intervals (red dots) for $VE_{VL}(V, \tau)$ for all $V \in \{4, 5, 6\}$, based on 1,000 simulated trials with $VE_S = 0$, Δ peak log-VL $= 0$

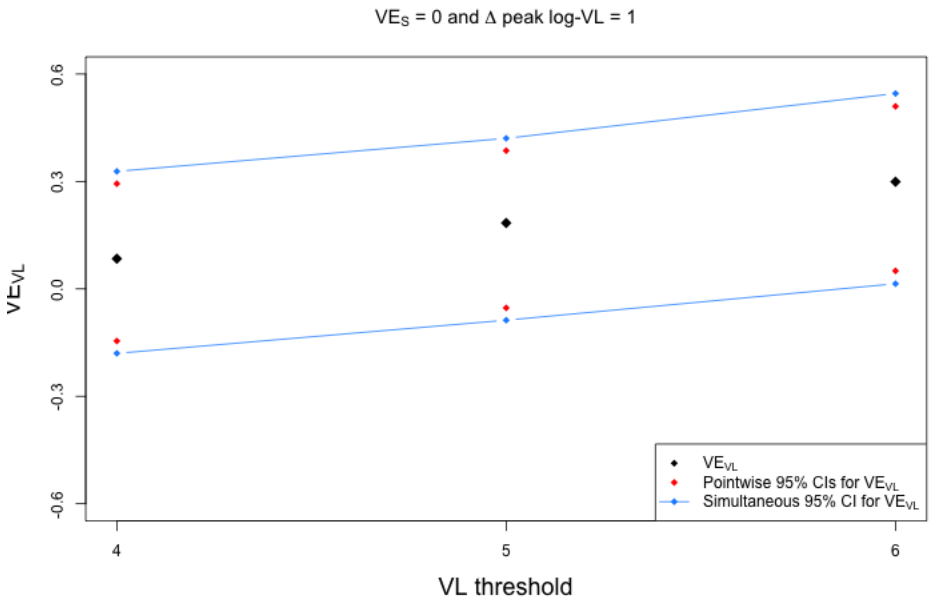

Figure A.3: Average 95% simultaneous confidence bands (blue lines) compared to average pointwise 95% confidence intervals (red dots) for $VE_{VL}(V, \tau)$ for all $V \in \{4, 5, 6\}$, based on 1,000 simulated trials with with $VE_S = 0$, Δ peak log-VL $= 1$

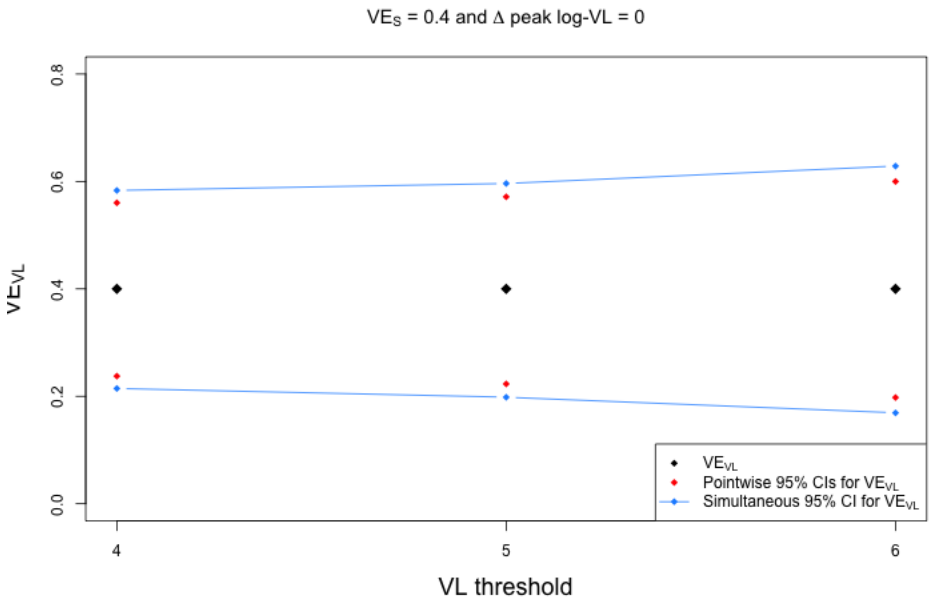

Figure A.4: Average 95% simultaneous confidence bands (blue lines) compared to average pointwise 95% confidence intervals (red dots) for $VE_{VL}(V, \tau)$ for all $V \in \{4, 5, 6\}$, based on 1,000 simulated trials with with $VE_S = 0.4$, Δ peak log-VL $= 0$