Non-Interventional Study

C4591065

Effectiveness of BNT162b2 formulations using state vaccine registry and insurance claims data

Aim 2

Statistical Analysis Plan

(SAP)

Version: 1.0

Final Author: PPD (PPD
Pfizer Inc., New York, NY, USA)

First draft by: PPD (PPD **Pfizer Inc, New York, NY, USA)**

Date: 10-Jun-2024

TABLE OF CONTENTS

1. AMENDMENTS FROM PREVIOUS VERSION(S)

NA. First version.

2. LIST OF ABBREVIATIONS

Abbreviation	Definition
aHR	Adjusted hazard ratio
CDC	Centers for Disease Control and Prevention
CI	Confidence interval
CPT	Current Procedural Terminology®
CVX	Clinical Vaccines Administered code
EUA	Emergency use authorization
FDA	U.S. Food and Drug Administration
HV	HealthVerity
ICD-10-CM	International Classification of Diseases, Tenth Revision, Clinical Modification
ICU	Intensive care unit
NDC	National Drug Codes
RSV	Respiratory syncytial virus
SAP	Statistical Analysis Plan
SMD	Standardized mean differences
NPI	National provider identifier
VE	Vaccine effectiveness

Note: in this document, any text taken directly from the non-interventional (NI) study protocol is ***italicized***.

3. RATIONALE AND BACKGROUND

Until recently, vaccine registries were not available for de-identification and linkage to secondary claims-based databases. However, through a collaboration with HealthVerity, a leading supplier of secondary health data, the State of California and the State of Louisiana have made their vaccine registries available for linkage for research purposes. California and Louisiana deidentify the registries, using HealthVerity's software to convert each patient's protected health information to a unique HealthVerity ID which can then be linked to the full HealthVerity database to gather clinical history and outcomes.

The purpose of this document is to detail the statistical methods and analyses that will be used in the HealthVerity claims database linked to the State of California/Louisiana's vaccine registries to address Aim 2 of the study protocol. Separate SAPs will be produced to address the other aims of the protocol.

4. STUDY OBJECTIVES AND HYPOTHESES

The objective of Aim 2 of this study is to *assess the vaccine effectiveness* (VE) *of BNT162b2 formulations in pediatrics against medically attended COVID-19, non-COVID-19 respiratory infections, mortality and healthcare resource utilization by age and adapted vaccine formulation.*

This study is descriptive in nature and as such does not include any pre-specified hypotheses.

5. RESEARCH METHODS

5.1. Study Design

This will be a retrospective cohort study using data from patients who were California or Louisiana residents in the HealthVerity claims database.

Cohort identification will begin from the date of Emergency Use Authorization (EUA) or U.S Food and Drug Administration (FDA) *approval for the BNT162b2 formulation. Study follow-up will occur from FDA EUA or approval date until COVID-19 diagnosis or other outcome of interest (hospitalization, emergency department visit, outpatient encounter, end of medical and/or pharmacy enrollment, or death [according to the mortality database]).*

The authorization/approval dates for each BNT162b2 formulation that will be used in this study are provided in Table 5.1.

Table 5.1 Timeline of Key BNT162b2 Regulatory Milestones Applicable to Pediatrics

Vaccine formulation	Age group	Event
BNT162b2 monovalent XBB.1.5 COVID-19 vaccine	Age 12 and older	September 11, 2023: U.S. Food and Drug Administration (FDA) approval
	Age 6 months to 11 years	September 11, 2023: FDA emergency use authorization (EUA)
BNT162b2 bivalent BA.4/5 COVID-19 vaccine	Age 12 and older	August 31, 2022: FDA EUA
	Age 5 to 11	October 12, 2022: FDA EUA
	Age 6 months to 4 years	December 8, 2022: FDA EUA

All persons will be considered unvaccinated from date of FDA date of vaccine authorization/approval until, where applicable, 13 days after receipt of vaccine. In order to allow for time after dose for meaningful immunologic protection to accrue, persons will be considered vaccinated 14 days after vaccination.

The following data are available, from 01 December 2018, with varying look-back periods as outlined below to define patient's characteristics, clinical history, risk factors, and healthcare utilization.

The outcomes of interest are medically attended COVID-19, non-COVID-19 respiratory infections, mortality and healthcare resource utilization and are described in Table 5.2.

*Main summary measure: VE = (1 – adjusted hazard ratio [aHR]) * 100, comparing vaccinated to unvaccinated persons*

5.2. Study Population

Patients who were California or Louisiana residents in the HealthVerity claims database will be defined using the following hierarchical definition:

1. *Persons who have their enrollment file record indicating a patient location of California or Louisiana in HealthVerity claims enrollment file.*

2. *Persons who have a patient state location other than California or Louisiana in HealthVerity claims enrollment but have one or more records in the California or Louisiana Immunization Registry. Examples include people who previously lived in Texas but moved to California and received a vaccine, or persons who crossed state lines for a vaccine and remain in their non-California state in claims data.*

Patients who meet the following criteria will be included in the analysis:

5.2.1. Inclusion Criteria

1. *Born in 2006-2023 (as age 0-17 years in 2023, owing to the data providing year but not month or day of birth).*

2. *At least six months of pharmacy and medical enrollment in HealthVerity prior to index date (FDA EUA date per age group vaccine). Six months was selected to allow for analysis to include children age <1yr. A gap of up to 30 days will be allowed.*

3. *Have been a State of California/Louisiana resident for at least six months prior to index date (FDA EUA date per age group and vaccine).* A gap of up to 30 days will be allowed.

5.2.2. Exclusion Criteria

1. *Currently pregnant (*with a code list for current pregnancy and *9 months lookback) or immunocompromised (*as per McGrath, Malhotra, Miles et al 2023*) at the time of study index* (definitions are provided in the Protocol Annex)

2. *Individuals with discrepancies in sex and/or year of birth between HealthVerity claims and California/Louisiana immunization registry datasets,* or missing sex in both datasets. In cases where sex or year of birth is unknown in either the claims or registry by available in the other, individuals will be included and the information from the dataset where year of birth or sex is available will be used, regardless of whether the information stemmed from claims or registry data.

3. *Persons with a diagnosis of COVID-19 in any setting ≤ 90 days prior to index (FDA EUA date per vaccine)*

4. *Receipt COVID-19 vaccine ≤ 90 days prior to index (FDA EUA date per vaccine).*

5. Persons who died before index date, operationalized as the last day in the month and year of death.

5.2.3. Subgroups and Stratification

Results will be stratified by age group (overall, age 6 months – 4 years, 5-11, 12-17).

Additionally, the analysis of the primary outcome will be stratified by state of residence. As tests and measures of heterogeneity can be affected by large sample sizes, clinical relevance will be used to assess the difference in VE estimates between the two registries. If a difference is deemed clinically relevant, secondary and exploratory outcomes may be additionally stratified by state of residence.

The following subgroups will be used to evaluate VE for the XBB.1.5 vaccine:

- Receipt of BA.4/5 bivalent vaccine
- Receipt of two prior doses of mRNA COVID-19 vaccine and no prior receipt of BA.4/5 bivalent vaccine
- No prior receipt of a COVID-19 vaccine

Of note, persons who received 1 mRNA wildtype dose, or received 2 wildtype doses <21 days apart, do not meet any of these criteria and will be included in the main analysis (which does not consider prior vaccination history) but excluded in these subgroup analyses of vaccination history.

5.3. Variables

5.3.1. Exposures

The exposure of interest is the BNT162b2 XBB.1.5 mRNA vaccine.

Information on vaccination status will continue to be collected on or after the authorization/ approval date, therefore vaccination status will be considered as a time-varying exposure. Patients will be assigned as vaccinated or unvaccinated according to their vaccination status by each person-time record during the follow-up. A patient will be considered vaccinated from 14 days after their BNT162b2 vaccination date till the remainder of their follow-up.

The source of the exposure variables are the California and Louisiana vaccine registries and the HV (HealthVerity) Claims database. Clinical Vaccines Administered code (CVX), National Drug Codes (NDC) and Current Procedural Terminology® (CPT) codes used to identify the XBB.1.5 and BA.4/5 vaccinations are listed in the Protocol Annex.

XBB vaccinations will be enumerated using the following hierarchical definition:

1. *Any person in HealthVerity claims data who has a record in the California/Louisiana Immunization Registry with a CVX code indicating having received the BNT162b2 XBB.1.5 monovalent mRNA vaccine will be considered vaccinated 14 days after registry event date.*
2. Any person in HealthVerity claims data who has a record in the California/Louisiana Immunization Registry that does not include a CVX code but lists the vaccine manufacturer as Pfizer with a vaccination date on or after 09/11/2023 indicating having received the BNT162b2 XBB.1.5 monovalent mRNA vaccine will be considered vaccinated 14 days after registry event date.

3. Any person in HealthVerity claims data who has a record in the California/Louisiana Immunization with an event indicating having received the BNT162b2 XBB.1.5 monovalent mRNA vaccine will be considered vaccinated 14 days after registry event date.

4. *Any person in HealthVerity claims data who does not have a record in the California/Louisiana Immunization Registry but has a pharmacy claim with an NDC code indicating a BNT162b2 XBB.1.5 monovalent mRNA vaccine dose will be considered vaccinated 14 days after claim date.*

5. *Any person in HealthVerity claims data who does not have a record in the California/Louisiana Immunization Registry and does not have a pharmacy claim with an NDC code indicating a BNT162b2 XBB.1.5 monovalent mRNA vaccine dose but has a medical claim with a CPT code indicating a BNT162b2 XBB.1.5 monovalent mRNA vaccine dose will be considered vaccinated 14 days after claim date.*

6. *Any person in HealthVerity claims data who does not have a record in the California/Louisiana Immunization Registry, does not have a pharmacy claim with an NDC code indicating a BNT162b2 XBB.1.5 monovalent mRNA vaccine dose and does not have a medical claim with a CPT code indicating a BNT162b2 XBB.1.5 monovalent mRNA vaccine dose will be considered BNT162b2 XBB.1.5 unvaccinated.*

BA.4/5 vaccinations will be enumerated in the same manner as XBB.1.5.

5.3.2. Outcomes

The outcomes of interest are presented in Table 5.2.

Table 5.2 Outcomes Variables

Outcome Type	Variable	Data source(s)	Operational definition
Primary outcome	COVID-19 diagnosis	HV Claims	Any encounter regardless of diagnosis setting with ICD-10-CM code: U07.1.
Secondary outcomes	COVID-19 outpatient encounter	HV Claims	Outpatient encounter with ICD-10-CM code: U07.1
	COVID-19 hospitalization	HV Claims	Hospitalizations "for COVID-19" were defined as those with an inpatient encounter at an acute care facility with ICD-10-CM code: U07.1 "COVID-19" that did not have an additional incidental code (for unintentional injury, physical trauma, poisoning, short-stay (<2 days) childbirth or severe and persistent mental illness) or a U07.1 hospitalization where treatment used solely for COVID-19 (eg, remdesivir) was identified regardless of accompanying diagnoses. As a sensitivity analysis, we will relax this definition to consider any hospitalization with a U07.1 code, without consideration of incidental findings.
	COVID-19 emergency department visit	HV Claims	Emergency department visit with ICD-10-CM code: U07.1
	COVID-19 critical illness	HV Claims	Intensive care unit [ICU] admission, mechanical ventilation, or inpatient death. If sample size allows (Section 5.5), we will evaluate VE against each component separately

Outcome Type	Variable	Data source(s)	Operational definition
	Non-COVID-19 respiratory infections (pneumonia, respiratory syncytial virus (RSV), rhinovirus, antibiotic prescriptions)	HV Claims	Any encounter regardless of setting with ICD-10-CM code for diagnosis or NDC codes for antibiotics (codes in the Protocol Annex). If sample size allows (Section 5.5), we will evaluate VE against each component separately
	Negative control outcome, such as accidental injury, ingrown toenail or atopic dermatitis	HV Claims	Any encounter regardless of setting, with ICD-10-CM code for negative control outcomes (codes in the Protocol Annex)
Exploratory outcomes	COVID-19 urgent care encounter	HV Claims	An urgent care encounter with ICD-10-CM code: U07.1 "COVID-19"
	All-cause mortality after COVID-19 encounter	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 encounter, regardless of setting
	All-cause mortality after COVID-19 outpatient encounter	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 outpatient encounter
	All-cause mortality after COVID-19 inpatient encounter	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 inpatient encounter

Outcome Type	Variable	Data source(s)	Operational definition
	All-cause mortality after COVID-19 emergency department encounter	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 emergency department encounter
	All-cause mortality after COVID-19 critical illness	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 critical illness
	All-cause mortality after COVID-19 urgent care encounter	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 urgent care encounter

Each outcome will be assessed independently and experiencing one outcome will not be a censoring event for another outcome. Where a patient has multiple instances of the same outcome, the first one will be used.

5.3.3. Covariates

The list of demographic and clinical characteristics that will be used to describe the cohort and considered for inclusion as covariates in models are provided in Table 5.3. Age, sex, state of residence and prior COVID-19 vaccinations will be obtained from the HealthVerity claims database and state vaccine registries. The HealthVerity claims database will be the sole source of all other variables.

All covariates will be assessed in the follow-up times outlined below. Categories may be collapsed or covariates may be excluded or combined if models fail to converge or if patient counts are low (Section 5.7.2). Additionally, categories with large patient counts may be split further.

Table 5.3 Demographic and Clinical Characteristics

Variable	Role(s)	Operational definition
Age	-Inclusion/ exclusion criteria -Descriptive (continuous and categorical) -Covariate (categorical) -Subgroup (categorical)	Age assessed at index date, as 2023 – year of birth. Categorized as: <5, 5-11 or 12-17 years.
Sex	-Inclusion/ exclusion criteria -Descriptive -Covariate	Sex assessed at index date. Male or female.
State of residence	-Inclusion/ exclusion criteria -Descriptive -Covariate -Subgroup	Defined in Section 5.2 California or Louisiana
Payor	-Descriptive -Covariate	Payor assessed at index date. Commercial, Medicaid, Medicare Advantage, Unknown/Missing

Variable	Role(s)	Operational definition
CDC-defined high risk for severe COVID-19(1)	-Descriptive (any high-risk condition and for each individual condition) -Covariate (each individual condition)	Assessed using all available lookback data. • Asthma • Cancer, as hematologic malignancies • Cerebrovascular disease • Chronic kidney disease: people receiving dialysis • Chronic lung diseases: bronchiectasis, COPD, interstitial lung disease, pulmonary embolism, pulmonary hypertension • Chronic liver diseases: cirrhosis, non-alcoholic fatty liver disease, alcoholic liver disease, autoimmune hepatitis • Cystic fibrosis • Diabetes mellitus, type 1 • Diabetes mellitus, type 2 • Disabilities, including Down syndrome • Heart conditions: heart failure, coronary artery disease, or cardiomyopathies • HIV • Mental health conditions: mood disorders including depression; Schizophrenia spectrum disorders • Neurologic conditions limited to dementia • Obesity (BMI $\geq$30 kg/m^2) • Primary immunodeficiencies • Smoking, current and former • Solid organ or blood stem transplantation • Tuberculosis • Use of corticosteroids or other immunosuppressive medications The lists of codes are provided in the Protocol Annex. Note: due to the limitations of claims data, "physical inactivity" will not be measurable and has been removed from the list.

Variable	Role(s)	Operational definition
Charlson Comorbidity Index using weights from Quan et al (2011)	-Descriptive (continuous and categorical) -Covariate (categorical)	Assessed in 6 months prior to index date. Conditions are scored as: 1 point for each of: • Chronic pulmonary disease • Rheumatologic disease • Diabetes mellitus with chronic complications • Renal disease 2 points for each of: • Congestive heart failure • Dementia • Mild liver disease • Any malignancy, including leukemia and lymphoma 4 points for each of: • Moderate or severe liver disease • HIV/AIDS 6 points for each of: • Metastatic solid tumor Categorized as: 0, 1 or 2+ The list of codes is provided in the Protocol Annex
Receipt of skilled nursing care and/or long-term care facility stay in year prior to index	-Descriptive -Covariate	Assessed in 6 months prior to index date Yes or No The list of codes is provided in the Protocol Annex
Wellness visit	-Descriptive -Covariate	Assessed in 6 months prior to index date Yes or No The list of codes is provided in the Protocol Annex
Decreased functional status	-Descriptive -Covariate	Assessed in 6 months prior to index date Yes or No The list of codes is provided in the Protocol Annex

Variable	Role(s)	Operational definition
Influenza vaccination in 6 months prior to index	-Descriptive -Covariate	Assessed in 6 months prior to index Yes or No The list of codes is provided in the Protocol Annex
Pneumococcal vaccination using all-available time prior to index	-Descriptive -Covariate	Assessed using all-available lookback data Yes or No The list of codes is provided in the Protocol Annex
Herpes zoster vaccination using all-available time prior to index	-Descriptive -Covariate	Assessed using all-available lookback data Yes or No The list of codes is provided in the Protocol Annex
Outpatient visits prior to index	-Descriptive (continuous and categorical) -Covariate (categorical)	Number of outpatient visits in 180 days prior to index Categorized as: 0, 1 or 2+ The list of codes is provided in the Protocol Annex
Inpatient admissions prior to index	-Descriptive (continuous and categorical) -Covariate (categorical)	Number of inpatient admissions in 180 days prior to index Categorized as: 0, 1 or 2+ The list of codes is provided in the Protocol Annex
Emergency department visits prior to index	-Descriptive (continuous and categorical) -Covariate (categorical)	Number of emergency department visits in 180 days prior to index Categorized as: 0, 1 or 2+ The list of codes is provided in the Protocol Annex
Telehealth visits prior to index	-Descriptive (continuous and categorical) -Covariate (categorical)	Number of telehealth encounters (CPT: 99201-99215) in 180 days prior to index Categorized as: 0, 1 or 2+
Number of documented SARS-CoV-2 tests in the 180 days prior to index	-Descriptive (continuous and categorical) -Covariate (categorical)	Number of visits in any setting with CPT in 180 days prior to index: 86408, 86409, 0225U, 0226U Categorized as: 0, 1 or 2+
Prior post-COVID-19 conditions	-Descriptive -Covariate	Assessed using all-available lookback data, any encounter regardless of setting with ICD-10-CM U09.9 "Post-COVID-19 condition" Yes or No

Variable	Role(s)	Operational definition
Prior COVID-19 diagnosis	-Descriptive -Covariate	Any encounter with ICD-10-CM code: U07.1 between 90 and 365 days prior to index Yes or No
Prior bivalent BA.4/5 COVID-19 vaccine	-Descriptive -Subgroup	Where the BNT162b2 XBB.1.5 vaccine is the exposure of interest. Using all-available lookback data, received bivalent BA.4/5 COVID-19 vaccine (regardless of brand) prior to index Yes or No The list of codes is provided in the Protocol Annex
At least two prior doses of mRNA vaccine and no bivalent BA.4/5 COVID-19 vaccine	-Descriptive -Subgroup	Where the BNT162b2 XBB.1.5 vaccine is the exposure of interest. Using all-available lookback data, received at least two prior doses of mRNA vaccine (regardless of formulation) and did not receive bivalent BA.4/5 COVID-19 vaccine. Yes or No Note: the doses must be at least 21 days apart. The list of codes is provided in the Protocol Annex
Never COVID-19 vaccinated	-Descriptive -Subgroup	Using all-available lookback data, no prior receipt of a COVID-19 vaccine (regardless of formulation) Note: Patients without prior bivalent BA.4/5 COVID-19 vaccine and either a single mRNA vaccine dose or two mRNA doses less than 21 days apart are categorized as "No", as per Section 5.2.3. Yes or No The list of codes is provided in the Protocol Annex
Time since last COVID-19 vaccine	-Descriptive -Subgroup	Using all-available lookback data, time in months between index date and prior receipt of a COVID-19 vaccine (regardless of formulation) The list of codes is provided in the Protocol Annex

5.4. Data Sources

The sources of the data are the State of California and the State of Louisiana vaccine registries linked to closed claims in HealthVerity.

5.4.1. California and Louisiana Immunization Registries

The California Immunization Registry is run by the California Department of Public Health and collects nearly all records of COVID-19 vaccinations administered in the state of California.

Data include a unique HealthVerity person identification code, vaccination event date and Clinical Vaccines Administered (CVX) code. CVX codes are unique to brand and formulation of vaccine.

Data is linked using HealthVerity's tokenization software, whereby specific patient identifiers are passed through the software and a de-identified patient ID is assigned. The same tokenization process is applied to closed claims sourced by HealthVerity, and patients can thus be linked across data sources. The tokenization process occurs outside of Pfizer, no personal identifying information will be transferred, and Pfizer will only have access to de-identified data. The data comply with Health Insurance Portability and Accountability Act (HIPAA) regulations.

The Louisiana Vaccine Registry is structured and accessed in the same way as California.

5.4.2. HealthVerity's Administrative Claims Data

HealthVerity closed claims will be limited to approximately 19 million patients residing in California and approximately 3.8 million in Louisiana.

HealthVerity's description of their data is that they are generally representative of the age and sex distribution of the population. For our specific research context, samples created from HealthVerity claims also generally follow California's demographic patterns with respect to age & sex.

The California subset of HealthVerity does not include patients in the Kaiser Permanente network. Insurers contributing to closed claims in both states include a mix of commercial payers, Medicare Advantage/Part C pains, and Medicaid Managed Care plans. Data elements include patient demographic information, inpatient/outpatient visit-level information such as diagnoses, procedures, and length of stay, hospital characteristics, and medication information. Owing to the nature of claims, the data represent the final set of diagnoses over the course of the hospitalization sent to the patient's insurer for reimbursement, with diagnosis prioritization assigned by clinicians or hospital staff. For inpatient encounters, the data will be assumed to represent the overarching events during hospitalization and may be less subject to rule out diagnosis codes than electronic health record data.

Death information is available and is sourced from over 40,000 public and private sources nationally. The coverage rate, estimated at over 90% of CDC reported deaths through present day, is based on the number of deaths in the Fact of Death Mortality Index compared to the number of deaths reported by the CDC. Due to HIPAA privacy restrictions, date of death is reported as month and year. Cause of death is not available.

5.5. Sample Size and Power Calculations

Since the current study utilizes retrospective deidentified data without ability to recruit to a target, sample size calculations are not applicable. Further, this is a descriptive study without a priori specified hypotheses. However, we examined the precision of the expected XBB.1.5 absolute VE estimates for the outcomes under various scenarios.

(25% of 12 million are pediatric)(16% uptake)*(60% of doses were Pfizer-BioNTech)*

~ 288,000 pediatric XBB doses in California.

The sample size used at start of follow-up was 3 million. Assuming similar absolute VEs to those seen with the BA.4/5 vaccine for symptomatic COVID-19 infections, we could anticipate estimates between 20 and 40%[4]. To assess the precision for a range of outcomes, incidence rates of 50, 100, 1000 and 10,000 cases per 100,000 unvaccinated persons during the study period were considered. A censoring rate of 5% by the end of the study period was used. A uniform vaccination rate between 14 days after study index and the end of follow-up was assumed.

Survival and censoring times were simulated assuming exponential distributions. Vaccination status was modelled as a time-varying covariate[5]. Hazard ratios were input as 1-(VE/100). Simulations were run a minimum of 10 times per scenario. Cox regression models were run on each simulated dataset to estimate the 95% CIs for the log HRs. Means were generated and exponentiated to obtain the 95% CIs for the VEs using (1- HR) x 100 and used to describe VE precision. Simulations were repeated on 25% of the initial sample size to represent the precision in a potential subgroup.

Table 5.4 Simulated 95% Confidence Intervals [Total Number of Events] for BNT162b2 XBB.1.5 Vaccine Effectiveness Estimates

Number at Start of Follow-up	Vaccine Effectiveness, %	Cases per 100,000 Unvaccinated Persons by End of Study			
		50	100	1000	10,000
3,000,000	20	2, 37 [1488]	7, 32 [2871]	16, 24 [28,678]	19, 21 [274,706]
	30	11, 45 [1429]	17, 41 [2849]	26, 34 [28,475]	29, 31 [272,600]
	40	22, 53 [1418]	29, 51 [2845]	37, 43 [28,256]	39, 41 [270,864]
750,000	20	-24, 49 [361]	-7, 43 [720]	12, 28 [7150]	17, 23 [68,646]
	30	-12, 57 [358]	2, 50 [713]	22, 37 [7097]	28, 32 [68,173]
	40	6, 67 [354]	14, 58 [712]	33, 47 [7077]	38, 42 [67,719]

Table 5.4 shows that there should be sufficient sample size in the California registry to produce 95% CIs for VEs that exclude 0 for the evaluated VEs and incidence rates, except for the rarer event scenarios in a subgroup. Assuming similar VEs in the two state registries, the inclusion of patients from the Louisiana registry would increase the precision further.

5.6. Missing Data

No imputation for missing values will be performed. Subject to patient counts (Section 5.7.2), missing or 'unknown' values will be reported and analyzed as separate categories and included in the totals.

5.7. Statistical Methodology and Analyses

5.7.1. General Considerations

Descriptive statistics will be used to summarize variables. For categorical variables, the frequency and percentage of patients in each category will be presented. Percentages will be based on the total number of relevant patients. For continuous variables, data will be presented as means, standard deviations, medians, 25th and 75th percentiles, minimums and maximums. Continuous variables may be additionally categorized and analyzed as categorical variables.

No statistical hypotheses are specified and there will be no correction for multiple comparisons.

Time in days between two dates will be calculated as (date2 – date1) + 1, where date2 is on or after date1. Where date2 is before date1, time in days between two dates will be calculated as (date2 – date1). Index date will be considered as Day 1. There is no Day 0.

Unless stated otherwise, one year is defined as 365 days, and one month is defined as 30 days.

All data analysis will be conducted using statistical software SAS version 9.4 or R version 4.1.0 or later.

5.7.2. Minimum Sample Size Requirements

In order to maintain patient de-identification, outcomes, covariates and subgroups will have at least 10 patients in each category to be included in the analysis.

In addition, a minimum of five events per variable (the number of degrees of freedom) will be required in each level of the outcome variable in the adjusted Cox regression models to minimize the potential for biased estimates [4]. Note, as an example, a categorical predictor variable with three levels contributes two degrees of freedom.

5.7.3. Index Date, Baseline and Follow-up

The index date for each individual will be defined as the BNT162b2 XBB.1.5 EUA or FDA approval date (Table 5.1). Age and sex will be assessed at index date. The baseline period used to assess other patient characteristics are specified for each covariate in Table 5.3.

Unvaccinated follow-up time will occur from the date of FDA authorization/approval until 13 days after the receipt of vaccination, the outcome of interest, or are censored at the first of the following: 6 months of follow-up, end of medical and/or pharmacy enrollment, receipt of a non-Pfizer BioNTech vaccine. Where mortality is not the outcome of interest, follow-up time will be additionally censored at death. Where mortality is the outcome of interest, follow-up time will be additionally censored at a death not meeting the outcome criteria.

Vaccinated follow-up time will occur from 14 days after the receipt of vaccination until the outcome of interest, or are censored at the first of the following: 6 months of total *follow-up* (ie, 11 March 2024)*, end of medical and/or pharmacy enrollment, receipt of a non-Pfizer BioNTech vaccine or receipt of a second BNT162b2 COVID-19 vaccine dose.* Where mortality is not the outcome of interest, follow-up time will be additionally censored at death. Where mortality is the outcome of interest, follow-up time will be additionally censored at a death not meeting the outcome criteria.

5.7.4. Eligibility, Demographic, Clinical and Exposure Characteristics

The number and percentage of patients meeting the eligibility criteria will be reported in a subject evaluation table.

Descriptive statistics will be used to summarize patient demographics, clinical characteristics, and healthcare utilization on or prior to index date. Descriptive statistics will be repeated by BNT162b2 vaccination status by the end of follow-up. Standardized mean differences (SMD) will be used to assess covariate balance by BNT162b2 vaccination status by the end of each month of follow-up. Overall SMDs will be presented for each covariate as the maximum absolute value across the months. SMDs ≤ 0.1 will be considered as evidence of negligible imbalance.

Summaries of the overall follow-up time in months will be presented and calculated as time from index date to the earliest of 6 months of follow-up, end of medical and/or pharmacy enrollment, receipt of a non-Pfizer BioNTech COVID vaccine, receipt of a second BNT162b2 COVID-19 vaccine dose or death.

The number of patients who are BNT162b2 vaccinated by the end of the overall follow-up will be reported. Kaplan-Meier analysis will be used to summarize the percentage of persons who are BNT162b2 vaccinated by the end of each month of follow-up. Patients will be censored at the earliest of the following: 6 months of follow-up, end of medical and/or pharmacy enrollment, receipt of a non-Pfizer BioNTech COVID vaccine or death.

The above analysis will be repeated by age group and state of residence.

5.7.5. Primary Analyses

Unvaccinated and unvaccinated person-time (Section 5.7.3) for each outcome of interest will be calculated for each patient, summed across all patients and presented in person-months. The total number of patients with each of the outcomes will be presented by BNT162b2 vaccination status at the time of the COVID-19 encounter. Crude incidence rates and 95% confidence intervals (CIs) will be presented as events per 100,000 person-months (or as appropriate) by BNT162b2 vaccination status. The above analysis will be presented overall and may also be presented by month of follow-up for the primary and secondary outcomes.

Unadjusted Cox regression models will be used to generate hazard ratios (HR) and 95% CIs for the effect of the BNT162b2 for each outcome, with BNT162b2 modelled as a time-varying exposure. Crude VE percentages will then be estimated based on (1-HR) x 100, where HR is obtained from the unadjusted model.

For the primary outcome, for each of incidence rates and hazard ratios, an additional stratification of interest will employ an interaction term between exposure and a time-varying calendar time covariate, to allow for separate estimates of incidence and hazard in the predominantly XBB (11 September – 08 December 2023), co-circulation

(09 December 2023 – 05 January 2024) and predominantly JN.1 (January 6, 2024 – end of data collection [March 11, 2024]) eras.

Adjusted Cox regression models will also be used to generate adjusted HRs (aHR) and adjusted VE percentages, with covariates listed in Table 5.3 included as time-fixed covariates. For analyses where the number of events per variable is low (Section 5.7.2), covariates with SMDs ≤0.1 may be excluded from regression models.

Measures of absolute and relative VE: For each outcome above, where sample size allows, *results will be calculated as:*

1. *Absolute VE comparing vaccinated to unvaccinated persons, regardless of prior vaccination*

2. *Relative VE will be calculated for each of the following mutually exclusive subgroups.*

    a. *XBB.1.5 vaccinated vs XBB.1.5 unvaccinated persons who received BA.4/5 bivalent vaccine* (regardless of brand).

    b. *XBB.1.5 vaccinated vs XBB.1.5 unvaccinated persons who did not receive BA.4/5 bivalent vaccine but did receive at least 2 doses of mRNA vaccine (regardless of* brand*).*

    c. *XBB.1.5 vaccinated vs never COVID-19 vaccinated* (of any formulation or brand).

The proportional hazards assumption for the exposure will be assessed graphically with the inspection of Schoenfeld residuals. If the assumption is not met, the follow-up may be restricted up to the time where the assumption is met or VEs may be generated during intervals using piecewise Cox models.

We will further use incidence of the COVID-19 endpoint in unvaccinated persons and apply VE estimates to estimate the number of cases averted per 100,000 person-months. This estimate of the vaccine-preventable disease incidence will be generated for each endpoint as the adjusted VE as a proportion multiplied by the incidence rate among unvaccinated persons.

The above analysis will be repeated by age group for each outcome, subject to patient counts (Section 5.7.2). Additionally, the analysis for the primary outcome will be repeated by state of residence. Subject to the distribution of time since last prior COVID-19 vaccine, relative VEs may also be repeated by subgroups of that variable.

5.7.6. Assessment of Residual Confounding

Negative control outcomes will be used to quantify residual and unmeasured confounding (Table 5.2). The analysis described above will be conducted for the negative control outcomes. Persons with a diagnosis of the negative control outcome within 90 days prior to index date will be excluded from the analysis for that outcome.

In the event aHR for negative control outcomes have a significant effect (95% CI do not include null value of 1), then the set of confounders will be adjusted to allow for calibration. The COVID-19 related outcome VE results may be additionally reported after correcting for estimates of the bias [5].

5.7.7. Sensitivity Analyses

Sensitivity analyses will consider the impact of study design and analytic assumptions, such as varying the exposure definition to 7 days post vaccination rather than 14 days and restricting the follow-up to 3 rather than 6 months.

6. REFERENCES

7. APPENDICES

7.1. Appendix A: Table Shells

The list of planned outputs and table shells will be presented in a separate document.

7.2. Appendix B: Code Lists

The list of ICD-10-CM, NDC and CPT codes to define each variable above is kept in a separate document, with regular updates to reflect release of new codes.

Document Approval Record

Document Name:	C4591065 Non Interventional Study Statistical Analysis Plan_FINAL_10June2024
Document Title:	C4591065 RAVEN 2024 SAP Aim 2_10June2024 FINAL

Signed By:	Date(GMT)	Signing Capacity
PPD	11-Jun-2024 17:59:12	Final Approval

Non-Interventional Study

C4591065

Effectiveness of mRNA COVID-19 Vaccine Formulations using State Vaccine Registry and Insurance Claims Data

Long Term Cost Analysis

Statistical Analysis Plan

(SAP)

Version: 1.0

Author: PPD

Date: 30-May-2024

TABLE OF CONTENTS

1. AMENDMENTS FROM PREVIOUS VERSION(S)

NA. First version.

2. LIST OF ABBREVIATIONS

Abbreviation	Definition
AIC	Akaike information criterion
CI	Confidence Interval
CCI	Charlson Comorbidity Index
CDC	Centers for Disease Control and Prevention
COPD	Chronic obstructive pulmonary disease
COVID-19	Coronavirus disease 2019
CPT	Current Procedural Terminology
CVX	Clinical Vaccines Administered code
ED	Emergency department
HCPCS	Healthcare Common Procedure Coding System
HCRU	Healthcare resource utilization
HIPAA	Health Insurance Portability and Accountability Act
ICD-10-CM	International Classification of Diseases, Tenth Revision, Clinical Modification
ICU	Intensive care unit
IMV	Invasive mechanical ventilation
IPCW	Inverse Probability of Censoring Weighting
LOS	Length of stay
NDC	National Drug Codes
Q1	First quartile, 25th percentile

Abbreviation	Definition
Q3	Third quartile, 75th percentile
SAP	Statistical analysis plan
SD	Standard deviation
SMD	Standardized mean differences
UC	Urgent care
US	United States
USD	US dollars

Note: in this document, any text taken directly from the non-interventional study protocol is ***italicized***.

3. RATIONALE AND BACKGROUND

Until recently, vaccine registries were not available for de-identification and linkage to secondary claims-based databases. However, through a collaboration with HealthVerity, a leading supplier of secondary health data, the State of California and the State of Louisiana have made their vaccine registries available for linkage for research purposes. California and Louisiana deidentify the registries, using HealthVerity's software to convert each patient's protected health information to a unique HealthVerity ID which can then be linked to the full HealthVerity database to gather clinical history and outcomes.

The purpose of this document is to detail the statistical methods and analyses that will be used in the HealthVerity claims database linked to the State of California/Louisiana's vaccine registries to address Aim 3B of the study protocol. Separate SAPs will be produced to address the other aims of the protocol.

4. STUDY OBJECTIVES AND HYPOTHESES

The objective of Aim 3B of this study is to estimate long-term (3B) healthcare resource utilization and costs associated with COVID-19 by vaccination status, age group and site of care.

This study is descriptive in nature and as such does not include any pre-specified hypotheses.

5. RESEARCH METHODS

5.1. Study Design

This will be a retrospective cohort study using data from patients who were California or Louisiana residents in the HealthVerity claims database.

5.2. Study Population

The study population will be comprised of adult (≥18 years of age) or pediatric (6 months to <18 years of age) patients who were diagnosed with COVID-19 between 01 June 2022 and 30 November 2022 and matched cohorts of control patients, who will be identified as those without a clinical diagnosis of COVID-19 during the matching time period. Patient matching will be described in greater detail in the latter section of the study protocol.

5.2.1. Inclusion Criteria

Patients must meet the following criteria to be eligible for the study:

All Patients:
Aim 3B criteria are different based on the pediatric and adult population analyses as follows:

5.2.1.1. Aim 3B Adult Inclusion Criteria

1. Born in 2004 or earlier (as age 18 or older as of the index date, owing to the data providing year but not month or day of birth).
2. Has COVID-19 diagnosis (ICD-10-CM code U07.1) in any setting, any position between 01 June 2022 and 30 November 2022 (first eligible code is the index date).
3. At least one year of continuous pharmacy and medical enrollment in HealthVerity prior to index date (ie, COVID-19 diagnosis). A gap of up to 30 days will be allowed.
4. Have been a State of California/Louisiana resident for at least one year prior to index date (ie, COVID-19 diagnosis)
5. Have continuous medical and prescription insurance coverage for at least 30 days after the index date (follow-up period).
    - The duration of the follow-up period will be required to be at least 1 month.
    - NOTE: since the evaluation of the long-term HCRU and cost starts from 1 month after the index date, this requirement for patients to have at least one month of follow-up is needed for this study.

5.2.1.2. Aim 3B Pediatric Inclusion Criteria

1. Born in 2005 or later (as age 0-17 years in 2023, owing to the data providing year but not month or day of birth)
2. Has COVID-19 diagnosis (ICD-10-CM code U07.1) in any setting, any position between 01 June 2022 and 30 November 2022 (first eligible code is the index date).
3. At least six months of pharmacy and medical enrollment in HealthVerity prior to index date. Six months was selected to allow for analysis to include children <1 year due to patient privacy measures in HealthVerity data. A gap of up to 30 days will be allowed
    - So, for the pediatric patients, the minimum age will be 6 months due to this requirement of six months of baseline period.
4. Have been a State of California/Louisiana resident for at least six months prior to index date
5. Have continuous medical and prescription insurance coverage for at least 30 days after the index date (follow-up period).
    - The duration of the follow-up period will be required to be at least 1 month.

NOTE: since the evaluation of the long-term HCRU and cost starts from 1 month after the index date, this requirement for patients to have at least one month of follow-up is needed for this study.

The index COVID-19 hospitalizations will be identified between 01 June 2022 and 30 November 2022 (COVID identification period)

- To avoid the incidental identification of COVID-19, hospitalizations for COVID-19 were defined as those with an inpatient encounter at an acute care facility with ICD-10-CM code U07.1. Furthermore, hospitalizations did not have an additional incidental code (for unintentional injury, physical trauma, poisoning, short-stay [<2 days] childbirth or serious psychiatric admissions), or a U07.1 hospitalization where treatment used solely for COVID-19 (eg, remdesivir) was identified regardless of accompanying diagnoses.

- The earliest claim will be identified as the index event with the corresponding date (hospital admission date) as the index date.

Patients who were California or Louisiana residents in the HealthVerity claims database will be identified using the following hierarchical definition:

1. Have a location of California or Louisiana indicated within their HealthVerity claims enrollment file record.

    a. Or, have a location other than California or Louisiana in their HealthVerity claims enrollment file record but have one or more records in the California or Louisiana Immunization Registry. Examples include people who previously lived in Texas but moved to California/Louisiana and received a vaccine, or residents of other states who crossed state lines to receive a vaccine in California or Louisiana.

Matched Control Patients:

1. Control patients will be exact-matched up to 4:1 to COVID-19 patients using the following matching criteria:
    a. Control patients will be assigned a random date between 01 June 2022 and 30 November 2022 as the index date. The control patients' index date will be further required to be within the ± 14 days of the index date of their corresponding matched COVID-19 patients.
        i. Have continuous medical and prescription insurance coverage between June 1, 2022 and November 30, 2022
        ii. To ensure that the follow-up period of control patients will not be post-COVID-19, control patients are required to have no COVID-19 diagnoses in any position in any inpatient or outpatient claims prior to 30 November 2022.

- Control patients are allowed to have a COVID-19 diagnosis after 30 November 2022 as long as all other patient matching criteria are met. In such cases, the follow-up period of these control patients will be censored at the date of the first inpatient or outpatient claim with a COVID-19 diagnosis in any position.

iii. Control patients will have similar continuous insurance enrollment criteria as the COVID-19 patients, described as below:

- Similar to their matched adult and pediatric COVID-19 patient counterparts, control patients will be required to have 12 and 6 months of continuous baseline enrollment, respectively.
- Control patients will be required to have continuous medical and prescription insurance coverage for at least 1 month after the index date (follow-up period).

b. Control patients will be matched 4:1 without replacement to COVID-19 patients using the following criteria (exact match unless otherwise specified).

i. Index date within ± 14 days of the index date of the COVID-19 patient

ii. Age

- Sample size permitting, an exact match of age will be used. If not, matched patients will be required to be within the same age group (see Table 1) AND have an age gap of no more than 2 years (control patient age = COVID-19 patient age ± 2 years).

iii. Gender

iv. State

v. Insurance type at the Index Date

vi. Charlson Comorbidity Index (CCI) groups during the baseline period (CCI score of 0, 1, 2, 3+)

vii. Groups of the number of all-cause hospitalizations during the baseline period (0, 1, 2, 3+ hospitalizations)

viii. Groups of the number of all-cause outpatient visits during the baseline period (0, 1, 2, 3+ outpatient visits)

c. The patient matching criteria may be further modified/relaxed if the proposed criteria are too restrictive (ie, collapsing and/or combining criteria).

5.2.2. Exclusion Criteria

Patients meeting any of the following criteria will be excluded from the study analyses:

1. Currently pregnant (identified with medical codes during the up to 9 months lookback prior to the index date) or immunocompromised (see algorithm below) at the time of study index

    - Immunocompromised at the time of study index defined as one inpatient and/or 2 outpatient claims 7 days apart for any of:

        - ICD-10-CM codes for HIV/AIDS using all-available lookback
        - ICD-10-CM codes for hematologic malignancy in prior year
        - ICD-10-CM codes for solid organ or bone marrow transplant in prior two years
        - NDC, HCPCS and/or ICD-10-PCS codes for immunosuppressive medication in 210 prior days. *Note: 210 days was chosen to add a 30-day grace period to 180 days, given some of the most immunosuppressive medications like rituximab and ocrelizumab are given semi-annually and we wanted to minimize misclassifications due to an overly strict definition.*

2. Individuals with discrepancies in sex and/or year of birth between HealthVerity claims and California/Louisiana immunization registry datasets.
3. Individuals with a diagnosis of COVID-19 in any setting ≤90 days prior to index.
4. Individuals who were vaccinated within 13 days prior to the index date.

5.2.3. Follow-up

Patients will be followed for between 1 and 12 months after the index date. Patients will be censored at the earliest of: 1) disenrollment from insurance coverage; 2) the date of the first claim with a pregnancy diagnosis in the follow-up; 3) the end of data availability on 30 November 2023 or 4) receipt of COVID vaccination. Control patients may be further censored at the date of the first claim with a COVID-19 diagnosis after 30 November 2022.

A sensitivity analysis will be carried out where COVID-19 patients may be further censored at the date of COVID-19 reinfection. This will be further described in the latter part of the study protocol.

5.3. Variables

All data will be extracted from the HealthVerity database; the study measurements are described in detail in Table 1-Table 2. Patient demographic and clinical characteristics will be evaluated at the index date. Comorbid conditions and Charlson Comorbidity Index (CCI) scores will be assessed during the baseline period. Comorbid conditions will be identified by diagnosis codes and/or medication usage. In the default study analysis both the charge and allowed amounts will be evaluated for cost data from the claim database. A sensitivity analysis will be carried out where the paid amounts from the open-source database will be used as the cost data (details described later). HCRU and cost measurements will be evaluated during the baseline and follow-up periods. Intensive care unit (ICU) admission and invasive mechanical ventilation (IMV) will be identified by hospital revenue codes, Healthcare Common Procedure Coding System (HCPCS) procedure codes, and Current Procedural Terminology, 4th Edition (CPT-4) procedure codes.

Table 1. Demographic, Clinical, and Baseline Characteristic Variables

Variable	Role	Data source	Operational definition
Age	Patient characteristic	HealthVerity	Age will be defined at the index date and will be used to assign patients to age groups: Adult (≥18 years): 18 – 29, 30 – 49, 50 – 64, 65 – 74, 75+ years; Pediatric (6 months – < 18 years): 0 – < 2, 2 – 4, 5 – 11, 12 – <18 years. Reported as minimum (min), maximum (max), median (1st – 3rd quartile [Q1-Q3]) and mean (standard deviation [SD]), as well as counts and percentages for age groups.
Gender	Patient characteristic	HealthVerity	Distribution of female and male, unknown patients. Evaluated at the index date and reported as counts and percentages.
State	Patient characteristic	HealthVerity	Distribution of patients' state information. Evaluated at the index date and reported as counts and percentages.
Insurance Type	Patient characteristic	HealthVerity	Distribution of the insurance payer type. Evaluated at the index date and reported as counts and percentages.
Index Year-Month	Patient characteristic	HealthVerity	Year-Month combination of the index date. Evaluated at the index date and reported as counts and percentages.
Site of Index COVID-19 Diagnosis	Clinical characteristic	HealthVerity	Inpatient, ED/urgent care (UC), Outpatient Non-ED/UC, Other/unknown. Evaluated at the index date and reported as counts and percentages.
Baseline Chronic Medical Conditions (Yes/No)	Clinical characteristic	HealthVerity	Proportions of patients with at least one medical coding for each of the chronic medical conditions listed below during the baseline period (not mutually exclusive) will be identified. Reported as counts and percentages. • Having any of the conditions below

Variable	Role	Data source	Operational definition
			○ Asthma ○ Cancer ○ Cerebrovascular disease ○ Chronic kidney disease ○ Chronic lung diseases limited to: ▪ Bronchiectasis ▪ COPD (Chronic obstructive pulmonary disease) ▪ Interstitial lung disease ▪ Pulmonary embolism ▪ Pulmonary hypertension ○ Chronic liver diseases limited to: ▪ Cirrhosis ▪ Non-alcoholic fatty liver disease ▪ Alcoholic liver disease ▪ Autoimmune hepatitis ○ Cystic fibrosis ○ Diabetes mellitus, type 1 ○ Diabetes mellitus, type 2 ○ Disabilities, including Down syndrome ○ Heart conditions (such as heart failure, coronary artery disease, or cardiomyopathies) ○ Mental health conditions limited to: ▪ Mood disorders, including depression ▪ Schizophrenia spectrum disorders ○ Neurologic conditions limited to dementia ○ Obesity (BMI >30 kg/m2 or >95th percentile in children) ○ Solid organ or blood stem cell transplantation ○ Tuberculosis *Note: racial/ethnic minority status is included in the high risk definition, but not included in HealthVerity data and therefore unable to be assessed. Pregnancy is included in the high risk definition, but is an exclusion criteria for each aim in this protocol and therefore is not listed here. Given the nature of administrative claims data, obesity islikely to be undercaptured.*
Quan-Charlson Comorbidity Index (CCI) Score	Clinical characteristic	HealthVerity	CCI will be defined during the baseline period and will be used to assign patients to CCI groups: 0, 1, 2, 3+. Reported as min, max, median (Q1-Q3) and mean (SD), as well as counts and percentages for CCI groups.
COVID-19 Vaccination Status	Clinical characteristic	HealthVerity	COVID-19 vaccination status (described later) during the baseline period. Categories will be reported as counts and percentages.
Number of All-cause	Patient	HealthVerity	Number of hospitalizations for any cause during the

Variable	Role	Data source	Operational definition
Hospitalization - Baseline	characteristic		baseline period. Reported as min, max, median (Q1-Q3) and mean (SD). The reported numbers will also be assigned to groups: 0, 1, 2, 3+, and reported as counts and percentages. NOTE: in the cases where there is any overlap of place of services, the following hierarchy will be used to determine the place of service: inpatient > emergency department > other outpatient services.
Total Length of Stay (LOS) of All-cause Hospitalization - Baseline	Patient characteristic	HealthVerity	Sum of the LOS (days) of all hospitalizations during the baseline period. Any overlapping inpatient days will be only counted once. Reported as min, max, median (Q1-Q3) and mean (SD).
Total Costs of All-cause Hospitalization - Baseline	Patient characteristic	HealthVerity	Sum of the costs of all hospitalizations during the baseline period. Reported as min, max, median (Q1-Q3) and mean (SD). NOTE: in the cases where there is any overlap of place of services, the following hierarchy will be used to determine the place of service: inpatient > emergency department > other outpatient services.
Number of All-cause Outpatient Visits - Baseline	Patient characteristic	HealthVerity	Number of outpatient visits during the baseline period. Reported as min, max, median (Q1-Q3) and mean (SD). The reported numbers will also be assigned to groups: 0, 1, 2, 3+, and reported as counts and percentages. Same measurements will be reported for the following sub-groups: office visits, emergency department, urgent care claims, others.
Total Costs of All-cause Outpatient Visits - Baseline	Patient characteristic	HealthVerity	Sum of the costs of all outpatient visits during the baseline period. Reported as min, max, median (Q1-Q3) and mean (SD). Same measurements will be reported for the following sub-groups: office visits, emergency department, urgent care claims, others.
Number of All Pharmacy Claims – Baseline	Patient characteristic	HealthVerity	Number of pharmacy claims during the baseline period. Reported as min, max, median (Q1-Q3) and mean (SD).
Total Costs of All Pharmacy Claims - Baseline	Patient characteristic	HealthVerity	Sum of the costs of all pharmacy claims during the baseline period. Reported as min, max, median (Q1-Q3) and mean (SD).
Total Costs of All Healthcare Claims - Baseline	Patient characteristic	HealthVerity	Sum of the costs of all healthcare, including inpatient, outpatient and pharmacy claims during the baseline period. Reported as min, max, median (Q1-Q3) and mean (SD).

Table 2. Follow-up Variables

Variable	Role	Data source	Operational definition
COVID-19 Hospitalization Status Within 30 Days (Yes/No)	Follow-up	HealthVerity	Whether patients had any inpatient COVID-19 hospitalizations within the 30-day period after the index date. Categories of Yes/No reported as counts and percentages.
Follow-up Duration	Follow-up	HealthVerity	Duration of the follow-up period (months). Reported as min, max, median (Q1-Q3) and mean (SD).
All-cause Hospitalization - Follow-up (Yes/No)	Outcome	HealthVerity	Proportions of patients with all-cause hospitalizations during the follow-up period. Reported as counts and percentages.
Number of All-cause Hospitalizations - Follow-up	Outcome	HealthVerity	Number of hospitalizations for any cause during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
Total LOS of All-cause Hospitalizations - Follow-up	Outcome	HealthVerity	Sum of the LOS of all hospitalizations during the follow-up period. Any overlapping inpatient days will only be counted once. Reported as min, max, median (Q1-Q3) and mean (SD).
Total Costs of All-cause Hospitalizations - Follow-up	Outcome	HealthVerity	Sum of the costs of all hospitalizations during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
All-cause Outpatient Visits - Follow-up (Yes/No)	Outcome	HealthVerity	Proportions of patients with all-cause outpatient visits during the follow-up period. Reported as counts and percentages.
Number of All-cause Outpatient Visits - Follow-up	Outcome	HealthVerity	Number of outpatient visits during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD). Same measurements will be reported for the following sub-groups: office visits, emergency department, urgent care claims, others.
Total Costs of All-cause Outpatient Visits - Follow-up	Outcome	HealthVerity	Sum of the costs of all outpatient visits during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD). Same measurements will be reported for the following sub-groups: office visits, emergency department, urgent care claims, others.
All Pharmacy Claims - Follow-up (Yes/No)	Outcome	HealthVerity	Proportions of patients with any pharmacy claims during the follow-up period. Reported as counts and percentages.
Number of All Pharmacy Claims – Follow-up	Outcome	HealthVerity	Number of pharmacy claims during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
Total Costs of All Pharmacy Claims - Follow-up	Outcome	HealthVerity	Sum of the costs of all pharmacy claims during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
Total Costs of All Healthcare Claims - Follow-up	Outcome	HealthVerity	Sum of the costs of all healthcare, including inpatient, outpatient and pharmacy claims during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
Number of COVID-19 Outpatient Visits -	Outcome	HealthVerity	Number of COVID-19 related outpatient visits during the follow-up period. COVID-19-related

Variable	Role	Data source	Operational definition
Follow-up			outpatient visits are identified by an outpatient visit diagnosis of COVID-19 in any position. Reported as min, max, median (Q1-Q3) and mean (SD). The same measurements will be reported for the following sub-groups: office visits, emergency department, urgent care claims, others.
Total Costs of COVID-19 Outpatient Visits - Follow-up	Outcome	HealthVerity	Sum of the costs of COVID-19 related outpatient visits during the follow-up period. COVID-19-related outpatient visits are identified by an outpatient visit diagnosis of COVID-19 in any position. Reported as min, max, median (Q1-Q3) and mean (SD). The same measurements will be reported for the following sub-groups: office visits, emergency department, urgent care claims, others.
Number of COVID-19 Hospitalization - Follow-up	Outcome	HealthVerity	COVID-19-related hospitalizations during the follow-up period. COVID-19 hospitalizations are identified by a hospital discharge diagnosis of COVID-19 in any position. Reported as min, max, median (Q1-Q3) and mean (SD).
Total LOS of COVID-19 Hospitalization - Follow-up	Outcome	HealthVerity	Sum of the LOS of COVID-19-related hospitalizations during the follow-up period. Any overlapping inpatient days will only be counted once. Reported as min, max, median (Q1-Q3) and mean (SD).
Total Costs of COVID-19 Hospitalization - Follow-up	Outcome	HealthVerity	Sum of the costs of COVID-19-related hospitalizations during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
COVID-19 Hospitalization Status (Yes/No) - Follow-up	Outcome	HealthVerity	Proportions of patients with or without COVID-19-related hospitalizations during the follow-up period. Reported as counts and percentages.
Total Costs of COVID-19 Medical Claims - Follow-up	Outcome	HealthVerity	Sum of the costs of all COVID-19-related medical claims, including inpatient and outpatient medical claims during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
IMV (Yes/No) – Within 30 Days	Outcome	HealthVerity	Proportions of patients with IMV (i.e., intubation and MV, extracorporeal membrane oxygenation [ECMO]) usage during the first 30 days of the follow-up period, among patients with any hospitalizations in the follow-up period. IMV will be identified by HCPCS and CPT-4 procedure codes. Reported as counts and percentages.
IMV Usage Duration – Within 30 Days	Outcome	HealthVerity	Number of days of IMV usage during the first 30 days of the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
ICU Admission (Yes/No) – Within 30 Days	Outcome	HealthVerity	Proportions of patients with ICU admission during inpatient admissions during the first 30 days of the follow-up period, among patients with any hospitalizations in the follow-up period. ICU admission will be identified by hospital revenue

Variable	Role	Data source	Operational definition
			codes, HCPCS, and CPT-4 procedure codes. Reported as counts and percentages.
ICU Duration – Within 30 Days	Outcome	HealthVerity	Number of days of ICU stay during the first 30 days of the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
IMV (Yes/No) - Entire Follow-up	Outcome	HealthVerity	Proportions of patients with IMV (i.e., intubation and MV, extracorporeal membrane oxygenation [ECMO]) usage during the follow-up period, among patients with any hospitalizations in the follow-up period. IMV will be identified by HCPCS and CPT-4 procedure codes. Reported as counts and percentages.
IMV Usage Duration - Entire Follow-up	Outcome	HealthVerity	Number of days of IMV usage during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
ICU Admission (Yes/No) - Entire Follow-up	Outcome	HealthVerity	Proportions of patients with ICU admission during inpatient admissions within the follow-up period, among patients with any hospitalizations in the follow-up period. ICU admission will be identified by hospital revenue codes, HCPCS, and CPT-4 procedure codes. Reported as counts and percentages.
ICU Duration- Entire Follow-up	Outcome	HealthVerity	Number of days of ICU stay during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
Oxygen Supplementation (Yes/No)	Outcome	HealthVerity	Proportions of patients with oxygen supplementation during the follow-up period, among patients with any hospitalizations in the follow-up period. Oxygen supplementation will be identified by HCPCS and CPT-4 procedure codes. Reported as counts and percentages.
Oxygen Supplementation Duration	Outcome	HealthVerity	Number of days of oxygen supplementation during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
Antiviral COVID Drug Usage Within 30 Days After the Index Date (Yes/No)	Outcome	HealthVerity	Proportions of patients with antiviral COVID drug use within the 30-day period after the index date. Reported as counts and percentages.
COVID-19 Reinfection (Yes/No) – Follow-up	Outcome	HealthVerity	Proportions of patients with COVID-19 reinfection during the follow-up period. The COVID-19 reinfection is defined as another COVID-19 diagnosis at least 90 days after the previous COVID-19 diagnosis. Reported as counts and percentages.
Time to COVID-19 Reinfection – Follow-up	Outcome	HealthVerity	Days from the index date to the first claim date of COVID-19 reinfection during the follow-up period. Reported as min, max, median (Q1-Q3) and mean (SD).
All-cause Mortality (Yes/No)	Outcome	HealthVerity	Proportions of patients with all-cause mortality during the index hospitalization or the follow-up

Variable	Role	Data source	Operational definition
			period. Reported as counts and percentages. *Note: Death information is available and is sourced from over 40,000 public and private sources nationally. The coverage rate, estimated at over 90% of CDC reported deaths through present day, is based on the number of deaths in the Fact of Death Mortality Index compared to the number of deaths reported by the CDC. Due to HIPAA privacy restrictions, date of death is reported as month and year. Cause of death is not available.*

COVID-19 Vaccination Status:

The exposure of interest is the BNT162b2 Bivalent vaccine.

A patient will be considered vaccinated from 14 days after their vaccination date.

The source of the exposure variables are the California and Louisiana vaccine registries and the HealthVerity claims database. Clinical Vaccines Administered code (CVX), National Drug Codes (NDC) and Current Procedural Terminology® (CPT) codes.

BNT162b2 Bivalent vaccinations will be enumerated using the following hierarchical definition:

1. *Any person in HealthVerity claims data who has a record in the California/Louisiana Immunization Registry with a CVX code indicating having received the BNT162b2 Bivalent vaccine will be considered vaccinated 14 days after registry event date.*

2. *Any person in HealthVerity claims data who does not have a record in the California/Louisiana Immunization Registry but has a pharmacy claim with an NDC code indicating an BNT162b2 Bivalent vaccine dose will be considered vaccinated 14 days after claim date.*

3. *Any person in HealthVerity claims data who does not have a record in the California/Louisiana Immunization Registry and does not have a pharmacy claim with an NDC code indicating an BNT162b2 Bivalent vaccine dose but has a medical claim with a CPT code indicating an BNT162b2 Bivalent vaccine dose will be considered vaccinated 14 days after claim date.*

4. *Any person in HealthVerity claims data who does not have a record in the California/Louisiana Immunization Registry, does not have a pharmacy claim with an NDC code indicating an BNT162b2 Bivalent vaccine dose and does not have a medical claim with a CPT code indicating an BNT162b2 Bivalent mRNA vaccine dose will be considered BNT162b2 Bivalent unvaccinated.*

Patients will be categorized into the following groups based on their COVID-19 vaccination status:

- **BNT162b2 Bivalent mRNA Vaccine:** Patients who received at least one dose of a BNT162b2 Bivalent mRNA vaccine in the baseline period, at least 14 days prior to the index date.

- **No Vaccine:** Patients who did not receive any COVID-19 vaccines in the up to 12 months prior to the index date. For patients with baseline periods of less than 12 months, all available records in the past up to 12 months periods will be searched.

- **Other Vaccines:** Patients who were not in the above described 2 vaccination groups. They may include patients who received any type of COVID-19 vaccine in the up to 12 months prior to index date, but did not meet the BNT162b2 Bivalent Vaccine group requirements described above. Some examples of these include patients 1) who had other types of mRNA COVID-19 vaccine and no doses of a BNT162b2 Bivalent vaccine in the baseline period; 2) who received non-mRNA vaccines, etc.

5.4. Data Sources

The source of the exposure variables are the California and Louisiana vaccine registries and the HealthVerity Claims database. Containing only de-identified data, the HealthVerity database is Health Insurance Portability and Accountability Act (HIPAA) compliant.

5.5. Study Size

It is expected that there are more than 50,000 potential COVID-19 patients for this study in the data source.

Given the expected number of COVID-19 patients in the study, power calculations were conducted for costs under various scenarios using two-part models.[1] The following assumptions were made: 50% of cases have zero costs, 60% of controls have zero costs, 4 controls matched to each case and a significance level of 0.05. Under various levels of standard deviation ($10,000 and under), means of non-zero costs among the controls ($100 and over) and relative differences in costs between cases and controls (1.15 and over), power exceeded 0.99. Therefore, even after

accounting for potential dropout and the use of robust standard errors due to matching, this sample size is sufficient for the study's purpose.

5.6. Data Analyses

Descriptive statistics will be utilized to summarize patient demographic and clinical characteristics in addition to the measured HCRU and cost outcomes. Counts and percentages will be reported for categorical variables. Continuous variables will be summarized using minimum, maximum, median (Q1-Q3) and mean (SD).

All cost measurements will be adjusted to 2024 USD using Consumer Price Index medical care component.

All study measurements will be compared between the COVID-19 and Control cohorts.

HCRU and Cost Outcomes at Different Time Periods:

For HCRU and cost outcomes, all measurements will be provided for the following time periods:

- Post-acute phase: 1-11 months (measured starting from 30 days after the index date)
  - 1 to <2 months (1 month post-acute period)
  - 1 to <4 months (3 month post-acute period)
  - 1 to <7 months (6 month post-acute period)
  - 1 to <10 months (9 month post-acute period)
  - 1 to <12 months (11 month post-acute period)

Within this study, 1 month is generally considered to be equivalent to 30 days, unless otherwise specified.

NOTE: HCRU and cost will not be measured during the acute phase of 0 to <1 month after the index date since this is not within the study scope.

Stratification Analysis:

Patients will be further stratified by the following variables:

- **Age group:**
  - Among the overall study population, patients will be stratified by their age at the index date as follows:

  - Adult and Pediatric
    - Adult: Patients ≥ 18 years at the index date.
    - Pediatric: Patients 6 months to < 18 years at the index date.
  - Detailed Age Groups
    - Adult: 18 – 29, 30 – 49, 50 – 64, 65 – 74, 75+ years
    - Pediatric: 0 – < 2, 2 – 4, 5 – 11, 12 – < 18 years.

- **COVID-19 Vaccination Status:**
  - BNT162b2 Bivalent mRNA Vaccine
  - No Vaccine
  - NOTE: Other Vaccines sub-group for vaccination status will not be used for this stratification analysis.
- **Site of Care: COVID-19 Hospitalization Status Within 30 Days (Yes/No):**
  - Yes
  - No
- **ICU Admission (Yes/No) – Within 30 Days:**
  - Among the overall study population, patients will be stratified by status of ICU Admission (Yes/No) – Within 30 Days
- **IMV Usage (Yes/No) – Within 30 Days:**
  - Among the overall study population, patients will be stratified by status of IMV usage (Yes/No) – Within 30 Days

Sensitivity Analysis:

Sensitivity analyses planned for the study will include:

1. **COVID-19 reinfection:** In the sensitivity analysis, a COVID-19 patient's follow-up period may be further censored at the first COVID-19 reinfection date (if any). The COVID-19 reinfection is defined as another COVID-19 diagnosis at least 90 days of gap after the previous COVID-19 diagnosis. The study measurements among the overall patient population (without stratifications) will be reported for this sensitivity population.

2. **Open-source Cost:** A portion of patients will have claims data that is linked to open-source cost information. A sensitivity analysis of the overall study analysis (not including the stratification analyses) will be carried out among such linked patients. The open-source cost data will be used to represent the cost information in this sensitivity analysis.

Statistical Analysis:

Descriptive Statistics:

Descriptive statistics will be utilized to summarize patient demographic, clinical characteristics, and HCRU and cost outcomes. Counts and percentages will be reported for categorical variables. Continuous variables will be summarized using minimum, maximum, median (Q1-Q3) and mean (SD). The count and percentage of patients with missing data will be reported.

Comparison Statistics:

All study measurements will be compared between the COVID-19 and Control cohorts. Standardized mean differences (SMD) will be reported for all variables.

Regression Analysis:

Multivariable regression analysis will be used to evaluate the potential predictors of long-term HCRU and costs associated with COVID-19.

1. The dependent variables of such regression models will include: number of all-cause hospitalizations during the follow-up period, total cost (charge and allowed amount) of all-cause hospitalizations during the follow-up period, number of COVID-19 hospitalizations during the follow-up period, total cost (charge and allowed amount) of COVID-19 hospitalizations during the follow-up period, total cost (charge and allowed amount) of all healthcare claims during the follow-up period, and total cost (charge and allowed amount) of COVID-19-related healthcare claims during the follow-up period.
2. The independent variables (model covariates) will include: patient cohort (COVID-19, Control), COVID-19 vaccination status, age, gender, state, insurance type, Charlson Comorbidity Index (CCI), number of all-cause hospitalizations during the baseline period, number of all-cause outpatient visits during the baseline period, total all-cause healthcare allowed amount during the baseline period, and index year-month.
3. Regression models:
    a. Generalized linear models with Negative Binomial distribution will be used for the analysis of all count data (eg, number of hospitalizations, etc.).
    b. Generalized linear models with log transformation and Gamma distribution will be used for the analysis of all cost data.

c. All regression models will produce point estimates and 95% confidence intervals (CIs). When reporting 95% CIs, the models will also add clustering on the patient matches to produce robust standard errors.

d. Akaike information criterion (AIC) will be used to evaluate whether a zero-inflated negative binomial model is a better fit for the count data. For cost-related regression models, AIC will be used to compare distributions and to compare with two-part models to see which models are more appropriate for handling zero costs.

Below is a summary table of such regressions:

Description	Exposure (aka independent variable)	Covariates	Outcome (aka dependent variable)	Type of regression
number of all-cause hospitalizations during the follow-up period, number of COVID-19 hospitalizations during the follow-up period	COVID-19, Control	COVID-19 vaccination status, age, gender, state, insurance type, Charlson Comorbidity Index (CCI), number of all-cause hospitalizations during the baseline period, number of all-cause outpatient visits during the baseline period, total all-cause healthcare allowed amount during the baseline period, and index year-month	number of all-cause hospitalizations during the follow-up period, number of COVID-19 hospitalizations during the follow-up period	Generalized linear model with negative binomial distribution
total cost (charge and allowed amount) of all-cause hospitalizations during the follow-up period, total cost (charge and allowed amount) of	COVID-19, Control	COVID-19 vaccination status, age, gender, state, insurance type, Charlson Comorbidity Index (CCI), number of all-cause hospitalizations during the baseline period, number of all-cause outpatient visits during the baseline period, total all-cause	total cost (charge and allowed amount) of all-cause hospitalizations during the follow-up period, total cost (charge and allowed amount) of COVID-19 hospitalizations during the follow-	Generalized linear model with gamma distribution

Description	Exposure (aka independent variable)	Covariates	Outcome (aka dependent variable)	Type of regression
COVID-19 hospitalizations during the follow-up period, total cost (charge and allowed amount) of all healthcare claims during the follow-up period, and total cost (charge and allowed amount) of COVID-19-related healthcare claims during the follow-up period		healthcare allowed amount during the baseline period, and index year-month	up period, total cost (charge and allowed amount) of all healthcare claims during the follow-up period, and total cost (charge and allowed amount) of COVID-19-related healthcare claims during the follow-up period	

Inverse Probability of Censoring Weighting (IPCW):

Imbalances in patient characteristics may develop over the follow-up period due to the artificial censoring of patients assigned as controls who later develop COVID-19 or have other censoring events. Inverse probability censored weighting (IPCW) can be used to compensate for such censoring. Therefore, the regression models may also be generated after applying IPCW as described below.[2,3]

Within each cohort, person-time records will be divided into short intervals depending on computational feasibility (eg, 10 days). Weights will be generated as the inverse of the probability of remaining uncensored at each time, as predicted from Cox regression models using the same baseline patient characteristics employed during matching. Weights will be stabilized by multiplying by the proportion of patients within the cohort who remain uncensored. Among controls, two models will be used: the first for future COVID-19 status and the second for

other censoring events. Stabilized weights from the two models will be multiplied to produce the final weights.

Summaries of the stabilized weights will be presented by cohort. Among patients assigned as COVID-19 cases, SMDs and baseline characteristics will be reported by administrative censoring status during the follow-up period, before and after IPCW. Among patients assigned as controls, SMDs and baseline characteristics before and after IPCW will be reported among those who, during the follow-up period, are: 1) not censored and remain undiagnosed with COVID-19; 2) become censored due to a COVID-19 diagnosis; and 3) become censored due to other events.

Other Statistical Descriptions:

P-value of 0.05 will be used as the threshold to determine the statistical significance of all comparisons and regressions.

All data analyses will be conducted using SAS 9.4 (SAS Institute, Cary, NC) or R (R Foundation, Indianapolis, IN).

6. REFERENCES

1. Lachenbruch PA. Power and sample size requirements for two-part models. Stat Med. 2001;20(8):1235-1238.

2. Hernán MA RJ. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC; 2020.

3. Maringe C, Benitez Majano S, Exarchakou A, et al. Reflection on modern methods: trial emulation in the presence of immortal-time bias. Assessing the benefit of major surgery for elderly lung cancer patients using observational data. Int J Epidemiol. 2020;49(5):1719-1729.

4. Kadri SS, Gundrum J, Warner S, et al. Uptake and Accuracy of the Diagnosis Code for COVID-19 Among US Hospitalizations. JAMA. 2020;324(24):2553-2554.

7. APPENDICES

7.1. Appendix A: Table shells

The list of planned outputs and table shells will be presented in a separate document.

Document Approval Record

Document Name:	C4591065 Non Interventional Study Statistical Analysis Plan (SA)) RA VEN Aim 3B_30May2024 FINAL
Document Title:	C4591065 Non Interventional Study Statistical Analysis Plan (SA)) RA VEN Aim 3B_30May2024 FINAL

Signed By:	Date(GMT)	Signing Capacity
PPD	30-May-2024 18:00:56	Final Approval

Non-Interventional Study

C4591065

Effectiveness of BNT162b2 formulations using state vaccine registry and insurance claims data

Aim 1

Statistical Analysis Plan

(SAP)

Version: 1.0

Author: PPD (PPD, **Pfizer Inc, New York, NY, USA)**

Date: 16-Jan-2024

TABLE OF CONTENTS

1. AMENDMENTS FROM PREVIOUS VERSION(S)

NA. First version.

2. LIST OF ABBREVIATIONS

Abbreviation	Definition
aHR	Adjusted hazard ratio
CDC	Centers for Disease Control and Prevention
CI	Confidence interval
CPT	Current Procedural Terminology®
CVX	Clinical Vaccines Administered code
EUA	Emergency use authorization
FDA	U.S. Food and Drug Administration
HV	HealthVerity
ICD-10-CM	International Classification of Diseases, Tenth Revision, Clinical Modification
ICU	Intensive care unit
NDC	National Drug Codes
RSV	Respiratory syncytial virus
SAP	Statistical Analysis Plan
SMD	Standardized mean differences
NPI	National provider identifier
VE	Vaccine effectiveness

Note: in this document, any text taken directly from the non-interventional (NI) study protocol is ***italicized***.

3. RATIONALE AND BACKGROUND

Until recently, vaccine registries were not available for de-identification and linkage to secondary claims-based databases. However, through a collaboration with HealthVerity, a leading supplier of secondary health data, the State of California and the State of Louisiana have made their vaccine registries available for linkage for research purposes. California and Louisiana deidentify the registries, using HealthVerity's software to convert each patient's protected health information to a unique HealthVerity ID which can then be linked to the full HealthVerity database to gather clinical history and outcomes.

The purpose of this document is to detail the statistical methods and analyses that will be used in the HealthVerity claims database linked to the State of California/Louisiana's vaccine registries to address Aim 1 of the study protocol. Separate SAPs will be produced to address the other aims of the protocol.

4. STUDY OBJECTIVES AND HYPOTHESES

The objective of Aim 1 of this study is to *evaluate the vaccine effectiveness* (VE) *of BNT162b2 formulations in non-pregnant, non-immunocompromised adults against medically attended COVID-19, mortality and healthcare resource utilization by age and adapted vaccine formulation.*

This study is descriptive in nature and as such does not include any pre-specified hypotheses.

5. RESEARCH METHODS

5.1. Study Design

This will be a retrospective cohort study using data from patients who were California or Louisiana residents in the HealthVerity claims database.

Cohort identification will begin from the date of Emergency Use Authorization (EUA) or U.S. Food and Drug Administration (FDA) *approval for the BNT162b2 formulation. Study follow-up will occur from FDA EUA or approval date until COVID-19 diagnosis or other outcome of interest (hospitalization, emergency department visit, outpatient encounter, end of medical and/or pharmacy enrollment, or death [according to the mortality database]).*

The authorization/approval dates for each BNT162b2 formulation that will be used in this study are provided in Table 5.1. Note that EUA (for children age 6 months to 11 years) and approval (age 12+) status for the bivalent vaccine was revoked on the date the XBB.1.5 formulation became available.

Table 5.1 Timeline of Key BNT162b2 Regulatory Milestones Applicable to Non-immunocompromised Adults

Vaccine formulation	Age group	Event
BNT162b2 monovalent XBB.1.5 COVID-19 vaccine	Age 12 and older	September 11, 2023: U.S. Food and Drug Administration (FDA) approval
BNT162b2 bivalent BA.4/5 COVID-19 vaccine	Age 12 and older	August 31, 2022: FDA EUA

All persons will be considered unvaccinated from date of FDA date of vaccine authorization/approval until, where applicable, 13 days after receipt of vaccine. In order to allow for time after dose for meaningful immunologic protection to accrue, persons will be considered vaccinated 14 days after vaccination.

The full data available, from December 1, 2018, will be used as the look-back period to define patient's characteristics, clinical history, risk factors, and healthcare utilization.

The outcomes of interest are medically attended COVID-19, mortality and healthcare resource utilization and are described in Table 5.2.

*Main summary measure: VE = (1 – adjusted hazard ratio [aHR]) * 100, comparing vaccinated to unvaccinated persons during the same time period.*

5.2. Study Population

Patients who were California or Louisiana residents in the HealthVerity claims database will be defined using the following hierarchical definition:

1. *Persons who have their enrollment file record indicating a patient location of California or Louisiana in HealthVerity claims enrollment file.*

2. *Persons who have a patient state location other than California or Louisiana in HealthVerity claims enrollment but have one or more records in the California or Louisiana Immunization Registry. Examples include people who previously lived in Texas but moved to California and received a vaccine, or persons who crossed state lines for a vaccine and remain in their non-California state in claims data.*

Patients who meet the following criteria will be included in the analysis:

5.2.1. Inclusion Criteria

1. *Born in 2005 or earlier (as age 18 or older in 2023, owing to the data providing year but not month or day of birth).*

2. *At least one year of pharmacy and medical enrollment in HealthVerity prior to index date (FDA approval/EUA date per vaccine). A gap of up to 30 days will be allowed.*

3. *Have been a State of California/Louisiana resident for at least one year prior to index date (FDA approval/EUA date per vaccine).* A gap of up to 30 days will be allowed.

5.2.2. Exclusion Criteria

1. *Currently pregnant (using 9 months lookback with a codelist for current pregnancy)*[1] *or immunocompromised (using 1 year lookback) at the time of study index* (definitions are provided in the Protocol Annex)

2. *Individuals with discrepancies in sex and/or year of birth between HealthVerity claims and California/Louisiana immunization registry datasets. In cases where sex or year of birth is unknown in either the claims or dataset but available in the other, individuals will be included and the information from the dataset where year of birth or sex is available will be used, regardless of whether the information stemmed for claims or registry data.*

3. *Persons with a diagnosis of COVID-19 (ICD-10-CM U07.1) in any setting ≤ 90 days prior to index (FDA approval/EUA date per vaccine)*

4. *Receipt of COVID-19 vaccine ≤ 60 days prior to index (FDA approval/EUA date per vaccine).*

5.2.3. Subgroups and Stratification

Results will be stratified by age group (overall, age 18-49, 50-64, 65+).

Additionally, the analysis of the primary outcome will be stratified by state of residence. As tests and measures of heterogeneity can be affected by large sample sizes, clinical relevance will be used to assess the difference in VE estimates between the two registries. If a difference is deemed clinically relevant, secondary and exploratory outcomes may be additionally stratified by state of residence.

The following subgroups will be used to evaluate relative VE for the XBB.1.5 vaccine:

- Prior receipt of BA.4/5 bivalent vaccine
- Receipt of two prior doses of mRNA COVID-19 vaccine and no prior receipt of BA.4/5 bivalent vaccine
- No prior receipt of any COVID-19 vaccine

5.3. Variables

5.3.1. Exposures

The exposure of interest is the BNT162b2 XBB.1.5 mRNA vaccine. *In the event that XBB uptake is low, we will instead examine BA.4/5 vaccine effectiveness* (Section 5.5).

Information on vaccination status will continue to be collected on or after the authorization/ approval date, therefore vaccination status will be considered as a time-varying exposure. Patients will be assigned as vaccinated or unvaccinated according to their vaccination status by each person-time record during the follow-up. A patient will be considered vaccinated from 14 days after their BNT162b2 vaccination date till the remainder of their follow-up.

The source of the exposure variables are the California and Louisiana vaccine registries and the HV (HealthVerity) Claims database. Clinical Vaccines Administered code (CVX), National Drug Codes (NDC) and Current Procedural Terminology® (CPT) codes used to identify the XBB.1.5 and BA.4/5 vaccinations are listed in the Protocol Annex.

XBB vaccinations will be enumerated using the following hierarchical definition:

1. *Any person in HealthVerity claims data who has a record in the California/Louisiana Immunization Registry with a CVX code indicating having received the BNT162b2 XBB.1.5 monovalent mRNA vaccine will be considered vaccinated 14 days after registry event date.*

2. *Any person in HealthVerity claims data who has a record in the California/Louisiana Immunization Registry with an event indicating having received the BNT162b2 XBB.1.5 monovalent mRNA vaccine will be considered vaccinated 14 days after registry event date.*

3. *Any person in HealthVerity claims data who does not have a record in the California/Louisiana Immunization Registry but has a pharmacy claim with an NDC code indicating a BNT162b2 XBB.1.5 monovalent mRNA vaccine dose will be considered vaccinated 14 days after claim date.*

4. *Any person in HealthVerity claims data who does not have a record in the California/Louisiana Immunization Registry and does not have a pharmacy claim with an NDC code indicating a BNT162b2 XBB.1.5 monovalent mRNA vaccine dose but has a medical claim with a CPT code indicating a BNT162b2 XBB.1.5 monovalent mRNA vaccine dose will be considered vaccinated 14 days after claim date.*

5. *Any person in HealthVerity claims data who does not have a record in the California/Louisiana Immunization Registry, does not have a pharmacy claim with an NDC code indicating a BNT162b2 XBB.1.5 monovalent mRNA vaccine dose and does not have a medical claim with a CPT code indicating a BNT162b2 XBB.1.5 monovalent mRNA vaccine dose will be considered BNT162b2 XBB.1.5 unvaccinated.*

BA.4/5 vaccinations will be enumerated in the same manner as XBB.1.5.

While XBB boosters are authorized for immunocompromised persons that are at least 2 months after first XBB dose, as well adults age 65+ at least 4 months after first XBB dose, only the first XBB vaccination will be analyzed. If a patient has multiple XBB.1.5 vaccination records, they will be censored at second dose.

5.3.2. Outcomes

The outcomes of interest are presented in Table 5.2.

Table 5.2 Outcomes Variables

Outcome Type	Variable	Data source(s)	Operational definition
Primary outcome	COVID-19 diagnosis	HV Claims	Any encounter regardless of diagnosis position or setting with ICD-10-CM code: U07.1.
Secondary outcomes	COVID-19 outpatient encounter	HV Claims	Outpatient encounter with ICD-10-CM code: U07.1
	COVID-19 hospitalization	HV Claims	Hospitalizations "for COVID-19" were defined as those with an inpatient encounter at an acute care facility with ICD-10-CM code U07.1 "COVID-19" that did not have an additional incidental code (for unintentional injury, physical trauma, poisoning, short-stay [<2 days] childbirth or serious psychiatric admissions), or a U07.1 hospitalization where treatment used solely for COVID-19 (eg, remdesivir) was identified regardless of accompanying diagnoses. As a sensitivity analysis, we will restrict this definition to be more conservative by requiring the incidental finding code be the most frequently claimed during the inpatient encounter, as a proxy for primary diagnosis as HealthVerity does not provide this information. As a sensitivity analysis, we will relax this definition to consider any hospitalization with a U07.1 code, without consideration of incidental findings.
	COVID-19 emergency department visit	HV Claims	Emergency department visit with ICD-10-CM code: U07.1

Outcome Type	Variable	Data source(s)	Operational definition
	COVID-19 critical illness	HV Claims	Intensive care unit [ICU] admission, mechanical ventilation, or inpatient death. If sample size allows (Section 5.5), we will evaluate VE against each component separately
	Negative control outcome, such as accidental injury, ingrown toenail or atopic dermatitis	HV Claims	Any encounter regardless of setting, with ICD-10-CM code for negative control outcomes (codes in in the Protocol Annex)
Exploratory outcomes	COVID-19 urgent care encounter	HV Claims	An urgent care encounter with ICD-10-CM code: U07.1 "COVID-19"
	All-cause mortality after COVID-19 encounter	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 encounter, regardless of setting
	All-cause mortality after COVID-19 outpatient encounter	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 outpatient encounter
	All-cause mortality after COVID-19 inpatient encounter	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 inpatient encounter
	All-cause mortality after COVID-19 emergency department encounter	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 emergency department encounter
	All-cause mortality after COVID-19 critical illness	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 critical illness

Outcome Type	Variable	Data source(s)	Operational definition
	All-cause mortality after COVID-19 urgent care encounter	HV Claims and HV Mortality Database	HealthVerity ID in Mortality Database within 3-months of COVID-19 urgent care encounter

Each outcome will be assessed independently and experiencing one outcome will not be a censoring event for another outcome. Where a patient has multiple instances of the same outcome, the first one will be used.

As exact dates of death are unavailable, deaths will be classified as occurring within 3 months of the COVID-19 encounter if they occurred in the same calendar month as the encounter or in the next three calendar months.

5.3.3. Covariates

The list of demographic and clinical characteristics that will be used to describe the cohort and considered for inclusion as covariates in models are provided in Table 5.3. Age, sex, state of residence and prior COVID-19 vaccinations will be obtained from the HealthVerity claims database and state vaccine registries. The HealthVerity claims database will be the sole source of all other variables.

All covariates will be assessed on or prior to index date and will therefore be considered as time-fixed variables during the follow-up. Categories may be collapsed or covariates may be excluded or combined if models fail to converge or if patient counts are low (Section 5.7.2). Additionally, categories with large patient counts may be split further.

Table 5.3 Demographic and Clinical Characteristics

Variable	Role(s)	Operational definition
Age	-Inclusion/ exclusion criteria -Descriptive (continuous and categorical) -Covariate (categorical) -Subgroup (categorical)	Age assessed at index date. Categorized as: 18-49, 50-64 or 65+ years.
Sex	-Inclusion/ exclusion criteria -Descriptive -Covariate	Sex assessed at index date. Male, female or unknown.
State of residence	-Inclusion/ exclusion criteria -Descriptive -Covariate -Subgroup	Defined in Section 5.2 California or Louisiana
Payor	-Descriptive -Covariate	Payor assessed at index date. Commercial, Medicaid, Medicare Advantage, Unknown, Missing

Variable	Role(s)	Operational definition
CDC-defined high risk for severe COVID-19[2]	-Descriptive (any high-risk condition and for each individual condition) -Covariate (each individual condition except age ≥ 50 years)	Assessed in the year prior to index • Age ≥ 50 years • Asthma • Cancer, as hematologic malignancies • Cerebrovascular disease • Chronic kidney disease: people receiving dialysis • Chronic lung diseases: bronchiectasis, COPD, interstitial lung disease, pulmonary embolism, pulmonary hypertension • Chronic liver diseases: cirrhosis, non-alcoholic fatty liver disease, alcoholic liver disease, autoimmune hepatitis • Cystic fibrosis • Diabetes mellitus, type 1 • Diabetes mellitus, type 2 • Disabilities, including Down syndrome • Heart conditions: heart failure, coronary artery disease, or cardiomyopathies • HIV • Mental health conditions: mood disorders including depression; Schizophrenia spectrum disorders • Neurologic conditions limited to dementia • Obesity (BMI ≥30 kg/m^2) • Primary immunodeficiencies • Smoking, current and former • Solid organ or blood stem transplantation • Tuberculosis • Use of corticosteroids or other immunosuppressive medications The lists of codes are provided in the Protocol Annex. Note: due to limitations of claims data, "physical inactivity" will not be measurable and has been removed from the list.

Variable	Role(s)	Operational definition
Charlson-Deyo Comorbidity Index[3]	-Descriptive (continuous and categorical) -Covariate (categorical)	Assessed in the year prior to index Conditions are scored as: 1 point for each of: • Myocardial infarction • CHF • Peripheral vascular disease • Cerebrovascular disease • Dementia • Chronic pulmonary disease • Connective tissue disease • Peptic ulcer disease • Mild liver disease • Diabetes mellitus without chronic complications 2 points for each of: • Hemiplegia • Moderate to severe renal disease • Diabetes with chronic complications • Cancer, including leukemia and lymphoma 3 points for moderate or severe liver disease 6 points for each of: • Metastatic carcinoma • HIV/AIDS Categorized as: 0, 1 or 2+ The list of codes is provided in the Protocol Annex
Receipt of skilled nursing care and/or long-term care facility stay in year prior to index	-Descriptive -Covariate	Assessed in the year prior to index date Yes or No The list of codes is provided in the Protocol Annex
Wellness visit	-Descriptive -Covariate	Assessed in the year prior to index date Yes or No The list of codes is provided in the Protocol Annex
Decreased functional status	-Descriptive -Covariate	Assessed in the year prior to index date Yes or No The list of codes is provided in the Protocol Annex

Variable	Role(s)	Operational definition
Influenza vaccination in year prior to index	-Descriptive -Covariate	Assessed in the year prior to index Yes or No The list of codes is provided in the Protocol Annex
Pneumococcal vaccination using all-available time prior to index	-Descriptive -Covariate	Assessed in the year prior to index Yes or No The list of codes is provided in the Protocol Annex
Herpes zoster vaccination using all-available time prior to index	-Descriptive -Covariate	Assessed in the year prior to index Yes or No The list of codes is provided in the Protocol Annex
Outpatient visits prior to index	-Descriptive (continuous and categorical) -Covariate (categorical)	Number of outpatient visits in 365 days prior to index Categorized as: 0, 1 or 2+ The list of codes is provided in the Protocol Annex, and includes office/clinic visits as well as telehealth.
Inpatient admissions prior to index	-Descriptive (continuous and categorical) -Covariate (categorical)	Number of inpatient admissions in 365 days prior to index Categorized as: 0, 1 or 2+ The list of codes is provided in the Protocol Annex
Emergency department visits prior to index	-Descriptive (continuous and categorical) -Covariate (categorical)	Number of emergency department visits in 365 days prior to index Categorized as: 0, 1 or 2+ The list of codes is provided in the Protocol Annex
Telehealth visits prior to index	-Descriptive (continuous and categorical) -Covariate (categorical)	Number of telehealth encounters (CPT: 99201-99215) in 365 days prior to index Categorized as: 0, 1 or 2+
Number of documented SARS-CoV-2 tests in the 180 days prior to index	-Descriptive (continuous and categorical) -Covariate (categorical)	Number of visits in any setting with CPT in 365 days prior to index: 86408, 86409, 0225U, 0226U Categorized as: 0, 1 or 2+
Prior post-COVID-19 conditions	-Descriptive -Covariate	Assessed in the year prior to index, any encounter regardless of setting with ICD-10-CM U09.9 “Post-COVID-19 condition” Yes or No

Variable	Role(s)	Operational definition
Prior COVID-19 diagnosis	-Descriptive -Covariate	Any encounter with ICD-10-CM code: U07.1 between 90 and 365 days prior to index Yes or No
Prior bivalent BA.4/5 COVID-19 vaccine	-Descriptive -Subgroup	Where the BNT162b2 XBB.1.5 vaccine is the exposure of interest. Using all-available lookback data, received bivalent BA.4/5 COVID-19 vaccine (regardless of brand) prior to index Yes or No The list of codes is provided in the Protocol Annex
At least two prior doses of mRNA vaccine and no bivalent BA.4/5 COVID-19 vaccine	-Descriptive -Subgroup	Where the BNT162b2 XBB.1.5 vaccine is the exposure of interest. Using all-available lookback data, received at least two prior doses of mRNA vaccine (regardless of formulation or brand) and did not receive bivalent BA.4/5 COVID-19 vaccine. Note: the doses must be at least 21 days apart. Yes or No The list of codes is provided in the Protocol Annex
Never COVID-19 vaccinated	-Descriptive -Subgroup	Using all-available lookback data, no prior receipt of a COVID-19 vaccine (regardless of formulation) Yes or No The list of codes is provided in the Protocol Annex
Time since last COVID-19 vaccine	-Descriptive -Subgroup	Using all-available lookback data, time in months between index date and prior receipt of a COVID-19 vaccine (regardless of formulation) The list of codes is provided in the Protocol Annex
Predominant variant	Time-varying covariate, to produce stratified analyses	Using a time-varying covariate, to indicate whether an event happened in the XBB era (September 11, 2023 – January 5, 2024) or JN1 era (January 6, 2024 – end of data collection [March 11, 2024])

5.4. Data sources

The sources of the data are the State of California and the State of Louisiana vaccine registries linked to closed claims in HealthVerity.

5.4.1. California and Louisiana Immunization Registries

The California Immunization Registry is run by the California Department of Public Health and collects nearly all records of COVID-19 vaccinations administered in the state of California.

Data include a unique HealthVerity person identification code, vaccination event date and Clinical Vaccines Administered (CVX) code. CVX codes are unique to brand and formulation of vaccine.

Data is linked using HealthVerity's tokenization software, whereby specific patient identifiers are passed through the software and a de-identified patient ID is assigned. The same tokenization process is applied to closed claims sourced by HealthVerity, and patients can thus be linked across data sources. The tokenization process occurs outside of Pfizer, no personal identifying information will be transferred, and Pfizer will only have access to de-identified data. The data comply with Health Insurance Portability and Accountability Act (HIPAA) regulations.

The Louisiana Vaccine Registry is structured and accessed in the same way as California.

5.4.2. HealthVerity's Administrative Claims Data

HealthVerity closed claims will be limited to approximately 19 million patients residing in California and approximately 3.8 million in Louisiana.

HealthVerity's description of their data is that they are generally representative of the age and sex distribution of the population. For our specific research context, samples created from HealthVerity claims also generally follow California's demographic patterns with respect to age & sex.

The California subset of HealthVerity does not include patients in the Kaiser Permanente network. Insurers contributing to closed claims in both states include a mix of commercial payers, Medicare Advantage/Part C pains, and Medicaid Managed Care plans. Data elements include patient demographic information, inpatient/outpatient visit-level information such as diagnoses, procedures, and length of stay, hospital characteristics, and medication information. Owing to the nature of claims, the data represent the final set of diagnoses over the course of the hospitalization sent to the patient's insurer for reimbursement, with diagnosis prioritization assigned by clinicians or hospital staff. For inpatient encounters, the data will be assumed to represent the overarching events during hospitalization and may be less subject to rule out diagnosis codes than electronic health record data.

Death information is available and is sourced from over 40,000 public and private sources nationally. The coverage rate, estimated at over 90% of CDC reported deaths through present day, is based on the number of deaths in the Fact of Death Mortality Index compared to the number of deaths reported by the CDC. Due to HIPAA privacy restrictions, date of death is reported as month and year. Cause of death is not available.

5.5. Sample Size and Power Calculations

Since the current study utilizes retrospective deidentified data without ability to recruit to a target, sample size calculations are not applicable. Further, this is a descriptive study without a priori specified hypotheses. However, we examined the precision of the expected XBB.1.5 absolute VE estimates for the outcomes under various scenarios.

Assuming uptake similar to that seen with the BA.4/5 vaccine, and non-differential uptake by age groups, we could anticipate in California:

(75% of 12 million are adult)(16% uptake)*(60% of doses were Pfizer-BioNTech)*

~ 864,000 adult XBB doses in California

Among these 9 million adults, 95% were assumed to be non-immunocompromised and used as the sample size at start of follow-up. Assuming similar absolute VEs to those seen with the BA.4/5 vaccine for symptomatic COVID-19 infections, we could anticipate estimates between 20 and 40%.[4] To assess the precision for a range of outcomes, incidence rates of 50, 100, 1000 and 10,000 cases per 100,000 unvaccinated persons during the study period were considered. A censoring rate of 5% by the end of the study period was used. A uniform vaccination rate between 14 days after study index and the end of follow-up was assumed.

Survival and censoring times were simulated assuming exponential distributions. Vaccination status was modelled as a time-varying covariate.[5] Hazard ratios were input as 1-(VE/100). Simulations were run a minimum of 10 times per scenario. Cox regression models were run on each simulated dataset to estimate the 95% CIs for the log HRs. Means were generated and exponentiated to obtain the 95% CIs for the VEs using (1- HR) x 100 and used to describe VE precision. Simulations were repeated on 25% of the initial sample size to represent the precision in a potential subgroup.

Table 5.4 Simulated 95% Confidence Intervals [Total Number of Events] for BNT162b2 XBB.1.5 Vaccine Effectiveness Estimates

Number at Start of Follow-up	Vaccine Effectiveness, %	Cases per 100,000 Unvaccinated Persons by End of Study			
		50	100	1000	10,000
8,550,000	20	7, 29 [4116]	12, 27 [8242]	18, 23 [81,702]	19, 21 [782,794]
	30	20, 40 [4104]	22, 36 [8144]	28, 32 [81,108]	29, 31 [777,275]
	40	30, 49 [4019]	33, 46 [8095]	38, 42 [80,490]	39, 41 [771,141]
2,137,500	20	-3, 39 [1026]	4, 34 [2072]	15, 24 [20,348]	18, 21 [195,697]
	30	10, 49 [1022]	14, 42 [2042]	25, 34 [20,297]	28, 31 [194,371]
	40	20, 57 [1016]	27, 52 [2022]	36, 44 [20,104]	39, 41 [192,827]

Table 5.4 shows that there should be sufficient sample size in the California registry to produce 95% CIs for VEs that exclude 0 for the evaluated VEs and incidence rates, except for the rare event scenario in a subgroup. Assuming similar VEs in the two state registries, the inclusion of patients from the Louisiana registry would increase the precision further.

5.6. Missing Data

No imputation for missing values will be performed. Subject to patient counts (Section 5.7.2), missing or 'unknown' values will be reported and analyzed as separate categories and included in the totals.

5.7. Statistical Methodology and Analyses

5.7.1. General Considerations

Descriptive statistics will be used to summarize variables. For categorical variables, the frequency and percentage of patients in each category will be presented. Percentages will be based on the total number of relevant patients. For continuous variables, data will be presented as means, standard deviations, medians, 25th and 75th percentiles, minimums and maximums. Continuous variables may be additionally categorized and analyzed as categorical variables.

No statistical hypotheses are specified and there will be no correction for multiple comparisons.

Time in days between two dates will be calculated as (date2 – date1) + 1, where date2 is on or after date1. Where date2 is before date1, time in days between two dates will be calculated as (date2 – date1). Index date will be considered as Day 1. There is no Day 0.

Unless stated otherwise, one year is defined as 365 days, and one month is defined as 30 days.

All data analysis will be conducted using statistical software SAS version 9.4 or R version 4.1.0 or later.

5.7.2. Minimum Sample Size Requirements

In order to maintain patient de-identification, outcomes, covariates and subgroups will have at least 10 patients in each category to be included in the analysis.

In addition, a minimum of five events per variable (the number of degrees of freedom) will be required in each level of the outcome variable in the adjusted Cox regression models to minimize the potential for biased estimates.[6] Note, as an example, a categorical predictor variable with three levels contributes two degrees of freedom.

5.7.3. Index Date, Baseline and Follow-up

The index date for each individual will be defined as the BNT162b2 EUA or FDA approval date (Table 5.1). Age and sex will be assessed at index date. The baseline period used to assess other patient characteristics are specified for each covariate in Table 5.3.

Unvaccinated follow-up time will occur from the date of FDA authorization/approval until 13 days after the receipt of vaccination, the outcome of interest, or are censored at the first of the following: 6 months of follow-up, end of medical and/or pharmacy enrollment, receipt of a non-Pfizer BioNTech vaccine. Where mortality is not the outcome of interest, follow-up time will be additionally censored at death. Where mortality is the outcome of interest, follow-up time will be additionally censored at a death not meeting the outcome criteria.

Vaccinated follow-up time will occur from 14 days after the receipt of vaccination until the outcome of interest, or are censored at the first of the following: 6 months of total follow-up (ie, 11 March 2024), *end of medical and/or pharmacy enrollment, receipt of a non-Pfizer BioNTech vaccine or receipt of a second BNT162b2 COVID-19 vaccine dose.* Where mortality is not the outcome of interest, follow-up time will be additionally censored at death. Where mortality is the outcome of interest, follow-up time will be additionally censored at a death not meeting the outcome criteria.

5.7.4. Eligibility, Demographic, Clinical and Exposure Characteristics

The number and percentage of patients meeting the eligibility criteria will be reported in a subject evaluation table.

Descriptive statistics will be used to summarize patient demographics, clinical characteristics, and healthcare utilization on or prior to index date. Descriptive statistics will be repeated by BNT162b2 vaccination status by the end of follow-up. Standardized mean differences (SMD) will be used to assess covariate balance by BNT162b2 vaccination status by the end of each month of follow-up. Overall SMDs will be presented for each covariate as the maximum absolute value across the months. SMDs ≤0.1 will be considered as evidence of negligible imbalance.

Summaries of the overall follow-up time in months will be presented and calculated as time from index date to the earliest of 6 months of follow-up, end of medical and/or pharmacy enrollment, receipt of a non-Pfizer BioNTech COVID vaccine, receipt of a second BNT162b2 COVID-19 vaccine dose or death.

The number of patients who are BNT162b2 vaccinated by the end of the overall follow-up will be reported. Kaplan-Meier analysis will be used to summarize the percentage of persons who are BNT162b2 vaccinated by the end of each month of follow-up. Patients will be censored at the earliest of the following: 6 months of follow-up, end of medical and/or pharmacy enrollment, receipt of a non-Pfizer BioNTech COVID-19 vaccine or death.

The above analysis will be repeated by age group and state of residence.

5.7.5. Primary Analyses

Vaccinated and unvaccinated person-time (Section 5.7.3) for each outcome of interest will be calculated for each patient, summed across all patients and presented in person-months. The total number of patients with each of the outcomes will be presented by BNT162b2 vaccination status at the time of the COVID-19 encounter. Crude incidence rates and 95% confidence intervals (CIs) will be presented as events per 100,000 person-months (or as appropriate) by BNT162b2 vaccination status. The above analysis will be presented overall and may also be presented by month of follow-up for the primary and secondary outcomes.

Unadjusted Cox regression models will be used to generate hazard ratios (HR) and 95% CIs for the effect of the BNT162b2 for each outcome, with BNT162b2 modelled as a time-varying exposure. Crude VE percentages will then be estimated based on (1-HR) x 100, where HR is obtained from the unadjusted model.

Adjusted Cox regression models will also be used to generate adjusted HRs (aHR) and adjusted VE percentages, with covariates listed in Table 5.3 included as time-fixed covariates. For analyses where the number of events per variable is low (Section 5.7.2), covariates with SMDs ≤0.1 may be excluded from regression models.

For the primary outcome, for each of incidence rates and hazard ratios, an additional stratification of interest will employ an interaction term between exposure and a time-varying calendar time covariate, to allow for separate estimates of incidence and hazard in the predominantly XBB (11 September – 08 December 2023), co-circulation (09 December 2023 – January 5, 2024) and predominantly JN.1 (06 January 2024 – end of data collection [11 March 2024]) eras.

Here is a hypothetical set of rows in the analytic file for an example person who got vaccinated on October 1 and got COVID on February 1. Note: for both SAS and R the (START, STOP] intervals are open on the left and closed on the right. This implies that the STOP time is included in the interval but the start time is not.

exposure	outcome	tvc_xbb	tvc_cocirc	tvc_jn1	t_exposure_start	t_exposure_end	t_outcome_start	t_outcome_end
0	0	1	0	0	2023-09-10	2023-10-14	2023-09-10	2023-10-14
1	0	1	0	0	2023-10-14	2024-03-11	2023-10-14	2023-12-08
1	0	0	1	0	2023-10-14	2024-03-11	2023-12-08	2024-01-05
1	1	0	0	1	2023-10-14	2024-03-11	2024-01-06	2024-02-01

Sample SAS code for predominantly XBB era:

```
PROC PHREG data = [data file name] COVSANDWICH(aggregate);
CLASS [list of categorical variables, if any];
MODEL (t_outcome_start, t_outcome_end)*outcome(0) = exposure [list of covariates, don't include time-varying covariates here]
        / TIES = EFRON RL ;
WHERE tvc_xbb = 1;
RUN;
```

Sample SAS code for co-circulation era:

```
PROC PHREG data = [data file name] COVSANDWICH(aggregate);
CLASS [list of categorical variables, if any];
MODEL (t_outcome_start, t_outcome_end)*outcome(0) = exposure [list of covariates, don't include time-varying covariate here]
        / TIES = EFRON RL;
WHERE tvc_cocirc = 1;
RUN;
```

Sample SAS code for predominantely JN1 era:

```
PROC PHREG data = [data file name] COVSANDWICH(aggregate);
CLASS [list of categorical variables, if any];
MODEL (t_outcome_start, t_outcome_end)*outcome(0) = exposure [list of covariates, don't include time-varying covariate here]
        / TIES = EFRON RL;
WHERE tvc_jn1 = 1;RUN;
```

Measures of absolute and relative VE: For each outcome above, results will be calculated as:

1. *Absolute VE XBB.1.5 comparing vaccinated to XBB.1.5 unvaccinated persons, regardless of prior vaccination*

2. *Relative VE will be calculated for each of the following mutually exclusive subgroups.*

    a. *XBB.1.5 vaccinated vs XBB.1.5 unvaccinated persons who received BA.4/5 bivalent vaccine* (regardless of brand).

    b. *XBB.1.5 vaccinated vs XBB.1.5 unvaccinated persons who did not receive BA.4/5 bivalent vaccine but did receive at least 2 doses of mRNA vaccine (regardless of brand).*

    c. *XBB.1.5 vaccinated vs never COVID-19 vaccinated* (of any formulation or brand).

The proportional hazards assumption for the exposure will be assessed graphically with the inspection of Schoenfeld residuals. If the assumption is not met, the follow-up may be restricted up to the time where the assumption is met or VEs may be generated during intervals using piecewise Cox models.

We will further use incidence of the COVID-19 endpoint in unvaccinated persons and apply VE estimates to estimate the number of cases averted per 100,000 person-months. This estimate of the vaccine-preventable disease incidence will be generated for each endpoint as the adjusted VE as a proportion multiplied by the incidence rate among unvaccinated persons.

The above analysis will be repeated by age group for each outcome, subject to patient counts (Section 5.7.2). Additionally, the analysis for the primary outcome will be repeated by state of residence. Subject to the distribution of time since last prior COVID-19 vaccine, relative VEs may also be repeated by subgroups of that variable.

5.7.6. Assessment of Residual Confounding

Negative control outcomes will be used to quantify residual and unmeasured confounding (Table 5.2). The analysis described above will be conducted for the negative control outcomes. Persons with a diagnosis of the negative control outcome within 90 days prior to index date will be excluded from the analysis for that outcome.

In the event aHR for negative control outcomes have a significant effect (95% CI do not include null value of 1), then the set of confounders will be adjusted to allow for calibration. The COVID-19 related outcome VE results may be additionally reported after correcting for estimates of the bias.[7]

5.7.7. Sensitivity Analyses

Sensitivity analyses will consider the impact of study design and analytic assumptions, such as varying the exposure definition to 7 days post vaccination rather than 14 days and restricting the follow-up to 3 rather than 6 months.

6. REFERENCES

1. Moll, K., et al., Validating Claims-Based Algorithms Determining Pregnancy Outcomes and Gestational Age Using a Linked Claims-Electronic Medical Record Database. Drug Saf, 2021. **44**(11): p. 1151-1164.

2. People with Certain Medical Conditions. 2023 11 MAY 2023 [cited 2023 06 OCT]; Available from: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html.

3. Quan, H., et al., Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care, 2005. **43**(11): p. 1130-9.

4. Link-Gelles, R., et al., Effectiveness of Bivalent mRNA Vaccines in Preventing Symptomatic SARS-CoV-2 Infection - Increasing Community Access to Testing Program, United States, September-November 2022. MMWR Morb Mortal Wkly Rep, 2022. **71**(48): p. 1526-1530.

5. Austin, P.C., Generating survival times to simulate Cox proportional hazards models with time-varying covariates. Stat Med, 2012. **31**(29): p. 3946-58.

6. Vittinghoff, E. and C.E. McCulloch, Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol, 2007. **165**(6): p. 710-8.

7. Tchetgen Tchetgen, E.J.S., Tamar; Richardson, David, Negative Outcome Control for Unobserved Confounding Under a Cox Proportional Hazards Model. Harvard University Biostatistics Working Paper Series. Working Paper 192., July 2015.

7. APPENDICES

7.1. Appendix A: Table shells

The list of planned outputs and table shells will be presented in a separate document.

Document Approval Record

Document Name:	C4591065 Non Interventional Statistical Analysis Plan Version 1.0, 16 January 2024
Document Title:	C4591065 RAVEN Non Interventional Statistical Analysis Plan Version 1.0, 16 January 2024

Signed By:	Date(GMT)	Signing Capacity
PPD	27-Mar-2024 13:22:22	Final Approval
	28-Mar-2024 14:19:47	Manager Approval