U.S. FOOD & DRUG ADMINISTRATION

Center for Biologics Evaluation and Research
Office of Biostatistics and Epidemiology

CBER Surveillance Program

Background Rates of Adverse Events of Special Interest for COVID-19 Vaccine Safety Monitoring

Protocol

January 12, 2021

This protocol is submitted as an addendum to the COVID-19 Vaccine Safety Surveillance: Active Monitoring Master Protocol (https://www.bestinitiative.org/wp-content/uploads/2020/12/C19-Vaccine-Safety-Protocol-2020.pdf).

Table of Contents

STUDY TEAM

FDA/CBER/OBE
Cindy Ke Zhou, PhD Hui-Lee Wong, PhD Deborah Thompson, MD, MSPH Rositsa Dimova, PhD Tainya Clarke, PhD Richard Forshee, PhD Azadeh Shoaibi, PhD Steven A. Anderson, PhD
IBM
Katie Fingar, PhD Keran Moll, PhD Shayan Hobbi, MS, MBA Frank Yoon, PhD Patrick Saunders-Hastings, PhD Tim Burrell, MD Jeff Beers, MD Kathryn Edwards, MD Steven Black, MD
Acumen, LLC & Stanford University
Michael Wernecke, BS Yoganand Chillarige, MPA Jiemin Liao, MS Mao Hu, BS Bradley Lufkin, MPA, MSES Ellen Tworkoski, MS, MPhil Laurie Feinberg, MD, MPH, MS Nirmal Choradia, MD Lorene Nelson, PhD (Stanford) Julia Simard, ScD (Stanford) Steven Goodman, MD, MHS, PhD (Stanford) Thomas MaCurdy, PhD
Centers for Medicare & Medicaid Services
Jeffrey Kelman, MD, MMS Marybeth Jason, MS
Optum
Kandace Amend, PhD, MPH Grace Yang, MPA, MA John Seeger, DrPH, PharmD Rachel P. Ogilvie, PhD, MPH Kathleen Skerry, BS

Version Control

Editors	Date	Version	Description
FDA, IBM, Acumen LLC	November 14, 2020	0.1	First draft
FDA, IBM, Acumen LLC	December 11, 2020	1.0	Second draft
FDA, IBM, Acumen LLC	December 17, 2020	1.1	Third draft
FDA, IBM, Acumen LLC	December 22, 2020	1.2	Fourth draft
FDA, Acumen LLC	December 31, 2020	1.3	Fifth draft
FDA, IBM	January 12, 2020	1.0	Final

1. Overview

Governments and private companies accelerated COVID-19 vaccine development in response to the COVID-19 pandemic that emerged in late 2019. The first human clinical trial for a COVID-19 vaccine began in March 2020. As of November 2020, there were 47 candidate vaccines in clinical evaluation across the globe, many of which use new technology platforms, such as messenger ribonucleic acid (mRNA) (WHO, 2020).[1]

Postmarket surveillance of adverse events is needed after vaccines are launched and use becomes widespread. Active safety monitoring methods often use background rates as a comparator for rapidly identifying potential increased risk of adverse events of special interest (AESIs) following vaccination. Thus, it is important to generate background rates of AESIs, overall and by subgroup, for populations that will potentially receive a COVID-19 vaccine.

2. Objectives

This protocol specifies the analytic approach to estimate background rates of AESIs in preparation for future sequential active monitoring for the safety of COVID-19 vaccines.

2.1 Primary Objectives

- To estimate incidence rates of AESIs by calendar year and data source over the period 2017 to 2020
- To estimate cumulative monthly incidence rates[a] of AESIs by data source over the period 2017 to 2020
- To estimate cumulative weekly and cumulative monthly incidence rates of COVID-19 by data source in 2020

2.2 Secondary Objectives

- To estimate yearly incidence rates of AESIs in the general population stratified by sex, age group, and race/ethnicity group (where reliably available) in each data source over the period of 2017 to 2020
- To estimate cumulative monthly incidence rates of AESIs in the general population stratified by sex, age group, and race/ethnicity group (where reliably available) in each data source over the period of 2017 to 2020
- To estimate cumulative monthly incidence rates of AESIs in the general population stratified by the time period of January 1, 2017 to February 29, 2020 (pre-COVID-19 period in the United States) and the time period of March 1, 2020 to December 11, 2020 (COVID- 19/pre-COVID-19 vaccine period) in each data source
- To estimate the cumulative weekly and cumulative monthly proportion of COVID-19 cases hospitalized around the incident COVID-19 diagnosis, by data source in 2020
- To estimate the cumulative weekly and cumulative monthly proportion of laboratory-confirmed COVID-19 incident cases using reverse transcription polymerase chain

[a] *Cumulative monthly incidence rate* is defined as the incident cases that occurred from January 1 through the month of evaluation divided by the total person-time at risk accumulated during the same period.

reaction (RT-PCR) laboratory results, by data source (for data sources with lab results available) in 2020

- To estimate incidence rates of AESIs in special populations of interest stratified by calendar year, sex, age group, and race/ethnicity (where reliably available) in each data source over the period 2017–2020. These populations will include:
    - Older adults (i.e., 65 years old and above at cohort entry)
    - Pediatric population (i.e., 0–17 years old at cohort entry)
    - Pregnant women
    - Individuals who received a seasonal influenza vaccine in the previous calendar year

3. Study Design

This is a retrospective, multi-database dynamic cohort study using secondary administrative claims data and electronic health record (EHR) data from 2017 (where available) through the last date of data reaching 80% complete (e.g., a 4-month data lag for the Blue Health Intelligence® [BHI®][b] and IBM® MarketScan® Research Databases, Table 1) up until release of the first COVID-19 vaccine. Data sources and data lags are described further below.

4. Data Sources

The current study will use administrative claims and electronic health records (EHRs). Table 1 briefly outlines currently available data sources and displays how often each data source is updated. Additional databases may be added when they are available to the Biologics Effectiveness and Safety (BEST) Initiative.

Table 1. Update Frequency for Each Data Source

Data Source		Update Frequency	Data Lag*	Population
Claim data sources	Blue Health Intelligence (BHI)	Monthly	4 months for >80% coverage of inpatient claims	National; >17 million beneficiaries annually who received a biologic product, were pregnant, or were born after October 1, 2015
	Centers for Medicare & Medicaid Services Medicare	Daily or weekly	30–70 days for >80% coverage of inpatient claims	National; 65+ years; >34 million beneficiaries annually
	IBM MarketScan Commercial Database	Monthly	4 months for >80% coverage of inpatient professional commercial claims	>25 million commercial enrollees annually

[b] Blue Health Intelligence® (BHI®) is a trade name of Health Intelligence Company, LLC, an independent licensee of the Blue Cross and Blue Shield Association.

Data Source		Update Frequency	Data Lag*	Population
	OptumServe Clinformatics Data Mart	Monthly	1.5 months for pharmacy claims, 3 months for outpatient claims, and 6 months for inpatient claims	National; 66 million patients and ~14 million patients annually
EHR data sources	MedStar	Real time	1–7 days for data processing	District of Columbia, Maryland, and Virginia; 1.5–2 million active patients annually
	OneFlorida	Quarterly for centralized PCORnet/ OMOP database	Up to 4 months for data processing and linkage across OneFlorida network	Florida; 2–3 million patients annually
	PEDSnet	Quarterly for centralized PCORnet/ OMOP database	Up to 4 months for data processing and linkage across PEDSnet network	Coverage for 12 states, with 1–2 million patients annually
	OptumServe	Quarterly	3 months prior to each refresh	National; 102 million patients, 30 million annually

* Data lag refers to the amount of time between the date of service and the date of availability for use by research teams.

4.1 Administrative Claims Data Sources

Available sources include adjudicated claims in BHI commercial claims data and the IBM MarketScan Commercial Database, adjudicated and pre-adjudicated[c] OptumServe commercial claims, as well as adjudicated and unadjudicated claims from Centers for Medicare & Medicaid Services (CMS) Medicare data.

BHI data provide HIPAA compliant, deidentified enrollment, demographic, and claims information from the majority of Blue Cross and Blue Shield commercial health insurance plans in the United States for the last ten years. While the BHI data set includes more than 200 million unique lives, detailed data for this study was limited to a cohort of all enrollees who received a biologic product, were pregnant, or were born after October 1, 2015. Pregnant women are identified via codes for prenatal care, gestational age, or pregnancy outcomes. Currently, the BHI cohort population for this study contains over 34 million individuals in total and about 17 million enrollees annually, on average. Approximately 350,000 pregnancy outcomes are observed annually. BHI data are updated monthly and are over 80% complete within 4 months of the service date.

The MarketScan Commercial Database contains more than 25 million annual enrollees for active employees, early retirees, Consolidated Omnibus Budget Reconciliation Act (COBRA)

[c] After a claim for medical service(s) or prescription drug(s) is submitted, the insurer determines their financial responsibility for the payment to the provider or the pharmacy. This process is referred to as claims adjudication. The insurer can decide to pay the claim in full, deny the claim, or to reduce the amount paid to the provider or the pharmacy. Due to the time requirement for the adjudication process, a database of adjudicated claims will have a longer lag time than the pre-adjudicated claims.

continuees, and dependents insured by employer-sponsored plans. The MarketScan Research Databases capture individual-level clinical utilization, expenditures and enrollment across inpatient, outpatient, prescription drug, and carve-out services from a selection of large employers, health plans, and government and public organizations.

The OptumServe Clinformatics Data Mart contains longitudinal health information for commercially insured and Medicare Advantage enrollees, and it includes more than 66 million lives since 2007. The commercial portion of the Clinformatics Data Mart includes approximately 14.5 million people on an annual basis and the median dwell time in the database for enrollees is approximately 2.5 years. The OptumServe Clinformatics Data Mart contains claims for physician, hospital, and prescription drug services. The OptumServe Clinformatics Data Mart is updated monthly and is 90% complete within 1.5 months for pharmacy claims, 3 months for outpatient claims, and 6 months for inpatient claims. The OptumServe pre-adjudicated claims database contains claims that are processed daily. The pre-adjudicated claims database includes data from January 2018.

CMS Medicare fee-for-service data contain enrollment, demographic, and claims information for all individuals enrolled in Parts A/B (since 1991), Part C (since 2012), and Part D (since 2006). Medicare data currently contains about 34 million current beneficiaries annually, on average. Medicare claims data undergo three stages of processing: enumeration, adjudication, and final payment. Medicare Shared Systems Data (SSD), which consists of claims sourced after enumeration, will be used in this study. SSD is updated daily and is over 80% complete within 30-70 days depending on setting and outcome.

4.2 Electronic Health Record Data Sources

Available EHR sources include direct EHR data from the MedStar Health network, the OneFlorida Clinical Research Consortium, the PEDSnet consortium, and OptumServe EHR.

MedStar Health is a non-profit, academic health system that serves patients in the District of Columbia, Maryland, and Virginia. It provides care through telemedicine eVisit services, physician offices, urgent care centers, regional ambulatory care centers, 10 hospitals, and the region’s largest home health agency. The MedStar Health network provides EHR data via a Cerner Corporation platform on approximately 6 million patients from 2010 to present, and about 1.5-2 million active patients annually.

The OneFlorida Clinical Research Consortium Data Trust contains claims and encounter data for Floridians enrolled in Medicaid and EHR data from public and private health care systems, including diagnoses, procedures, medications, patient demographics, unique patient codes for re-identification by consortium partners and other data elements in both the PCORnet and the Observational Medical Outcomes Partnership (OMOP) Common Data Models (CDM). Providers include large integrated healthcare delivery networks, 13 large hospitals, and ambulatory care and primary care facilities. In total, the Data Trust contains EHR data for about 6 million Floridians or about 75% of Floridians from 2012 to present, with about 2-3 million patients annually on average.

PEDSnet is a national network of eight pediatric hospitals and healthcare organizations spanning across 12 states. The consortium's data contains patient-level EHR data from public and private health care systems, including diagnoses, procedures, medications, patient demographics, unique patient codes for re-identification by consortium partners and other data elements in both the PCORnet and OMOP CDMs. In total, the consortium covers over 6 million patients with about 1-2 million patients annually on average.

The OptumServe EHR repository is derived from dozens of large healthcare provider organizations across the United States treating patients with a variety of health insurance coverage. The OptumServe EHR data contains patient demographics, diagnoses, procedures, medications prescribed and administered, immunizations, laboratory results, microbiology results, vital signs, clinical and inpatient stay administrative data and unique patient codes for re-identification by consortium partners and other data elements in the OMOP CDM. In total, the OptumServe EHR covers more than 102 million patients with approximately 30 million patients having an encounter annually.

4.3 Linked Claims-EHR Data Sources

The linked claims-EHRs include approximately 1 million patients annually from the Florida Medicaid claims linked to the respective EHR data in the OneFlorida EHR system.

The OptumServe linked claims-EHR includes a total of approximately 25 million patients from 2007 to 2018, with approximately 2 million active annual enrollees. Optum's integrated claims-EHR database combines the adjudicated claims data with electronic health record data.

5. Outcomes

The list of AESIs has been developed based on a comprehensive literature review, consultation with subject matter experts, and examination of lists developed by organizations with a mandate to consider or assess COVID-19 vaccine safety, particularly the vACCine Covid-19 monitoring readinESS (ACCESS) and Safety Platform for Emergency vACcines (SPEAC). These AESIs have not been associated with COVID-19 vaccines based on available pre-licensure evidence.

5.1 Adverse Events of Special Interest

A list of prespecified COVID-19 vaccine AESIs is presented in Table 2. An independent assessment of historical background rates will be conducted for each AESI.

An International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM)-based definition (or *algorithm*) was developed for each AESI listed. Algorithms were developed based on comprehensive literature reviews to identify ICD-10-CM algorithms that had been subject to independent validation. Because validated ICD-10-CM-based algorithms were not available for most AESIs, a validated ICD-9-CM algorithm was converted to ICD-10-CM via forward-backward mapping, using the General Equivalence Mappings for reference[d]. These draft algorithms were then subject to review and refinement by clinical subject matter experts. In

[d] Additional information about the General Equivalence Mappings (GEMs) and the methodology for forward and backward mapping can be found at Centers for Medicare & Medicaid Services. https://www.cms.gov/Medicare/Coding/ICD10/2018-ICD-10-CM-and-GEMs.

each case, a more inclusive, general algorithm was developed first and then tailored for a COVID-19 vaccine-specific application via further consultation with clinicians.

For assessment of AESIs in the clean window (i.e., an interval used to define incident outcomes where an individual enters the study cohort only if an AESI did not occur during that interval), certain EHR databases record entries in the Problem List[e] in the Systematized Nomenclature of Medicine (SNOMED). For these databases, the AESIs in the clean window will be assessed based on SNOMED codes mapped from the ICD-10-CM-based algorithms using the OMOP vocabulary tables, as well as diagnoses recorded during the clean window. For these EHR databases, absence of a diagnosis or an entry in the problem list for an AESI would satisfy the clean window requirements.

Claims from inpatient facilities (IP), outpatient facilities in the emergency department (OP-ED), and all outpatient facilities and individual providers or professionals (OP/PB) will be used to capture AESIs. A more detailed description of these settings is provided in Appendix, Table A1. Table 2 specifies the care settings used for identification of diagnosis (diagnosis codes may be on any position on a claim, unless otherwise specified), age group of interest when counting AESIs, and length of clean window to define incident AESIs. This list is subject to change or expansion over time as new knowledge becomes available. A similar approach will be taken for any new AESI algorithms to be developed, with definitions based on comprehensive literature reviews and clinical subject matter expert consultation.

Table 2. List of Adverse Events of Special Interest (AESIs)

Safety AESI	Age Group of Interest	Setting	Clean Window**
General Population Outcomes			
Guillain-Barré syndrome	All	IP- primary position only	365 days*
Bell's palsy	All	IP, OP/PB	183 days*
Anaphylaxis	All	IP, OP-ED	30 days
Encephalomyelitis/encephalitis	All	IP	183 days*
Narcolepsy	All	IP, OP/PB	365 days*
Appendicitis	All	IP, OP-ED	365 days*
Non-hemorrhagic stroke	All	IP	365 days*
Hemorrhagic stroke	All	IP	365 days*
Acute myocardial infarction	All	IP	365 days*
Myocarditis/pericarditis	All	IP, OP/PB	365 days*
Deep vein thrombosis	All	IP, OP/PB	365 days*
Pulmonary embolism[#]	All	IP, OP/PB	365 days*
Disseminated intravascular coagulation	All	IP, OP-ED	365 days*
Immune thrombocytopenia	All	IP, OP/PB	365 days*
Transverse myelitis	All	IP, OP-ED	365 days*

[e] Problem lists facilitate continuity of patient care by providing a comprehensive and accessible list of patient problems in one place. Problem lists used within health records are a list of illnesses, injuries, and other factors that affect the health of an individual patient, usually identifying the time of occurrence or identification and resolution.

Safety AESI	Age Group of Interest	Setting	Clean Window**
Multisystem inflammatory syndrome	18 years and older	IP, OP-ED	365 days*
Pediatric Outcomes			
Multisystem inflammatory syndrome in children (MIS-C)	0–17 years	IP, OP-ED	365 days*
Febrile seizures	6–60 months[2]	IP, OP/PB	42 days
Kawasaki disease	1–5 years	IP, OP/PB	365 days
Pregnancy and Birth Outcomes			
Spontaneous abortion[3]	Women, 12–55 years	IP, OP/PB	Pregnancy spacing requirements are specified by the algorithm.
Preterm birth[3]	Women, 12–55 years	IP, OP/PB	
Stillbirth[3]	Women, 12–55 years	IP, OP/PB	
Full-term birth[3]	Women, 12–55 years	IP, OP/PB	

Abbreviations: ED, emergency department; AESI, adverse event of special interest; IP, inpatient; N/A, not applicable; Setting Definitions: IP refers to inpatient facility claims. OP-ED refers to a subset of outpatient facility claims occurring in the emergency department. OP/PB refers to all outpatient facility claims, and professional/provider claims except those professional/provider claims with a laboratory place of service. References for these windows could not be found in the literature and are instead based on input from clinicians.

** Clean window is defined as the time period prior to cohort entry, during which no AESI was observed.

[#] If an individual has both deep vein thrombosis (DVT) and pulmonary embolism (PE) (i.e., the DVT progressed to PE), the case will be deduplicated in the analysis stage and assigned only PE. The PE onset date is determined by the date the PE code is reported in the database.

5.2 COVID-19 Diagnosis

COVID-19 disease cases will be identified by capturing the first COVID-19 diagnosis codes for a patient. Additionally, COVID-19 diagnoses confirmed by reverse transcription polymerase chain reaction (RT-PCR) testing will be identified among patients who had COVID-19 RT-PCR results within +/-7 days of the diagnosis in claims data sources with lab result data (e.g., MarketScan commercial data) or in the EHR data. COVID-19 lab codes will be identified using Logical Observation Identifiers Names and Codes, or LOINC. Cumulative incidence of COVID-19 diagnoses will be assessed in study data sources on a monthly and weekly basis starting in March of 2020 to the end of the study period (Section 6) for data that are at least 80% complete. Cumulative weekly and cumulative monthly proportions of COVID-19 cases with a hospitalization between -7 and +14 days before/after the incident diagnosis will also be assessed by data source in 2020.

5.3 Negative Control Events

This study will also assess negative control events (Table 3), posited to be unrelated to potential AESIs of COVID-19 vaccination but for which healthcare utilization may or may not change as a result of lockdowns or care-seeking behavior during the pandemic. Urgent negative control events will include colonic diverticulitis. Because this condition requires immediate treatment, healthcare utilization for this outcome is expected to remain stable during the pre-COVID-19 and COVID-19 periods. Non-urgent conditions will include hypertension; screenings for well-child

visits; and those for breast, cervical, and colon cancer. It is hypothesized that healthcare utilization for these conditions and visits may decrease during the COVID-19 period because of delayed healthcare-seeking behavior.

Table 3. Negative Control Events

Negative Control Event	Age Group of Interest	Setting	Continuous Enrollment or Clean Window
Colonic diverticulitis	18 years and older	IP, OP/PB	365 days*(clean window)
Hypertension	18 years and older	IP, OP/PB	
Well-child and well-care visits	All ages	IP, OP/PB	365 days**(continuous enrollment)
Colonoscopies for colorectal cancer screening	45 years and older	IP, OP/PB	
Mammograms for breast cancer screening	Women, 40 years and older	IP, OP/PB	
Cervical cancer screenings	Women, 21 years and older	IP, OP/PB	

Abbreviations: ED, emergency department; AESI, adverse event of special interest; IP, inpatient; N/A, not applicable; Setting Definitions: IP refers to inpatient facility claims. OP/PB refers to all outpatient facility claims, and professional/provider claims except those professional/provider claims with a laboratory place of service.

* References for these windows could not be found in the literature and are instead based on input from clinicians. The clean window is meant to increase comparability of these negative control events to the safety AESIs.

** A clean window is not implemented for preventive care visits or screenings to not exclude patients who sought preventive care in the previous year.

6. Study Period

The study period will start on January 1, 2017, and end on December 11, 2020, when the FDA issued an emergency use authorization (EUA) for the Pfizer-BioNTech COVID-19 vaccine, the first COVID-19 vaccine available in the US. The study period will be divided into two parts because COVID-19 pandemic has affected healthcare utilization; Reductions in the use of preventive care, elective care and the number of visits to physician offices and hospitals have been associated with the introduction of shelter in place policies.[4]

The pre-COVID-19 period extends from January 1, 2017, through February 29, 2020. The COVID-19/pre-vaccine period begins on March 1, 2020 and ends on December 11, 2020. March 1, 2020 was chosen as the start date of this analysis based on preliminary assessments in the data sources showing COVID-19 diagnoses and RT-PCR testing accrual starting on March 1, 2020.

Assessment of pregnancy and clean periods will use data from 2016, although follow-up for the study cohort will start in 2017.

7. Study Population in Claims Data Sources

7.1 General Population of Interest

Annual cohorts will be identified for calendar years 2017, 2018, 2019, and 2020. In each annual cohort, the study population includes any individual enrolled with medical benefits in one of the data sources for at least 1 day during the calendar year and who met the clean period requirement before cohort entry, as specified below. AESIs will be identified in medical claims. Requiring medical benefits without also requiring pharmacy coverage improves the inclusion of special populations of interest. It is expected that most enrollees with medical benefits also have pharmacy benefits. Hereafter, *enrollment* refers to medical benefit coverage, and *continuous enrollment* allows a coverage gap of up to 31 days.

Cohort entry: For each AESI in each calendar year, and those who were not in the database at birth, an individual enters the AESI-year-specific cohort the date they meet the clean period requirement and is still enrolled on the entry date. Cohort entry date is no earlier than January 1 of the given calendar year. The *clean period requirement* is defined as having (1) continuous enrollment for the length of the prespecified clean window (i.e., a baseline period specific to each AESI to establish an incident AESI [see Table 2]) prior to the cohort entry date and (2) no diagnosis from the care setting(s) (as specified in Table 2) for the AESI during the clean period. Individuals may be included in more than one AESI-specific cohort if they meet all the inclusion criteria.

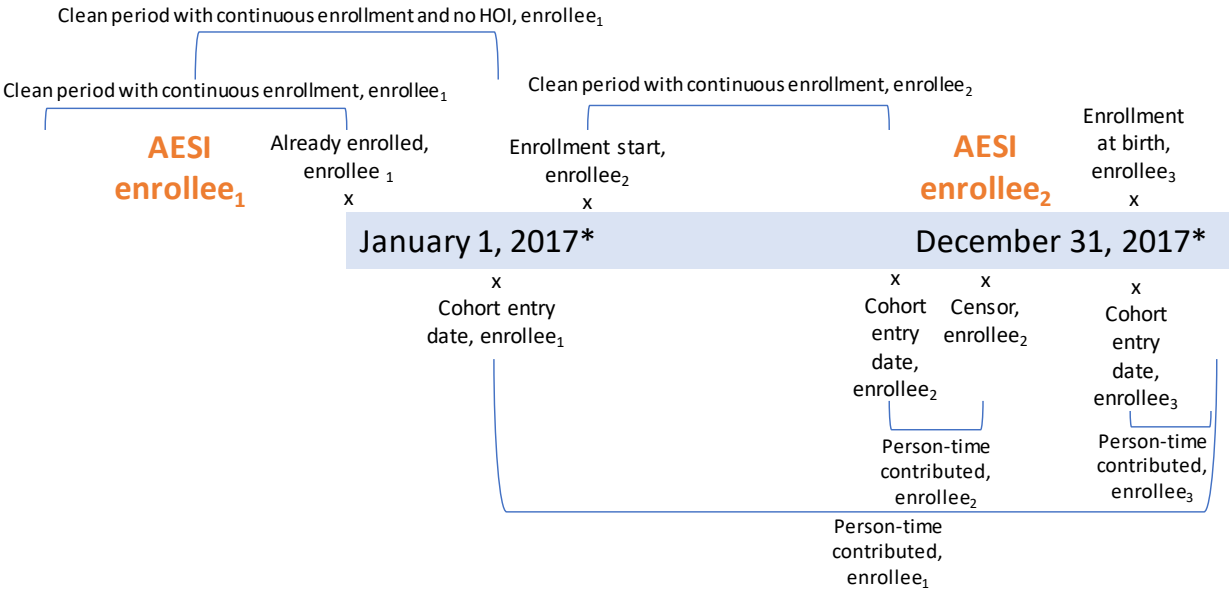
If a birth occurred before January 1 and the infant is continuously enrolled from birth but has not reached the full length of the clean period on January 1, the entry date for the infant is January 1, unless an AESI has occurred during the shortened clean period prior to entry. For those who were born in the year and started enrollment within 31 days of birth, the cohort entry date is the date of birth (this is because no clean period is needed for newborns; AESIs, if any, are all incident cases by definition). For databases with only birth year or partial birth date, infants can enter the cohort at the later date of the start of the enrollment or the start of the analytic calendar year, unless an AESI has occurred prior to the start of the analytic calendar year following the start of the enrollment.

Cohort re-entry: If an individual is censored, the individual may re-enter the same annual AESI cohort if another clean period requirement is met during the same calendar year. If an individual is censored at an occurrence of an AESI and remains enrolled, the individual may re-enter the cohort after another clean period elapses without an AESI; if an individual is censored due to disenrollment, a clean period from the re-enrollment date must be met before re-entry. Individuals who re-enter the cohort are treated as separate observations, and multiple occurrences of AESIs among re-entered individuals are assumed to be incident cases.

Follow-up and censoring: Individuals in each annual cohort will be followed from each cohort entry date until the earliest date of AESI occurrence (first date of service with the AESI diagnosis in the specified care setting), death (identified by date of death if available in a data source, or the discharge date for death that occurred in a hospital or hospice based on the discharge status), disenrollment, the end of the calendar year, or the date of the latest data that reached approximately 80% completion (applicable to the later months of 2020). Additional inclusion/exclusion criteria for the calculation of incidence rates in claims data are provided in Section 8.

Figure 1 presents three examples to illustrate rules for cohort entry/re-entry, follow-up and censoring.

Figure 1. General Population Cohort and Follow-up



This figure illustrates three examples of cohort entry and the accumulation of person-time at risk:

Enrollee$_1$ was enrolled as of January 1, 2017. However, the enrollee had an AESI in 2016, which does not allow a full clean period to incur by January 1, 2017, therefore, this enrollee's cohort entry date is delayed into 2017 until a full clean period with no AESI is met. From the cohort entry date, enrollee 1 remained enrolled and had no censoring events in 2017. This enrollee's person-time at risk accumulated until the end of year 2017.

Enrollee$_2$ started enrollment during 2017 and entered the 2017 annual cohort after a clean period with no AESI is met. The enrollee subsequently had an AESI and censored at the date of the AESI. The enrollee's person-time at risk accumulated from entry date to the AESI date. This enrollee's AESI will contribute to the numerator for the calculation of the incidence rate of this AESI.

Enrollee$_3$ was born during 2017 and enrolled at birth. This enrollee enters the 2017 annual cohort at birth and remained enrolled without no censoring events until the end of 2017. This enrollee's person-time at risk accumulates from date of birth to the end of 2017.

*Annual cohorts will be defined for 2017, 2018, 2019, and 2020.

7.2 Special Populations of Interest

Special populations of interest include the older adult group (i.e., 65+ years at cohort entry), pediatric populations (i.e., <18 years at cohort entry), pregnant women, and individuals who received an influenza vaccination in the prior calendar year. The two latter populations are described in detail in the following subsections.

Patients in Medicare FFS will only be included if they are age 65 or above at the time of vaccination. Private insurance claims databases will be restricted to patients under age 65 to reduce the chance that individuals are double-counted across databases.

For pediatric AESIs that are limited to a population of interest within an age range (e.g., Kawasaki disease among children aged 1-5 years old), an individual must be within the age range on the cohort entry date. Therefore, the cohort entry data is January 1, or the date meeting clean period

requirement, or the date the individual reached the specified age range for an AESI, whichever happens later, in a given calendar year. During follow-up, in addition to other censoring events as specified in Section 7.1, an individual will also be censored when the individual reaches beyond the specified age range for an AESI during a calendar year (e.g., an individual will be censored from the Kawasaki cohort when reaching 6 years old).

7.2.1 Pregnant Women

COVID-19 vaccines are not recommended, nor are they contraindicated for pregnant women. It is a decision for women in consultation with her healthcare provider. Incidentally or intentionally pregnant women may be exposed to COVID-19 vaccine(s). Background rates of pregnant outcomes and AESIs among pregnant women will help contextualize potential safety signals.

The CBER BEST Initiative has developed and validated a set of algorithms to identify gestational age and pregnancy outcomes with good to excellent performance among women aged 12–55 years at the time of the outcome.[5] The algorithms identify pregnancies by capturing pregnancy endpoints (full-term live births, preterm live births, spontaneous abortions, and stillbirths) using ICD-10-CM/Procedure Coding System (PCS) and other service codes (Healthcare Common Procedure Coding System codes [HCPCS] and diagnosis-related group [DRG] codes) in women's claims. They then estimate a pregnancy start date and gestational age at the outcome using relevant ICD-10-CM/PCS and HCPCS codes on prenatal and delivery claims. The algorithms group obstetric services into pregnancy episodes based on hierarchical and spacing requirements. Evidence of high clinical accuracy for gestational age (such as the timing of assisted reproductive technology procedures and gestational age determined during first trimester ultrasound) are prioritized when determining the pregnancy start date. The pregnancy outcomes and estimate gestational age will be identified using ICD-10-CM/PCS, HCPCS, and DRGs codes. The time period between the pregnancy start date and outcome date defines a pregnancy episode. A woman may have more than one pregnancy episode during the study period or within each calendar year.

For claims data sources, pregnancy episodes will be constructed based on the algorithm described above using data from 2016 to the end of the study period. Data from 2016 are needed to identify all pregnancies that lasted into 2017.

Cohort entry for the pregnant subpopulation: The cohorts will be defined differently, depending on whether the outcome of analysis is a pregnancy outcome (preterm live birth, full-term live birth, stillbirth, or spontaneous abortion) or the other non-pregnancy-related safety AESIs specified in Table 2.

For pregnancy outcomes, for women who were pregnant on January 1 of a given calendar year, the entry date is January 1. For pregnancy episodes that started during the year, cohort entry date is the start of the pregnancy.

For other safety AESIs, the entry date is the pregnancy start date or when women meet the clean period requirement during pregnancy, whichever is later.

Cohort re-entry: A woman may contribute person-time from multiple pregnancies and enter the pregnancy cohort multiple times during the same calendar year. Re-entry is defined similarly as in Section 7.1.

Follow-up and censoring among the pregnant subpopulation: For pregnancy outcomes,

pregnant women will be followed from the cohort entry date until the end of the pregnancy or end of the calendar year, whichever comes first. For other safety AESIs, they will be followed from cohort entry date until the date of the AESI, the end of the pregnancy, or the end of the calendar year, whichever occurs first. Women must remain continuously enrolled during the entire pregnancy episode in order to apply the pregnancy algorithms and be included in the analysis, therefore, censoring by death or disenrollment is not needed among the pregnant subpopulation.

7.2.2 Individuals with Recent Influenza Vaccination

This study will also evaluate the background rates of AESIs among a subpopulation with a recent history of a seasonal influenza vaccine. The recent history of flu vaccination may serve as a proxy for health-seeking behavior, vaccination acceptance, and comorbidities. We will evaluate comparability between this population and the population receiving COVID-19 vaccines in the active monitoring protocol. In each annual cohort, we define the subpopulation with a recent history of a flu vaccine as those who had at least one flu vaccine observed during the prior calendar year. An individual is not required to have a certain length of continuous enrollment during the prior calendar year to be selected into the subpopulation, however only for those individuals that have at least one day of enrollment will an influenza vaccine be observed. The cohort entry date, follow-up time, and censoring are defined in the same manner as in the general population (Section 7.1).

7.3 Populations to Evaluate Negative Control Events

Annual cohorts for the evaluation of negative control events will include all individuals who are enrolled for at least 1 day in the calendar year.

The cohort entry requirement is different for negative control events of hypertension and colonic diverticulitis and for those that are preventive care services (well-child and well-care visit, colonoscopy for colorectal cancer screening, mammogram for breast cancer screening, or cervical cancer screening).

For hypertension and colonic diverticulitis, the cohort entry date is the later date of January 1 of a given calendar year or the date an individual meets the clean period requirement, or the date an individual reaching 18 years old.

For preventive care services, the cohort entry date is the later date of January 1 of a given calendar year, the date an individual has 1 year of continuous enrollment, or the date the individual reaching the specified age range (e.g., 45 years old and older for screening colonoscopy). The absence of an event in the period prior to the cohort entry is not required. This is an effort to not exclude patients who sought preventive care in the previous year. For preventive services, some of which of which are recommended annually, an incidence case—defined as the first occurrence—is less meaningful in tracking the healthcare utilization pattern changes.

The re-entry, follow-up time, and censoring are defined in the same manner as in the general population (Section 7.1).

7.4 Population to Evaluate COVID-19 Disease Incidence

A population for the evaluation of COVID-19 disease incidence will include all individuals who are enrolled for at least 1 day for the period of March 2020 to the date with at least 80% complete data availability at the time of analysis or the end of the study period, whichever is earlier. March 1, 2020 was chosen as the start date of this analysis based on preliminary assessments in the data sources showing COVID-19 diagnoses and RT-PCR testing accruing starting on March 1, 2020.

The cohort entry date is March 1, 2020, or start of enrollment or birth during 2020, whichever happens later. No clean period or continuous enrollment prior to entry will be required for this population.

Cohort re-entry is not permitted because it would be difficult to distinguish between ongoing COVID-19 care after the initial diagnosis and reinfection with COVID-19. In addition, there only have been rare reports on reinfections in the same individual for COVID-19 thus far.

Follow-up and censoring: Individuals will be followed from entry date until the date of the first COVID-19 diagnosis code, death, disenrollment, date of latest data with at least 80% completeness, or end of the year or study period, whichever happens first.

8. Analysis in Claims Data Sources

8.1 Background Rate Estimation in Claims Data Sources

8.1.1 General Population

Person-time at risk is defined as the person-time of an individual from the cohort entry (or re-entry) date to the date of censoring, as specified in Section 7. Person-time at risk is defined the same way in subpopulations (Section 8.1.2), populations to evaluate negative control events (Section 8.1.3) and populations to estimate COVID-19 incidence (Section 8.1.4).

Denominator: For each AESI, in each annual cohort, person-time at risk will accumulate from the cohort entry date until censoring in the calendar year. Censoring will occur at the AESI event, death, disenrollment, the end of the calendar year, or the latest date with data of approximately 80% completeness at the time of analysis, whichever comes first.

For the annual rates, all person-time at risk in a given calendar year will be summed to form the denominator.

For cumulative monthly rates, all person-time at risk during the calendar year that overlaps the time period from January 1 through the month of evaluation will be summed to form the denominator. For example, the denominator for the cumulative monthly rate through June is the total person-time at risk from January 1 to June 30. Person-time will first be calculated in person-days and then scaled to person-years or person-months for annual and cumulative monthly rates, respectively.

Numerator: The number of incident cases among individuals contributing person-time to the denominator will be enumerated for a calendar year to form the numerator for the annual rates. The numerator of incident cases from January 1 of a given calendar year through the end of the

month of evaluation among individuals contributing person-time in the same time period will be summarized to form the numerator of the cumulative monthly incidence rates.

Rate calculation: The numerators will be divided by the corresponding aggregated person-time denominators by calendar year or cumulative month and expressed as an incidence rate per 100,000 person-years or per 100,000 person-months. The 95% confidence interval of the annual and cumulative monthly rates will be estimated using the exact Poisson confidence limit method.[6]

8.1.2 Special Populations of Interest

8.1.2.1 Pediatric Populations and Older Adults

The population is classified into the pediatric and older adult subpopulations based on age at the cohort entry (or re-entry) date within each annual cohort. The denominator and numerator for rates estimation will be limited to children younger than 18 years old and adults aged 65 years and older, respectively. Otherwise, the rate calculation is consistent with the methods described above for the general population in Section 8.1.1.

For pediatric AESIs that are limited to a specific age range of interest (as described in Section 7.2), the denominator and the numerator will be limited to the person-time at risk within the age range and incident cases occurring within the age range, respectively.

8.1.2.2 Pregnant Population

Denominator: For pregnancy outcomes, person-time at risk during pregnancy is accumulated starting from cohort entry or re-entry date until date of pregnancy outcome or end of the calendar year, whichever happens first.

For annual rates, person-time during pregnancy in the calendar year will be summed to form the denominator.

For cumulative monthly rates, the denominator will include person-time during pregnancy starting on the cohort entry date for the pregnancy through the end of pregnancy or the end of the month of evaluation, whichever happens first.

For assessment of the AESIs that are not pregnancy related but that occur during pregnancy (nonpregnancy AESIs), person-time at risk accumulation will start from the pregnancy cohort entry date to the date of nonpregnancy AESI, the end of pregnancy, or the end of the calendar year, whichever happens first. As mentioned in prior sections, a woman may enter the cohort multiple times if she has multiple pregnancies. For the annual rates, all person-time at risk during pregnancy during a given calendar year will be summed to form the denominator. For cumulative monthly rates, the denominator will include person-time at risk during pregnancy from cohort entry date through the month of evaluation.

Numerator: The number of pregnancy outcomes (preterm live births <37 weeks of gestation, full-term live births 37+ weeks of gestation, spontaneous abortions <20 weeks of gestation, stillbirths 20+ weeks of gestation) and the number of AESIs occurring during pregnancy will be counted annually and cumulatively by month.

Rate calculation: The numerator will be divided by the corresponding person-time denominator calculated above. Rates will be expressed as the number of incident cases per 100,000 gestational weeks. The 95% confidence interval of the rates will be estimated as specified in Section 8.1.1.

8.1.2.3 Individuals with Recent Influenza Vaccination

Denominator: For each AESI in each calendar year, a subset of the overall population who had an influenza vaccination in the prior calendar year will be sampled. For annual and cumulative monthly rates, the person-time at risk for the denominator will accumulate from cohort entry (re-entry) date until the occurrence of the AESI, or the end of the year/month, whichever happens first.

Numerator: The number of incident AESIs will be enumerated in a similar fashion as specified for the general population in Section 8.1.1, and includes incident cases among those who contribute person-time at risk to the denominator.

Rate calculation: The numerator will be divided by the corresponding summed person-time at risk (by year and cumulative month) and expressed as an incidence rate per 100,000 person-years or person-months. The 95% confidence interval of the rates will be estimated as specified in Section 8.1.1.

8.1.3 Negative Control Events

The annual and cumulative incidence rates for negative control events will be calculated in the same manner as specified for the general population in Section 8.1.1 among the populations eligible for the negative control events analysis.

8.1.4 COVID-19 Disease Incidence

Among the population for the evaluation of COVID-19 diagnosis as specified in Section 7.4, person-time at risk for the denominator will be summed from March 1, 2020, cumulatively by week and by month, until the end of 2020. Incidence rates will be calculated by dividing the number of the incident COVID-19 cases (cumulative weeks and cumulative months) by the corresponding denominators.

In data sources with lab results, for a subset of patients with a COVID-19 incident diagnosis and a COVID-19 RT-PCR test result within +/-7 days of the diagnosis, the proportion of those who had a positive COVID-19 result will be reported, by cumulative week and month.

Among patients with a COVID-19 incident diagnosis, the proportion of the patients who had a hospitalization overlapping the time period of -7 to +14 days of the incident diagnosis will be reported, by cumulative week and month.

8.2. Summary of Stratification and Subgroup Analyses

To fulfill the primary objectives, AESI incidence rates will be estimated by calendar year and data source over the period of 2017 to 2020. Additionally, in 2020, cumulative weekly and cumulative monthly incidence rates of COVID-19 will be estimated by data source.

In secondary analyses, cumulative monthly incidence rates of the AESIs are estimated within each calendar year of 2017 to 2020. AESI incidence rates are further estimated by stratifying by age, sex, and race (where reliably available) and among special subpopulations, including pregnant women, older adults, pediatric populations, and individuals with a recent history of influenza vaccination.

AESI incidence is likely to differ by population characteristic of the study population. Therefore, background rates will be estimated within strata defined by the following demographic characteristics:

- Calendar year or month (or week for incident COVID-19 cases)
- Age group
    - Pediatric populations: 0–5 months, 6–23 months, 24 months–4 years, 5–11, 12–15, and 16–17 years for data sources with birth date; 0–1 year, 24 months–4 years, 5–11, 12–15, and 16–17 years for data sources with birth year and/or partial birth date
    - Adults: 18–25, 26–35, 36–45, 46–55, and 56–64 years
    - Older adults: 65–74, 75–84, and 85+ years
- Sex
- Race/ethnicity (where reliably available)

Additional finer stratifications (i.e., age*sex and age*sex*race/ethnicity) will be conducted as part of background rate estimation. The results will not only inform knowledge of rates pre-COVID-19 vaccines but also how they vary across populations.

Planned stratified analyses in the general population and subpopulations are summarized in Table 4. Subpopulations are subsets of the general population based on age groups, pregnancy outcomes, or exposure to an influenza vaccine in the past year. Details of the study populations are described in Section 7.

Table 4. Planned Stratified Analyses in the General Population and Subpopulations for the Outcomes of Interest in Claims Data Sources

Populations	Safety AESIs	Negative Control Events	Pregnancy Outcomes	COVID-19
General population	Stratification: by data source, and by age*sex*data source or age*sex*race/ethnicity (where reliably available) Time period: cumulative monthly and annual incidence rates from 2017–last data available until December 11, 2020; for the 2020 cohort pre-COVID-19 period (January and February 2020) vs. COVID-19 period	Stratification: by data source, and by age*sex*data source or age*sex*race/ethnicity (where reliably available) Time period: cumulative monthly and annual incidence rates from 2017–last data available until December 11, 2020; for the 2020 cohort pre-COVID-19 period (January and	N/A	Stratification: by data source Time period: cumulative weekly (where reliably available) and cumulative monthly incidence rates from March 1, 2020–last data available until December 11, 2020 Data sources: BHI, Medicare, MarketScan, Optum

Populations	Safety AESIs	Negative Control Events	Pregnancy Outcomes	COVID-19
	Data sources: BHI, Medicare, MarketScan, Optum	February 2020) vs. COVID-19 period Data sources: BHI, Medicare, MarketScan, Optum		
Pediatric	Stratification: by data source Time period: cumulative monthly and annual incidence rates from 2017–last data available until December 11, 2020 Data sources: BHI, MarketScan, Optum	N/A	N/A	Stratification: by data source Time period: cumulative monthly incidence rates from March 1, 2020–last data available until December 11, 2020 Data sources: BHI, MarketScan, Optum
Older adults	Stratification: by data source Time period: cumulative monthly and annual incidence rates from 2017–last data available until December 11, 2020 Data sources: Medicare	N/A	N/A	Stratification: by data source Time period: cumulative weekly (Medicare only) and cumulative monthly incidence rates from March 2020–last data available until December 11, 2020 Data sources: Medicare
Pregnant women	Stratification: by data source Time period: cumulative monthly and annual incidence rates from 2017–last data available until December 11, 2020 Data sources: BHI, MarketScan, Optum	N/A	Stratification: by data source Time period: cumulative monthly and annual incidence rates from 2017–last data available until December 11, 2020 Data sources: BHI, MarketScan, Optum	Stratification: by data source Time period: cumulative monthly incidence rates from March 2020–last data available until December 11, 2020 Data sources: BHI, MarketScan, Optum

Populations	Safety AESIs	Negative Control Events	Pregnancy Outcomes	COVID-19
Influenza-vaccinated population	Stratification: by data source, and by age*sex*data source or age*sex*race/ethnicity (where reliably available) Time period: cumulative monthly and annual incidence rates from 2017–last data available until December 11, 2020 Data sources: BHI, Medicare, MarketScan, Optum	N/A	N/A	N/A

Abbreviations: BHI, Blue Health Intelligence; AESI, adverse event of special interest; N/A, not applicable.

9. Study Population in EHR Sources

9.1 General Population of Interest

Similar to claims data sources, annual cohorts will be identified by calendar year within the study period, in 2017 (where available), 2018, 2019, and 2020.

Unlike claims data sources, nonintegrated EHR data sources lack the concept of claims-based enrollment to estimate person-time. Instead, the patient population in EHR networks comprise those that interact with the healthcare providers that feed their data to EHR systems. For this study, eligible individuals are defined as care-seekers that had at least one encounter in any of the EHR data sources in a given period. Incidence proportions instead of person-time incidence rates will be calculated.

Each annual cohort will include any individual who had at least one health encounter in a given data source during the calendar year and met the clean period requirement before the cohort entry, as specified below. The *EHR-specific clean period requirement* is defined as having no observed AESI during the length of the clean period prior to January 1, either based on diagnosis recorded on encounters during the clean period or on the problem list with onset dates within the clean period. This is in an effort to capture incident cases for all care-seeking patients in each analysis year, regardless of their care-seeking behavior in the prior calendar year. Some EHR sources, such as MedStar, include a large network of urgent care providers. Patients having an encounter and, therefore selected for a given calendar year, are not expected to also have an encounter in the same network in the prior year. The use of the problem list will increase the number of patients included for analysis, even though we note the potential inaccuracy of the problem list.

9.2 Special Populations of Interest

Special populations of interest include older age groups (i.e., 65+ years on January 1 of each calendar year), pediatric populations (i.e., <18 years on January 1 of each calendar year), pregnant women, and individuals who received the influenza vaccination in the prior calendar year. The two latter populations are described in subsequent sections.

For pediatric AESIs that are limited to a population of interest within an age range (e.g., Kawasaki disease among 1-5 years old), an individual must be within the age range for at least 1 day during a calendar year to be included in a given annual cohort, in addition to meeting the EHR-specific clean period requirement.

9.2.1 Pregnant Women

Given the challenges of observability of health care encounters outside of the network, there are no reported algorithms to construct pregnancy episodes in nonintegrated EHR data systems. For assessment of pregnancy outcomes in each calendar year, the pregnant women include those who are aged 12–55 years on January 1 of the calendar year and have any of the pregnancy outcome codes any time during the calendar year. In the same manner specified in the algorithms[3] described in Section 7.2.1, clinical codes indicating pregnancy endpoints including live birth, spontaneous abortion, elective abortion, stillbirth, ectopic, trophoblastic pregnancy and unknown pregnancy outcome will be identified. A set of hierarchical and spacing rules as described in the algorithms[3] will be applied to distinguish outcomes codes from the same pregnancy vs separate pregnancies and to place different outcomes of the same woman on the timeline. No further construction of pregnancy episodes will be conducted, because it requires longitudinal observability during the gestational period. Live births will be classified into preterm and full-term live births based on clinical codes indicating gestational age during the same encounter of the outcome (preterm live births <37 weeks of gestation, full-term live births 37+ weeks of gestation). If no gestational age information is available for a live birth, the pregnancy outcome will remain as a live birth. The unit of analysis is pregnancy, defined by each pregnant outcome in a calendar year. A woman with multiple pregnancy outcomes will be included multiple times for a given year. Nonpregnancy safety AESIs during pregnancy will not be assessed because pregnancy episodes cannot be constructed using EHR data.

9.2.3 Individuals with Recent Influenza Vaccination

The subpopulation with a recent history of influenza vaccination includes the subset of the general population, for each AESI, that had an observed influenza vaccination in the prior calendar year based on the presence of influenza vaccination codes.

9.3 Population to Evaluate Negative Control Events

Cohort inclusion criteria are different for negative control events of hypertension and colonic diverticulitis and for those that are preventive care services (well-child and well-care visit, colonoscopy for colorectal cancer screening, mammogram for breast cancer screening, or cervical cancer screening).

For hypertension and colonic diverticulitis, the annual cohort includes those who had at least one encounter during the calendar year, met the EHR-specific clean period requirement and were over 18 years old for at least 1 day during the calendar year.

For preventive care services, the annual cohort includes those who had at least one encounter during the calendar year and were within the specified age range for the preventive care service for at least 1 day during the calendar year.

9.4 Population to Evaluate COVID-19 Disease Incidence

The population to evaluate COVID-19 incidence proportion includes individuals who had at least one health encounter between March 1, 2020, and December 11, 2020. No clean period will be required for this population because incidence of COVID-19 will likely not have been observed before this period.

10. Analysis in EHR Data Sources

10.1 Background Incidence Proportion Estimation in EHR Data Sources

10.1.1 General Population

Denominator: Individuals who meet the EHR-specific clean period requirement as specified in Section 9 in each calendar year will remain in the annual cohort for the calendar year. Individuals will not be censored or excluded at the occurrence of an AESI. The denominator for the annual proportion estimate is the count of eligible individuals in the annual cohort. The denominator is the count of individuals who had at least one encounter between January 1 and the month of evaluation. For example, the denominator for the June proportion is the number of individuals who had at least one encounter between January 1 through June 30.

Numerator: The number of incident cases among individuals in the denominator will be enumerated for a calendar year to form the numerator for the annual incidence proportion. The number of incident cases from January 1 of the year through the end of the month evaluation among individuals in the same time period will be summarized to form the numerator of the cumulative monthly proportions.

Incidence proportion calculation: The numerators will be divided by the corresponding denominators (by year and cumulative month) and expressed as an incidence proportion per 100,000 individuals. The 95% confidence interval of the annual and cumulative monthly proportions will be calculated using the Agresti-Coull approximate binomial interval that improves on the normal approximation by achieving the nominal coverage probability[7,8].

Sensitivity analysis: As some patients may only have one health care encounter due to the outcome of interest in a given period, the denominator will be restricted to patients who had at least one encounter due to reasons other than the outcome of interest in a given period to define the population at risk. The numerator will enumerate the number of incident cases among individuals in the denominator in the same period. The incidence proportion will be calculated in the same fashion above.

10.1.2 Special Populations of Interest

10.1.2.1 Pediatric Populations and Older Adults

The annual and cumulative monthly incidence proportion calculations among the pediatric and older adult populations are consistent with the methods described above for the general population in Section 10.1.1.

For pediatric AESIs that are limited to a specific age range of interest (as described in Section 9.2), the denominator will be limited to the individuals who were within the age range for at least 1 day in a given calendar year or cumulative months; the numerator will be limited to incident AESIs occurring on a date when the individual was within the age range.

10.1.2.2 Pregnant Population

Denominator: Denominator for the annual incidence proportion will include all pregnancy outcomes identified during the calendar year as specified in Section 9.2.1. Denominator for the cumulative monthly proportion will include those pregnancy outcomes between January 1 and the end of the month evaluation.

Numerator: Numerator is the number of pregnancy outcomes of interest (preterm births, full-term live births, spontaneous abortions, stillbirths). They will be counted annually and cumulatively through the month evaluation.

Proportion calculation: The numerator will be divided by the corresponding denominator calculated above. Proportions will be expressed as incident cases per 100,000 pregnancies. The 95% confidence interval of the proportions will be estimated as defined in Section 10.1.1.

10.1.2.3 Individuals with Recent Influenza Vaccination

For the subpopulation with a recent history of a seasonal influenza vaccine as defined in Section 9.2.3, the annual and cumulative monthly incidence proportions will be calculated using the same methods specified in Section 10.1.1.

10.1.3 Negative Control Events

The annual and cumulative monthly incidence proportions for negative control events will be calculated in the same manner as specified for the general population in Section 10.1.1 among the population eligible for negative control events analysis.

10.1.4 COVID-19 Disease Incidence

Among the population for COVID-19 evaluation as specified in Section 9.4, the denominator for the cumulative weekly and cumulative monthly proportion will include the number of individuals who had at least one encounter from March 1, 2020, through the week/month of evaluation. Incidence proportions will be calculated by dividing the number of the incident COVID-19 cases (cumulative weeks and months) by the corresponding denominators. Sensitivity analysis will restrict to the denominator to individuals who had at least one encounter due to reasons other than COVID-19 from March 1, 2020, through the week/month of evaluation.

For a subset of patients with an incident COVID-19 diagnosis and a COVID-19 RT-PCR test result within +/-7 days of the diagnosis, the proportion of those who had a positive COVID-19 result will be reported, by cumulative week and month.

Among patients with a COVID-19 incident diagnosis, the proportion of the patients who had a hospitalization overlapping the time period of -7 and +14 days from the incident diagnosis will be reported, by cumulative week and month.

10.2 Summary of Stratification and Subgroup Analyses in EHR Data Sources

Analysis stratifications in the general population and subpopulations will be conducted in the same manner as specified in Section 8.2. These analyses are summarized in Table 5. Details of

the study populations are described in Section 9. All available data sources in Section 4.2 will be queried.

Table 5. Planned Stratified Analyses in the General Population and Subpopulations for the Outcomes of Interest in EHR Data Sources

Populations	Safety AESIs	Negative Control Events	Pregnancy Outcomes	COVID-19
General population	Stratification: by data source, and by age*sex*data source or age*sex*race/ethnicity (where reliably available) Time period: Cumulative monthly and annual incidence proportions from 2018–last data available until December 11, 2020; for the 2020 cohort pre-COVID-19 period (January and February 2020) vs. COVID-19 period	Stratification: by data source, and by age*sex*data source or age*sex*race/ethnicity (where reliably available) Time period: Cumulative monthly and annual incidence proportions from 2018–last data available until December 11, 2020; for the 2020 cohort pre-COVID-19 period (January and February 2020) vs. COVID-19 period	N/A	Stratification: by data source Time period: cumulative weekly and monthly incidence proportions from March 2020–last data available until December 11, 2020
Pediatric	Stratification: by data source Time period: cumulative monthly and annual incidence proportions from 2018–last data available until December 11, 2020	N/A	N/A	Stratification: by data source Time period: cumulative weekly and monthly incidence proportions from March 2020–last data available until December 11, 2020
Older adults	Stratification: by data source Time period: Cumulative monthly and annual incidence proportions from 2018–last data available until December 11, 2020	N/A	N/A	Stratification: by data source Time period: cumulative weekly and monthly incidence proportions from March 2020–last data available until December 11, 2020

Populations	Safety AESIs	Negative Control Events	Pregnancy Outcomes	COVID-19
Pregnant women	N/A	N/A	Stratification: by data source Time period: cumulative monthly and annual incidence proportions from 2018–last data available until December 11, 2020	N/A
Influenza-vaccinated population	Stratification: by data source, and by age*sex*data source or age*sex*race/ethnicity (where reliably available) Time period: cumulative monthly and annual incidence rates from 2018– last data available until December 11, 2020	N/A	N/A	N/A

Abbreviations: EHR, electronic health record; AESI, adverse event of special interest; N/A, not applicable.

11. Limitations

11.1 Identification of Pregnancies in Claims Data Sources

The subpopulation of pregnant women of interest specified in this protocol includes only those pregnancies that had an observed outcome during the study period. Ongoing pregnancies without a recorded outcome are not included, which particularly affects the estimates of person-time at risk during pregnancy in calendar year 2020. The algorithm will only include those pregnancies ending in live birth, stillbirth, and spontaneous abortion, not other rarer outcomes such as ectopic, molar or other abnormal products of conceptions, or elective abortion. Claims-based algorithms cannot comprehensively capture all early losses such as spontaneous abortions, especially those that occur early during pregnancy before the pregnancy is detected by the mother. Some claims have evidence of delivery (e.g., cesarean section) without documentation of an outcome; these also are not included. A small percentage of pregnancies are excluded because of conflicting coding of outcome types or the estimated gestational age not agreeing with the outcome type. For example, some stillbirths with an estimated gestational of <20 weeks are excluded. Although they may have been miscoded spontaneous abortions, no attempt was made to reclassify them (if they were reclassified, they will likely add about 1% of spontaneous abortions based on prior data). Pregnancy outcomes for which a gestational age estimate could not be determined using information in the claims are excluded. For example, a missed abortion with no information of gestational age is not included because pregnancy start date cannot be assigned.

To accurately determine the time at risk during pregnancy, a pregnancy start date is required, hence the construction of pregnancy episodes. Pregnancy episodes include only those for women continuously enrolled during the pregnancy episode and with an estimated pregnancy start date.

For the reasons mentioned above, the nonpregnancy AESI background rates estimated among the pregnant population may not be generalizable to those pregnancies not captured by the algorithms. For pregnancy outcomes, the incidence rate of spontaneous abortions may be underestimated for all calendar years. For calendar year 2020, the incidence rates of all pregnancy outcomes are may be overestimated due to the underestimation of the denominator.

11.2 Independence of AESIs

This protocol will assess each AESI separately and apply the clean period for each AESI. Thus, the AESIs are assumed to be independent of each other, and the analyses do not take competing risks into consideration. This assumption may not be met. For instance, some of the circulatory and vascular AESIs, such as stroke, acute myocardial infarction, pulmonary embolism, and deep vein thrombosis, may be correlated within the same individual.

11.3 Definition of Incidence

This protocol uses a clean period in which the AESI did not occur before an individual can enter the study cohort. The clean periods are set at one year for the majority of AESIs. Individuals who were diagnosed with the AESI prior to the beginning of the clean period and who did not receive a diagnosis for the AESI during the clean period would be included in estimates of incident AESIs. Thus, incident cases during the follow-up period are not necessarily *new* diagnoses, and individuals who satisfy the clean period multiple times may enter subsequent follow-up periods and can have additional incidence AESIs during the study period. Chronic conditions such as narcolepsy of an individual may be captured multiple times as incident cases within a same individual, if the diagnoses occurred far enough part to allow additional clean period requirement to be met. This definition may differ from other ways incidence has been operationalized in epidemiologic studies. However, it was chosen because it aligns with the definition of incidence that will be implemented in the active monitoring study protocol. This ensures comparability between the calculation of the background rates and the calculation of the rates of AESIs during the vaccination period.

11.4 Limitations of EHRs

There is a lack of observability in nonintegrated EHR networks for vaccinations, safety AESIs, pregnancy outcomes, AESIs during pregnancy, negative control events, and COVID-19. This is because patients may seek care out of network. As a result, this protocol estimates an incidence proportion instead of a rate because entry into a study cohort and the follow-up period cannot be determined reliably. Measurement of incident cases is among individuals who had at least one EHR activity during the calendar year, and by doing so, it assumes that people who have EHR activities will seek care for the AESI in the same network during the same period and therefore are observable. This assumption is untestable; thus, the estimated incidence proportion may not reflect the incidence proportions in the underlying population. To define population at risk and address the “look-ahead” bias, sensitivity analysis is proposed by requiring at least one health care encounter due to reasons other than the outcome of interest. However, such restriction results in a smaller and likely sicker population.

11.5 Generalizability

The results produced by this protocol should be interpreted based on the generalizability of the populations included. For BHI, this is limited to enrollees in Blue Cross and Blue Shield insurance plans. While the BHI data set includes more than 200 million unique lives, detailed data for this study was limited to a cohort of enrollees who received a biologic product, were pregnant, or were born after October 1, 2015. For MarketScan data, the population is limited to individuals with employer-sponsored private insurance coverage. CMS data are limited to Medicare enrollees. Various EHR data cover different segments of the U.S. population with respect to region (e.g., Medstar: District of Columbia, Maryland, Virginia; OneFlorida: Florida). Still, the collection of these sources covers wide-ranging populations in the United States to prepare for and implement active monitoring of COVID-19 vaccines.

12. References

[1] World Health Organization (2020). Draft landscape of COVID-19 candidate vaccines. Available at: https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines. Accessed November 13, 2020.

[2] Duffy, J., et al., *Febrile seizure risk after vaccination in children 6 to 23 months.* Pediatrics, 2016. **138**(1).

[3] Moll, K., et al., *Validating claims-based algorithms determining pregnancy outcomes and gestational age using a linked claims-electronic medical record database.* Unpublished results, 2020.

[4] Cantor, J., et al., *The Impact of the COVID-19 Pandemic and Policy Response on Health Care Utilization: Evidence from County-level Medical Claims and Cellphone data*. NBER Working Paper, (w28131).

[5] Moll, K., et al., *Final report: Validating pregnancy outcomes and gestational age in a claims-EMR linked database*. FDA BEST Initiative. https://www.bestinitiative.org/wp-content/uploads/2020/08/Validating_Pregnancy_Outcomes_Linked_Database_Report_2020-1.pdf

[6] Ulm, K., *Simple method to calculate the confidence interval of a standardized mortality ratio (SMR)*. American Journal of Epidemiology. 1990;131(2):373-375. doi:10.1093/oxfordjournals.aje.a115507

[7] Agresti, A. and Coull, B. A. (1998). *Approximate is better than "exact" for interval estimation of binomial proportions*. The American Statistician 52; 119-126

[8] Brown, Cai, and DasGupta (2001). Interval estimation for a binomial proportion. Statistical Science 16(2); 101-133

13. Appendix

13.1 Care Setting Definitions in Claims

The following table summarizes how each setting will be defined for AESI and vaccine exposure identification in claims data sources. Note that the OP-ED setting is a subset of the OP/PB setting.

Table A1. Care Setting Definitions in Claims

Setting	Definition
Inpatient (IP)	Inpatient acute facility claim (e.g. UB-04 with type of bill = 11x) discharge diagnoses
Outpatient Emergency Department (OP-ED)	Outpatient facility claim (e.g. UB-04) in ED diagnoses
Outpatient & Professional (OP/PB)*	Outpatient facility claim (UB-04) diagnoses OR Professional claim (CMS-1500) diagnoses with non-lab place of service[#]

*Including all sources of professional claims (e.g. urgent care etc.)
[#]Independent laboratory place of service code = 81