Analysis of Coronavirus Prevention Network COVID-19 vaccine efficacy trials: Do vaccine side effects predict better efficacy to prevent COVID-19? Statistical Analysis Plan

Yichen Wang

1 Introduction

Coronavirus Disease 2019 (COVID-19) pandemic, caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has had an unprecedented impact on global health, economies, and daily life since its outbreak in late 2019. The virus spreads primarily through respiratory droplets and has led to widespread morbidity and mortality, straining healthcare systems worldwide. Vaccination is considered an efficient approach to combat the spread of COVID-19. As of today, several COVID-19 vaccines based on different technologies are available, including mRNA-based vaccines (e.g., Pfizer-BioNTech and Moderna), viral vector vaccines (e.g., Johnson & Johnson and AstraZeneca), and protein subunit vaccines (e.g., Novavax and Sanofi), to meet the global need for vaccine, the wide geographic diversity of the pandemic, and the emergence of variants (Corey et al. [2020]). Vaccination against SARS-CoV-2 has been shown to be effective in reducing infections, hospitalizations, and mortality. Some widely used vaccines, namely Moderna, Johnson & Johnson, AstraZeneca, Novavax, and Sanofi, were developed in the six harmonized phase 3 COVID-19 vaccine efficacy trials conducted in public-private partnerships funded by NIAID and BARDA (USG Phase 3 program). The primary publications of all of the USG Phase 3 program can be seen in Baden et al. [2021], El Sahly et al. [2021], Sadoff et al. [2021, 2022], Falsey et al. [2021], Dunkle et al. [2022], Dayan et al. [2023b,a].

Correlates of protection (CoP) is an immune marker statistically correlated with vaccine efficacy (Plotkin and Gilbert [2012]). It can be used to reliably predict vaccine efficacy against clinically relevant endpoints and, thus, is widely pursued in vaccine research, including COVID-19. Some previous research has investigated the association between antibody immune markers and COVID-19 stems from the six harmonized phase 3 COVID-19 VE trials, as mentioned in the previous paragraph. This research was conducted by the NIAID- and BARDA-funded COVID-19 Vaccine Correlates of Protection Program (USG CoP Program). The immune correlates statistical analysis plan (SAP) specified statistical methods within each of the following statistical frameworks, all of which have been applied to some or all of the trial datasets: (1) nonparametric covariate-marginalized assessment of univariable marker correlates of risk (van der Laan et al. [2022]); (2) nonparametric and semiparametric univariable marker controlled vaccine efficacy causal analysis (Gilbert et al. [2022b]); (3) nonparametric univariable marker stochastic interventional vaccine efficacy causal analysis (Hejazi et al. [2021]); (4) nonparametric univariable/multivariable marker mediation analysis based on natural direct and indirect effects (Benkeser et al. [2021]), (5) nonparametric and semiparametric principal stratification correlates of vaccine efficacy analysis (Gilbert et al. [2020], Huang et al. [2022]); (5) nonparametric stochastic interventional vaccine efficacy analysis (Hejazi et al. [2021]); and (6) multivariable marker superlearning correlates of risk analysis that considered all of the measured antibody markers in the same machine learning analysis (Benkeser et al. [2023b]). All immune correlates analyses conducted by all methods supported both IgG binding antibodies to the Spike protein and neutralizing antibody titers as correlates of COVID-19 in vaccine recipients, and also supportive of these antibody markers for use as partially validated surrogate endpoints (Earle et al. [2021], Fong et al. [2022, 2023], Khoury et al. [2022], Benkeser et al. [2023a,b], Hejazi et al. [2023], Huang et al. [2023]). These blinded-phase correlates studies had scope (1) to study antibody markers measured 2–4 weeks post primary vaccination series in SARS-CoV-2 negative individuals (inferred to be never previously infected with SARS-CoV-2); (2) to study the primary study endpoint of virologically-confirmed symptomatic COVID-19 through to between 2.5 and 7 months post

primary vaccination series depending on the trial; (3) to study COVID-19 endpoints caused by ancestral and many pre-omicron variants. The paper "A COVID-19 milestone attained – a correlate of protection for vaccines" summarized the significance and impact of these results, as well as important open challenges that remain to be addressed with future correlates of protection research (Gilbert et al. [2022a]). The results to date have supported that binding antibodies to SARS-CoV-2 spike, binding antibodies to the spike receptor-binding domain (RBD), and pseudovirus neutralizing antibody titers ("ID50 titer") measured 2–4 weeks after the primary vaccination series are all significant inverse correlates of risk and supported to have value as partially valid surrogate endpoints (immune correlate of protection), with ID50 titer standing out as a consistent and highest quality independent correlate of protection across trials and vaccine platforms. These results have been impactful, with ID50 titer being used as a surrogate endpoint for many vaccine recommendation, authorization, and approval decisions.

The immune correlates results provided the insight that generally across the vaccine platforms, vaccine recipients with lowest antibody levels (near or below assay detection limits) were particularly vulnerable to failure of vaccination to protect, with impact to raise special concern for sub-populations with low frequencies of detectable neutralizing antibodies. Despite predicting the efficacy of the vaccines, safety is another big concern in vaccine development. Previous research has studied the relationship between immunogenicity and reactogenicity (solicited adverse reactions). The results in Siangphoe et al. [2023] suggest that solicited systemic adverse reactions (ARs) predict higher neutralizing antibody (nAb) geometric mean titers (GMTs) after a second mRNA-1273 injection. Another research studied the association between self-reported solicited ARs after receiving SARS-CoV-2 mRNA vaccines (Pfizer-BioNTech and Moderna) and serum concentration of IgG antibody against the spike protein (IgG spike antibody) among participants in Framingham Heart Study (FHS). The result suggests that systemic solicited ARs are associated with greater antibody response (Hermann et al. [2022]). Moreover, Dutcher et al. [2024] examined the association between solicited ARs and neutralizing antibody (nAB) response at 1 month and 6 months after the second vaccine dose among participants who received either BNT162b2 or mRNA-1273. The study found that chills, tiredness, feeling unwell, and headache after the second dose were each associated with higher nAB at 1 and 6 months after vaccination. The study also suggests a significant positive association between the number of distinct symptoms, vaccination-induced change in skin temperature and heart rate, and nAB at both follow-up time points. In contrast, the research conducted by Choi et al. [2023] among mRNA-1273 recipients found that neither patients experienced any solicited local nor systemic ARs after both injections were predictive of higher antibody response (IgG spike antibody). Only participants with a headache after the first injection showed significantly stronger antibody response after 8 weeks.

Despite the prior publications investigating the predictiveness of antibody markers on SARS-CoV-2 infection as well as the association between reactogenicity and antibody response, little is known about the direct association between reactogenicity and the clinically relevant endpoints (SARS-CoV-2 infection). Therefore, we aim to study the reactogenicity side effects as predictors of COVID-19 as well as exploring the connection of side effects to causal vaccine effects on COVID-19. This project will analyze multiple phase 3 trials, including the P3001 Moderna COVE trial, the P3004 Novavax PREVENT-19 trial, the P3005 Sanofi VAT00008 Stage 1 trial, and the P3005 Sanofi VAT00008 Stage 2 trial, in the US Government's Phase 3 COVID-19 Vaccine Efficacy Trials program. It will be another cross-protocol analysis of multiple phase 3 trials, with other research projects with this objective published in Lora et al. [2023], Theodore et al. [2023], Rick et al. [2023], Turley et al. [2023].

2 Overview and Objectives of Study Design

This project addresses the following objectives:

- Objective 1 (risk prediction and variable importance): To characterize associations of reactogenicity variables with future risk of COVID-19 among vaccine and placebo recipients separately
- Objective 2 (affect of vaccine-caused reactogenicity on COVID-19): To characterize the vaccine vs. placebo effect on COVID-19 in subgroups defined by reactogenicity outcome if assigned vaccine (principal stratification causal inference)

Addressing Objective 1 includes quantification of COVID-19 classification accuracy and variable importance analysis for estimating which reactogenicity variables have the greatest influence in predicting COVID-19.

Addressing Objective 2 focuses on placebo recipients with no reported reactogenicity, comparing vaccine efficacy in the subgroup with a given reactogenicity outcome if assigned vaccine (but not if assigned placebo) to vaccine efficacy in the subgroup without a given reactogenicity outcome under both randomization assignments.

Each objective is evaluated separately for the six treatment cohorts described in Section 3.

3 Study cohorts

For the Moderna and Novavax trials, the cohort analyzed is the protocol-specified per-protocol cohort which is the same cohort used for primary efficacy analysis, consisting of randomized baseline SARS-CoV-2 negative participants who received all planned immunizations without major protocol violations. In both cases, the per-protocol analysis set and baseline negative vs. positive are defined the same as in the protocol and primary papers, as also defined in Rick et al. [2023]. Note that for Moderna trial, the primary paper Baden et al. [2021] used the data cutoff on November 11, 2020 (interim analysis), while the primary paper El Sahly et al. [2021] used the data cutoff on March 26, 2021 (final analysis). In our work, we focus on the most update-to-date as in El Sahly et al. [2021]. Similarly, for Novavax trial, the primary paper Dunkle et al. [2022] used the data cutoff on April 19, 2021, while in our work, we focus on the data cutoff on September 27, 2021.

For both Sanofi stage 1 and Sanofi stage 2 trials, separate analyses are done for participants naive at both D01 and D22 in the modified full analysis set post-dose2 (mFAS-PD2) and for participants non-naive at either D01 or D22 in mFAS-PD2. The naive and non-naive cohorts are defined the same as in the protocols and primary analyses Dayan et al. [2023b] and Dayan et al. [2023a], involving serology testing data at D01 and D22. Participants with undetermined infection status were excluded. The data cut-off date we used for Sanofi trials is June 05, 2023. In addition, since reactogenicity data were only collected among participants in the reactogenicity subset (see more details in Section 5.1), all the analyses for Sanofi stage 1 and Sanofi stage 2 are implemented among participants who are in mFAS-PD2 and also in the reactogenicity subset. Since the total number of cases and the incidence rate are both low, additional analyses were conducted on the entire Sanofi Stage 2 cohort, regardless of the previous infection status. This includes participants from both the Sanofi Stage 2 naive analysis set and the Sanofi Stage 2 non-naive analysis set.

4 Endpoints

For Moderna and Sanofi trials, the primary endpoint is the protocol-specified primary endpoint, virologically-confirmed symptomatic COVID-19 (referring to as COVID-19), starting 14 days after dose 2 and throughout the blinded period of study follow-up. For the Novavax trial, the primary endpoint is the protocol-specified primary endpoint, virologically-confirmed symptomatic COVID-19, starting 7 days after dose 2 and throughout the blinded period of study follow-up.

5 Background on statistical analyses

5.1 Description of reactogenicity variables

In the P3001 Moderna COVE trial, solicited local and systemic ARs were monitored and recorded for each participant and for 7 days after each injection. Solicited ARs will be recorded daily using eDiaries (Baden et al. [2021], El Sahly et al. [2021]). The types of solicited ARs collected in this trial were arthralgia, chills, erythema, fatigue, fever, headache, myalgia, nausea, pain, swelling, and underarm gland swelling or tenderness.

In the P3004 Novavax PREVENT-19 trial, solicited ARs for each participant were recorded using eDiaries starting on the days of the vaccinations and for a total of 7 days after each vaccine injection (Dunkle et al.

[2022]). The types of solicited ARs collected in this trial were joint pain, redness, fatigue, fever, headache, muscle pain, nausea/vomiting, pain, swelling, malaise, and tenderness.

In the P3005 Sanofi VAT00008 Stage 1 trial, participants in the reactogenicity subset were collected for solicited ARs for 7 days after each vaccination using eDiaries, and 25% of the initial 8000 participants will be randomly assigned to the reactogenicity subset (Dayan et al. [2023b]). Similarly, in the P3005 Sanofi VAT00008 Stage 2 trial, participants in the reactogenicity subset were collected for solicited ARs for 7 days after each vaccination, and the first 4000 participants along with all participants above 60 years old will be assigned to the reactogenicity subset (Dayan et al. [2023a]). In both Sanofi Stage 1 trial and Sanofi Stage 2 trial, the types of solicited ARs collected were arthralgia, chills, erythema, fever, headache, myalgia, pain, swelling, and malaise.

The reactogenicity datasets for all these trials collect the worst severity grade and the duration of each type of solicited AR after each injection. For each vaccination, the worst severity grade of a certain AR is defined as the maximum severity grade of all reported severity grades of that AR among the 7 days after the injection; The duration of a certain AR is defined as the sum of duration of all reported duration of that AR among the 7 days after the injection. The severity grading of solicited ARs was based on grading scales modified from the FDA Toxicity Grading Scale for Healthy Adult and Adolescent Volunteers Enrolled in Preventive Vaccine Clinical Trials. In addition, the datasets also record the worst severity grade and the duration of three composite solicited ARs: any solicited AR, any solicited local AR, and any solicited systemic AR. The worst severity grade of any solicited AR after each injection is defined as the maximum worst severity grade of all types of solicited ARs collected in the trial, and the duration is denoted as the sum of duration of all types of solicited ARs collected. The worst severity grade and duration for any solicited local AR and any solicited systemic AR were defined similarly as for any solicited AR. In this project, the reactogenicity variables we are interested in are the occurrence of any solicited AR, the occurrence of any solicited local AR, the occurrence of any solicited systemic AR, and fever for each vaccination. The occurrence of certain AR after each vaccination is denoted as an indicator variable where 1 refers to either having duration for at least 1 day or having worst severity grade greater than or equal to grade 1, and 0 refers to having 0 duration and worst severity grade 0. The definition of worst severity grade 0 can be seen in Section 6.3.

5.2 Description of potential confounders of the effect of reactogenicity variables on study endpoints

The confounders adjusted for in the analyses to examine the association between reactogenicity variables and future risk of COVID-19 for each trial are adapted from Gilbert et al. [2022c], Fong et al. [2023].

In the Moderna trial, the baseline factors used for confounder adjustment are baseline risk scores, community of color, and risk of exposure to SARS-CoV-2 as per OSHA (Occupational Safety and Health Administration). The baseline risk scores adjusted for were built using data only from baseline negative per-protocol placebo recipients of the Moderna trial (Gilbert et al. [2022c]).

In the Novavax trial, the baseline factors used for confounder adjustment are baseline risk scores, geographic region, and risk of exposure to SARS-CoV-2 as per OSHA. The baseline risk scores adjusted for were built from ensemble super-learning of the placebo arms of the Novavax trial, restricting to the US participants in the Per-protocol cohort (Fong et al. [2023]).

For the Sanofi trial, the baseline risk score and the force of infection (FOI) score are adjusted. The baseline risk scores were developed based on the placebo arm pooling over the Stage 1 and Stage 2 trials, developing a separate risk score for Naive Per-protocol cohort and Non-naive Per-protocol cohort participants. The definition of FOI score can be seen in the Sanofi immune correlates SAP (Peter B. Gilbert [2024]).

6 Statistical analyses

6.1 Variable Importance: Prediction of COVID-19

We will implement the nonparametric variable importance technique from Wolock et al. [2023] to predict the risk of COVID-19 infection in each of the six study cohorts. The analysis will be done for each injection and

for each treatment arm. We will use the landmark time area under the ROC curve (AUC) for predictiveness measure and implement the analysis using cross-fitting with global survival stacking for nuisance estimation. The algorithms included in the global survival stacking are SL.mean, SL.glm, and SL.glmnet. The results of the marginal variable importance measures (VIM) analysis will be presented. For this analysis, the full feature vector includes the baseline characteristics plus all other features or feature groups of interest, and the reduced feature vector includes only the baseline characteristics. The feature or feature of group included in each trial can be seen in Table 1 - Table 3. The marginal VIM analysis adjusts for enrollment date where it is denoted as an indicator variable based on the median of the enrollment date among all participants in the certain analysis set. R code available at Charles Wolock's Github repository *surv_vim_supplementary* using *survML* R package is used to implement the analyses.

The analysis will be conducted for each study cohort, treatment arm, and injection. In addition, we require each analysis set to have at least 15 COVID-19 cases. For each type of reactogenicity endpoint, we also require at least one case for each level of the reactogenicity variable. Therefore, the vaccine recipients in the Novavax trial are excluded from this analysis (14 cases). The occurrence of fever is excluded from the Moderna post-dose one analysis, Novavax post-dose one and post-dose two analysis, and Sanofi Stage 2 Naive post-dose one analysis.

Group	Feature(s)
1	occurrence of any type of solicited AR within 7 days after each injection
2	occurrence of any solicited local AR within 7 days after each injection
3	occurrence of any solicited systemic AR within 7 days after each injection
4	occurrence of fever within 7 days after each injection
5	Baseline characteristics: baseline risk score, community of color, risk of exposure to SARS-CoV-2 as per OSHA

Table 1: Groups of Features included in the Moderna Trial

Group	Feature(s)
1	occurrence of any type of solicited AR within 7 days after each injection
2	occurrence of any solicited local AR within 7 days after each injection
3	occurrence of any solicited systemic AR within 7 days after each injection
4	occurrence of fever within 7 days after each injection
5	Baseline characteristics: baseline risk score, risk of exposure to SARS-CoV-2 as per OSHA

Table 2: Groups of Features included in the Novavax Trial

Group	Feature(s)
1	occurrence of any type of solicited AR within 7 days after each injection
2	occurrence of any solicited local AR within 7 days after each injection
3	occurrence of any solicited systemic AR within 7 days after each injection
4	occurrence of fever within 7 days after each injection
5	Baseline characteristics: baseline risk score, force of infection (FOI)

Table 3: Groups of Features included in the Sanofi Trials

6.1.1 Definition of the landmark times

Each of the analyses estimates a conditional survival curve $P(T \leq t_0|X)$ with T the time between the time origin and the COVID-19 endpoint and t_0 the last time by which a COVID-19 endpoint is counted, such

that a definition of t_0 is needed for each analysis. For the analyses of the Sanofi vaccines, t_0 is set to 181 days after the time origin for the analyses of post-dose two reactogenicity. For post-dose one analyses, t_0 is defined as the same value plus 21 days after the time origin for post-dose one analyses, therefore 202 days after the time origin. These choices mean that for each analysis approximately 6 months follow-up post dose two is used. For post-dose one analyses, the time origin is 7 days post dose 1. For post-dose two analyses, the time origin is 7 days post dose 2. The landmark times for each injection are the same across the Sanofi trials.

For the analyses of the Moderna trial, for the post-dose two reactogenicity analyses, t_0 is defined as the latest day of follow-up after the time origin such that at least 5 COVID-19 endpoints occurred in each of the vaccine and placebo arms. In general, follow-up only includes follow-up prior to unblinding, such that (of course) at least 5 COVID-19 endpoints must occur during the blinded phase of follow-up. Next, for the Moderna post-dose one reactogenicity analyses, t_0 is defined using the same algorithm as that for post two analyses. The same algorithm is used to define t_0 for post-dose two and post-dose one analyses in Novavax trial. Thus a total of twelve landmark times t_0 are defined, six for each injection. Table 4 lists the values of the twelve landmark times.

Vaccine Cohort	Post Dose One (days)	Post Dose Two (days)
Moderna	144	115
Novavax	67	46
Sanofi Stage 1 Naive	202	181
Sanofi Stage 1 Non-Naive	202	181
Sanofi Stage 2 Naive	202	181
Sanofi Stage 2 Non-Naive	202	181

Table 4: Landmark time since time origin for each trial. The time origin for post-dose one analyses is 7 days post dose 1. The time origin for post-dose two analyses is 7 days post dose 2.

6.2 Principal stratification: Assessment of vaccine efficacy in subgroups defined by reactogenicity potential outcomes under assignment to vaccine and placebo

Let R be a binary reactogenicity variable measured after dose one or after dose two, with $R(a)$ the potential outcome under assignment to vaccine ($a = 1$) and under assignment to placebo ($a = 0$). Let Y be the indicator of COVID-19 after the time point that R is measured, with $Y(a)$ the potential outcomes for $a = 0, 1$. We consider causal estimands of interest defined exactly as in Gilbert et al. [2020], with only difference they used notation S for the intermediate variable instead of R. In their notation, our objective is to estimate VE(1,0)=CEP(1,0) and VE(0,0)=CEP(0,0) where:

- VE(1,0) $:= 1 - P(Y(1) = 1|R(1) = 1, R(0) = 0)/P(Y(0) = 1|R(1) = 1, R(0) = 0)$ is vaccine efficacy in the subgroup with $R(1) = 1$ and $R(0) = 0$
- VE(0,0) $:= 1 - P(Y(1) = 1|R(1) = 0, R(0) = 0)/P(Y(0) = 1|R(1) = 0, R(0) = 0)$ is vaccine efficacy in the subgroup with $R(1) = 0$ and $R(0) = 0$.

That is, VE(1,0) is vaccine efficacy in the subgroup with vaccine-caused reactogenicity, and VE(0,0) is vaccine efficacy in the subgroup without reactogenicity under both treatment assignments. It is also of interest to estimate ratios of strata-specific vaccine efficacy, specifically, the relative risk ratio [1-VE(0,0)]/[1-VE(1,0)] which is interpreted as the degree (on the multiplicative scale) to which vaccine-reduction of COVID-19 is better for the (1,0) subgroup than for the (0,0) subgroup (i.e., quantifying a vaccine-protection advantage of having vaccine-caused reactogenicity).

For estimation and inference, much of the available literature on principal stratification methods focuses on the setting where $R(0)$ is constant. While the NEE-VB method described in Gilbert et al. [2020] does apply to estimate the three parameters VE(1,0), VE(0,0), and [1-VE(0,0)]/[1-VE(1,0)], the R package *psbinary* only includes an implementation for applications with $R(0)$ constant.

The NEE-CB method is applied to the datasets for both post-dose one and post-dose two analyses where NEE refers to the version of the methods that assume "no early efficacy" of the vaccine vs. placebo on COVID-19 by the time reactogenicity is measured, and CB standards for "constant biomarker", which in our context means that $R(0)$ takes the same constant value for all placebo recipients: $R(0) = 0$ (i.e., $P(R(0) = 0) = 1$). Given this assumption, only placebo recipients with no reported reactogenicity are included in the analyses. The NEE-CB method is used for post-dose one analysis because the vaccines have no efficacy over the first 7 days post-dose one. The NEE-CB method is also used for post-dose two analysis since no participant encountered symptomatic COVID-19 infection before 7 days post dose two. The NEE-CB method requires one user-specified sensitivity parameter, β_0, which can represent different degrees of post-randomization selection bias when set to different ranges. The details of setting this sensitivity parameter can be seen in Section 6.2.1.

Now, it would also be interesting to estimate vaccine efficacy parameters for subgroups defined by $R(1)$ only, that is

$\text{VE}(1) := 1 - P(Y(1) = 1|R(1) = 1)/P(Y(0) = 1|R(1) = 1)$ and

$\text{VE}(0) := 1 - P(Y(1) = 1|R(1) = 0)/P(Y(0) = 1|R(1) = 0)$.

This would be possible, based on the fact that the subgroup $R(1) = 1$ is the union of the two subgroups $\{R(1) = 1, R(0) = 0\}$ and $\{R(1) = 1, R(0) = 1\}$, and the $R(1) = 0$ subgroup equals the $\{R(1) = 0, R(0) = 0\}$ subgroup, leveraging the monotonicity assumption that $R(1) = 0$ implies $R(0) = 0$. In other words, results on estimates of the conditional probability pieces $P(Y(a) = 1|R(1) = j, R(0) = k)$ used for estimation of VE(1,0), and VE(0,0) yield plug-in estimates of VE(1) and VE(0). However, the available R package does not include this functionality, and because considerable work would be required, for now the SAP does not include estimation of VE(1) and VE(0).

The analyses will be implemented for study cohorts in which for both post-dose one and post-dose two analyses, at least 15 vaccine cases occurred before the certain landmark time for a given reactogenicity value and at least one vaccine case occurred for each level of the reactogenicity variable. The analyses will be implemented for study cohorts in which for a given dosing, there are at least 15 vaccine cases, defined as vaccine cases occurred before the certain landmark time, for the given reactogenicity variable and at least one vaccine case for each level of the reactogenicity variable. Therefore, participants in the Novavax trial and Sanofi Stage 1 Non-naive cohort are excluded from this analysis since only 8 and 13 vaccine recipients had symptomatic COVID-19 infection before a certain landmark time, respectively. In addition, we only consider implementing this analysis among trials that have vaccine efficacy of at least 30%, which excludes the Sanofi Stage 1 Naive cohort from this analysis (Dayan et al. [2023b]). The occurrence of fever is excluded from Moderna post-dose one analysis, Sanofi Stage 2 Naive post-dose one analysis, and Sanofi Stage 2 Non-naive post-dose two analysis.

6.2.1 Implementation details for the NEE-CB principal stratification analysis

The data analysis is done for the same overall binary reactogenicity variables studied for the prediction objective, that measure any reactogenicity, local reactogenicity, systemic reactogenicity, or fever. As for the prediction analyses, the analyses are done for post-dose one reactogenicity and for post-dose two reactogenicity. The cohorts, landmark times, and included COVID-19 endpoints are the same as for the prediction analyses. The same decisions made for accounting for missing reactogenicity variables for the prediction objectives are carried over to the principal stratification objective. R code analyze_NEE.Rd for the *psbinary* R package available at Bryan Blette's Github repository is used to carry out the data analysis.

To estimate and make inference, sensitivity analyses are conducted with the sensitivity parameter β_0 set to different ranges. First, the sensitivity parameter is set to vary from $\log(1.0) = 0$ to $-\log(1.0) = 0$ (none robustness), and the point estimate and 95% CI are calculated. Then, the sensitivity parameter is set to vary from $\log(0.75)$ to $-\log(0.75)$ (median robustness) and $\log(0.5)$ to $-\log(0.5)$ (high robustness), and the ignorance interval (IGI) and 95% estimated uncertainty interval (EUI) are calculated. More details about the definition of these parameters can be seen in Gilbert et al. [2020].

6.3 Missing data

Participants without primary endpoint documented were treated as censoring. The censoring rules for each trial can be seen in the primary papers Baden et al. [2021], El Sahly et al. [2021], Dunkle et al. [2022], Dayan et al. [2023b,a].

The manipulation of missing values in the risk of exposure to SARS-CoV-2 as per OSHA in the Moderna trial can be seen in Theodore et al. [2023]. The missing values in baseline risk scores in each analysis set are imputed by the median of baseline risk scores, and the force of infection (FOI) in Sanofi trials are imputed by the median of the FOI. The missing values in community of color are impuated by the most frequent category. There's no missing value in other baseline variables.

In the Moderna trial, having reactogenicity with severity grade 0 is defined as not having the reactogenicity or having more mild symptoms than grade 1 according to the grading scale (Baden et al. [2021], El Sahly et al. [2021]). For participants who have (worst) severity grade 0 for a certain reactogenicity in the reactogenicity dataset, the corresponding duration of this reactogenicity was coded as missing. In this case, we convert the missing duration to 0 since the patient did not have a certain reactogenicity or the symptom was too mild. In the Novavax trial, the severity of reactogenicity was graded from grade 1 to grade 4 (Dunkle et al. [2022]). Therefore, in the reactogenicity dataset, only grade 0 was coded for any solicited AR which means not having reactogenicity, whereas for all other types of reactogenicity recorded in the dataset, not having that specific type of reactogenicity was coded as missing. Similarly to the manipulation applied to the Moderna reactogenicity dataset, for participants in the Novavax reactogenicity dataset who have any solicited AR with the worst severity grade of 0, the missing values of worst severity grade and duration for all other types of reactogenicity are converted to 0. In the Sanofi trials, both severity grade 0 and duration 0 are included in the reactogenicity dataset, so no further manipulation is needed. In addition, there are several cases in each study cohort that do not have any solicited systemic AR and are missing the values of fever. In this case, the values of fever are coded as 0. Other than the missingness mentioned, there remains a small amount of missing data in the reactogenicity dataset for each trial, where those participants do not have any records of reactogenicity, meaning all reactogenicity variables are missing in the dataset. All other missing cases not mentioned here are kept as missing.

Furthermore, while the Moderna and Novavax trials collected reactogenicity data from all participants, our reactogenicity dataset includes only a subset of these individuals. Specifically, it excludes participants who missed duration data for both injections, and we lack additional information regarding their severity grades. Since we do not have sufficient evidence of potential factors that can reasonably predict the missing reactogenicity variables, we decide to proceed with all the analyses using complete-case analysis.

6.4 Multiplicity considerations

We do not use multiplicity correction techniques to control the type I error rate for multiple hypothesis testing due to the small number of endpoints (COVID-19) in our trials. We do not want to lose potential signals in the pursuit of higher power. Therefore, we acknowledge that our power is low, and the results may be considered hypothesis-generating, warranting further replications of our analyses.

References

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.

D Benkeser, I Diaz, and J Ran. Inference for natural mediation effects under case-cohort sampling with applications in identifying COVID-19 vaccine correlates of protection. *arXiv*, arXiv:2103.02643 [q-bio.QM] [Preprint] March 5 2021. Cited 13 March 2021. Available from https://arxiv.org/pdf/2103.02643.pdf, 2021.

David Benkeser, Youyi Fong, Holly E Janes, Elizabeth J Kelly, Ian Hirsch, Stephanie Sproule, Ann Marie Stanley, Jill Maaske, Tonya Villafana, Christopher R Houchens, et al. Immune correlates analysis of a phase 3 trial of the AZD1222 (ChAdOx1 nCoV-19) vaccine. *npj Vaccines*, 8(1):36, 2023a.

David Benkeser, David C Montefiori, Adrian B McDermott, Youyi Fong, Holly E Janes, Weiping Deng, Honghong Zhou, Christopher R Houchens, Karen Martins, Lakshmi Jayashankar, et al. Comparing antibody assays as correlates of protection against COVID-19 in the COVE mRNA-1273 vaccine efficacy trial. *Science translational medicine*, 15(692):eade9078, 2023b.

Min Joo Choi, Jung Yeon Heo, Yu Bin Seo, Young Kyung Yoon, Jang Wook Sohn, Ji Yun Noh, Hee Jin Cheong, Woo Joo Kim, Ju-yeon Choi, Young Jae Lee, et al. Predictive value of reactogenicity for anti-sars-cov-2 antibody response in mrna-1273 recipients: a multicenter prospective cohort study. *Vaccines*, 11(1):120, 2023.

Lawrence Corey, John R Mascola, Anthony S Fauci, and Francis S Collins. A strategic approach to covid-19 vaccine r&d. *Science*, 368(6494):948–950, 2020.

Gustavo H Dayan, Nadine Rouphael, Stephen R Walsh, Aiying Chen, Nicole Grunenberg, Mary Allen, Johannes Antony, Kwaku Poku Asante, Amit Suresh Bhate, Tatiana Beresnev, et al. Efficacy of a bivalent (d614+ b. 1.351) sars-cov-2 recombinant protein vaccine with as03 adjuvant in adults: a phase 3, parallel, randomised, modified double-blind, placebo-controlled trial. *The Lancet Respiratory Medicine*, 11(11): 975–990, 2023a.

Gustavo H Dayan, Nadine Rouphael, Stephen R Walsh, Aiying Chen, Nicole Grunenberg, Mary Allen, Johannes Antony, Amit Suresh Bhate, Tatiana Beresnev, Matthew I Bonaparte, et al. Efficacy of a monovalent (d614) sars-cov-2 recombinant protein vaccine with as03 adjuvant in adults: a phase 3, multi-country study. *Eclinicalmedicine*, 64, 2023b.

Lisa M Dunkle, Karen L Kotloff, Cynthia L Gay, Germán Áñez, Jeffrey M Adelglass, Alejandro Q Barrat Hernández, Wayne L Harper, Daniel M Duncanson, Monica A McArthur, Diana F Florescu, et al. Efficacy and safety of nvx-cov2373 in adults in the united states and mexico. *New England Journal of Medicine*, 386(6):531–543, 2022.

Ethan G Dutcher, Elissa S Epel, Ashley E Mason, Frederick M Hecht, James E Robinson, Stacy S Drury, and Aric A Prather. Covid-19 vaccine side effects and long-term neutralizing antibody response: A prospective cohort study. *Annals of Internal Medicine*, 2024.

Kristen A Earle, Donna M Ambrosino, Andrew Fiore-Gartland, David Goldblatt, Peter B Gilbert, George R Siber, Peter Dull, and Stanley A Plotkin. Evidence for antibody as a protective correlate for covid-19 vaccines. *Vaccine*, 39(32):4423–4428, 2021.

Hana M El Sahly, Lindsey R Baden, Brandon Essink, Susanne Doblecki-Lewis, Judith M Martin, Evan J Anderson, Thomas B Campbell, Jesse Clark, Lisa A Jackson, Carl J Fichtenbaum, et al. Efficacy of the mrna-1273 sars-cov-2 vaccine at completion of blinded phase. *New England Journal of Medicine*, 385(19): 1774–1785, 2021.

Ann R Falsey, Magdalena E Sobieszczyk, Ian Hirsch, Stephanie Sproule, Merlin L Robb, Lawrence Corey, Kathleen M Neuzil, William Hahn, Julie Hunt, Mark J Mulligan, et al. Phase 3 safety and efficacy of azd1222 (chadox1 ncov-19) covid-19 vaccine. *New England Journal of Medicine*, 385(25):2348–2360, 2021.

Youyi Fong, Adrian B McDermott, David Benkeser, Sanne Roels, Daniel J Stieh, An Vandebosch, Mathieu Le Gars, Griet A Van Roey, Christopher R Houchens, Karen Martins, et al. Immune correlates analysis of the ensemble single ad26. cov2. s dose vaccine efficacy clinical trial. *Nature microbiology*, 7(12):1996–2010, 2022.

Youyi Fong, Yunda Huang, David Benkeser, Lindsay N Carpp, Germán Áñez, Wayne Woo, Alice McGarry, Lisa M Dunkle, Iksung Cho, Christopher R Houchens, et al. Immune correlates analysis of the PREVENT-19 COVID-19 vaccine efficacy clinical trial. *Nature communications*, 14(1):331, 2023.

Peter B Gilbert, Bryan S Blette, Bryan E Shepherd, and Michael G Hudgens. Post-randomization biomarker effect modification analysis in an hiv vaccine clinical trial. *Journal of Causal Inference*, 8(1):54–69, 2020.

Peter B Gilbert, Ruben O Donis, Richard A Koup, Youyi Fong, Stanley A Plotkin, and Dean Follmann. A Covid-19 milestone attained—a correlate of protection for vaccines. *New England Journal of Medicine*, 387(24):2203–2206, 2022a.

Peter B Gilbert, Youyi Fong, Avi Kenny, and Marco Carone. A controlled effects approach to assessing immune correlates of protection. *Biostatistics*, page kxac024, 2022b.

Peter B Gilbert, David C Montefiori, Adrian B McDermott, Youyi Fong, David Benkeser, Weiping Deng, Honghong Zhou, Christopher R Houchens, Karen Martins, Lakshmi Jayashankar, et al. Immune correlates analysis of the mRNA-1273 COVID-19 vaccine efficacy clinical trial. *Science*, 375(6576):43–50, 2022c.

N Hejazi, MJ van der Laan, HE Janes, PB Gilbert, and DC Benkeser. Efficient nonparametric inference on the effects of stochastic interventions under two-phase sampling, with applications to vaccine efficacy trials. *Biometrics*, 77(4):1241–1253, 2021.

Nima S Hejazi, Xiaoying Shen, Lindsay N Carpp, David Benkeser, Dean Follmann, Holly E Janes, Lindsey R Baden, Hana M El Sahly, Weiping Deng, Honghong Zhou, et al. Stochastic interventional approach to assessing immune correlates of protection: Application to the cove messenger rna-1273 vaccine trial. *International Journal of Infectious Diseases*, 137:28–39, 2023.

Emilia A Hermann, Benjamin Lee, Pallavi P Balte, Vanessa Xanthakis, Beth D Kirkpatrick, Mary Cushman, and Elizabeth Oelsner. Association of symptoms after covid-19 vaccination with anti–sars-cov-2 antibody response in the framingham heart study. *JAMA network open*, 5(10):e2237908–e2237908, 2022.

Ying Huang, Yingying Zhuang, and Peter Gilbert. Sensitivity analysis for evaluating principal surrogate endpoints relaxing the equal early clinical risk assumption. *The Annals of Applied Statistics*, 16(3):1774–1794, 2022.

Ying Huang, Nima S Hejazi, Bryan Blette, Lindsay N Carpp, David Benkeser, David C Montefiori, Adrian B McDermott, Youyi Fong, Holly E Janes, Weiping Deng, et al. Stochastic interventional vaccine efficacy and principal surrogate analyses of antibody markers as correlates of protection against symptomatic covid-19 in the cove mrna-1273 trial. *Viruses*, 15(10):2029, 2023.

David S Khoury, Timothy E Schlub, Deborah Cromer, Megan Steain, Youyi Fong, Peter B Gilbert, Kanta Subbarao, James A Triccas, Stephen J Kent, and Miles P Davenport. Correlates of protection, thresholds of protection, and immunobridging in sars-cov-2 infection. *medrxiv*, pages 2022–06, 2022.

Alfredo J Mena Lora, Jessica E Long, Yunda Huang, Lindsey R Baden, Hana M El Sahly, Dean Follmann, Paul Goepfert, Glenda Gray, Beatriz Grinsztejn, Karen Kotloff, et al. Rapid development of an integrated network infrastructure to conduct phase 3 covid-19 vaccine trials. *JAMA network open*, 6(1):e2251974–e2251974, 2023.

Youyi Fong David Benkeser Jia Jin Kee Yiwen Lu Chenchen Yu Bhavesh Borate Marco Carone Lindsay N. Carpp Nima S. Hejazi Michal Juraska Li Li Lars van der Laan Brian D. Williamson Dean Follmann Peter B. Gilbert, Ying Huang. Statistical analysis plan for assessing immune orrelates in the vat08 trial of the sanofi-gsk covid-19 vaccines: Stages 1 and 2. Technical report, Fred Hutchinson Cancer Reserach Center, 2024.

SA Plotkin and Peter B Gilbert. Nomenclature for immune correlates of protection after vaccination. *Clinical Infectious Diseases*, 54:1615–1617, 2012. PMCID: PMC3348952.

Anne-Marie Rick, Matthew B Laurens, Ying Huang, Chenchen Yu, Thomas CS Martin, Carina A Rodriguez, Christina A Rostad, Rebone M Maboa, Lindsey R Baden, Hana M El Sahly, et al. Risk of covid-19 after natural infection or vaccination. *EBioMedicine*, 96, 2023.

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.

Jerald Sadoff, Glenda Gray, An Vandebosch, Vicky Cárdenas, Georgi Shukarev, Beatriz Grinsztejn, Paul A Goepfert, Carla Truyers, Ilse Van Dromme, Bart Spiessens, et al. Final analysis of efficacy and safety of single-dose ad26. cov2. s. *New England Journal of Medicine*, 386(9):847–860, 2022.

Uma Siangphoe, Lindsey R Baden, Hana M El Sahly, Brandon Essink, Kashif Ali, Gary Berman, Joanne E Tomassini, Weiping Deng, Rolando Pajon, Roderick McPhee, et al. Associations of immunogenicity and reactogenicity after severe acute respiratory syndrome coronavirus 2 mrna-1273 vaccine in the cove and teencove trials. *Clinical Infectious Diseases*, 76(2):271–280, 2023.

Deborah A Theodore, Angela R Branche, Lily Zhang, Daniel S Graciaa, Madhu Choudhary, Timothy J Hatlen, Raadhiya Osman, Tara M Babu, Samuel T Robinson, Peter B Gilbert, et al. Clinical and demographic factors associated with COVID-19, Severe COVID-19, and SARS-CoV-2 infection in adults: A secondary cross-protocol analysis of 4 randomized clinical trials. *JAMA Network Open*, 6(7):e2323349–e2323349, 2023.

Christine B Turley, Trevon Fuller, Lisa J Sanders, Hyman Scott, Amaran Moodley, Amanda Woodward Davis, Brett Leav, Jacqueline Miller, Kathryn Schoemaker, An Vandebosch, et al. Modifiers of covid-19 vaccine efficacy: results from four covid-19 prevention network efficacy trials. *Vaccine*, 41(33):4899–4906, 2023.

Lars van der Laan, Wenbo Zhang, and Peter B Gilbert. Nonparametric estimation of the causal effect of a stochastic threshold-based intervention. *Biometrics*, pages 1–22, 2022.

Charles J Wolock, Peter B Gilbert, Noah Simon, and Marco Carone. Nonparametric variable importance for time-to-event outcomes with application to prediction of hiv infection. *arXiv preprint arXiv:2311.12726*, 2023.