

**Virtual Network: Investigating the Risk of COVID-19-Associated  
Outcomes and COVID-19 Vaccine Effectiveness Using Integrated  
Medical and Public Health Records (VISION-COVID)**

**Centers for Disease Control and Prevention  
& Westat**

**DRAFT vs 4.0: 6/9/2021**

## **1.0 Study Summary**

### **1.1 Title of Study**

**Investigating the Risk of COVID-19-Associated Outcomes and COVID-19 Vaccine Effectiveness Using Integrated Medical and Public Health Records**

### **1.2 Investigators**

*Centers for Disease Control and Prevention*

- Catherine Bozio
- Mark Thompson
- Eduardo Azziz-Baumgartner
- Lenee Blanton
- Sue Reynolds
- Jill Ferdinands
- Zijing Guo
- Palak Patel
- Eric Griggs
- Carrie Reed
- Alicia Fry

*Westat*

- Sarah Ball
- Patricia Shifflett
- Rebecca Birch
- Rebecca Fink
- Saloni Sapru

- Beth Mittl
- Keith Umbel
- Sarah Bevan
- Carly Hallowell
- Matthew Levy
- Maria Demarco

*Participating sites*

- HealthPartners Institute
- Kaiser Permanente Northwest
- Universities of California at Irvine, Los Angeles, San Diego, San Francisco and Davis
- University of Colorado, Denver
- Atrium Health
- Intermountain Healthcare
- Regenstrief Institute
- Columbia University
- Baylor, Scott and White Health

## **2.0 Protocol Objective**

We aim to develop a multi-site collaboration to utilize electronic medical and public health records to address key questions about the clinical epidemiology of COVID-19 and COVID-19 vaccine effectiveness. Specifically, we will integrate additional sites (specifically, Atrium Healthcare, Intermountain Healthcare, Regenstrief Institute, Columbia University, and Baylor Scott and White Health) into an existing virtual network of

healthcare systems that follows an IRB-approved protocol. Through this collaboration of healthcare systems and research organizations, the primary aim is to estimate the effectiveness of COVID-19 vaccination in preventing laboratory-confirmed COVID-19 associated health and medical utilization outcomes as well as the incidence of COVID-19 associated outcomes by socio-demographic and high-risk groups. Similar objectives will be accomplished for influenza virus disease and vaccination during periods of local influenza circulation. Deidentified individual-level data will be extracted from medical, laboratory, and vaccination records from each site and defined using a common codebook to accomplish these objectives.

### **3.0 Scientific Background**

#### ***Burden of SARS-CoV-2***

SARS-CoV-2 continues to be of broad global public health importance and concern due to the global burden of COVID-19 illness and other associated medical complications and severe outcomes identified to date, since the initial identification of the virus in Wuhan, China in December 2019. As of November 23, 2020 in the United States, SARS-CoV-2 infections have resulted in more than 12.1 million cases of COVID-19, with a current overall cumulative COVID-19 hospitalization rate of 228.7 per 100,000, and more than 255,000 deaths.<sup>1</sup> Early evidence points to significant burden of severe illness related to SARS-CoV-2 infection by race and ethnicity, age, socio-economic status, and underlying health condition status.<sup>2</sup> Yet, there are numerous knowledge gaps regarding how COVID-19 and especially severe manifestations of disease impact on different socio-demographic and health risk groups.

Emerging data on race and ethnicity suggest an overrepresentation of non-Hispanic black patients among those hospitalized for COVID-19.<sup>3</sup> A study using integrated electronic health records identified adults with suspected and confirmed COVID-19 found that compared with non-Hispanic white patients, black patients had 2.7 times the odds of hospitalization, after adjusting for age, sex, comorbidities, and income.<sup>4</sup>

While all people are susceptible to infection with this novel virus, older adults have elevated rates of COVID-19-associated hospitalization and the majority of persons hospitalized with COVID-19 have underlying medical conditions.<sup>5</sup> In contrast to influenza disease which can be severe in both young children and older adults<sup>6</sup>, this has not been true with COVID-19. Population data from China and Italy indicate that children are mildly affected in comparison to adults, representing approximately 5% of cases and less than 1% of admissions to hospital.<sup>7</sup> A report describing the burden of COVID-19 infection in North American pediatric intensive care units confirmed that severe COVID-19 can occur in children, though this occurs far less frequently than among adults.<sup>8</sup>

COVID-19 has also notably resulted in severe and life-threatening disease among working-age adults and those without prior underlying medical conditions. Further information is needed on the different manifestations or phenotypes of severe COVID-19 across age, health, and other risk groups.

### ***Burden of other respiratory viruses and COVID-19***

Respiratory viruses, including influenza and respiratory syncytial virus (RSV) among others commonly circulate in the United States during a typical influenza season. However, the incidence of other respiratory viruses during the 2020-21 influenza season and our ability to conduct surveillance may be impacted by COVID-19 in multiple ways. The circulation of

seasonal influenza viruses may be reduced as a result of altered health behaviors (e.g., wearing masks, washing hands) and social distancing measures in response to COVID-19. For example, a quasi-experimental study assessing the trends in seasonal influenza cases from the 2014-2015 season to the 2019-2020 season in 11 countries and regions, found that in East Asia, the number of seasonal influenza cases in the 2019-20 season was lower after the onset of COVID-19 transmission compared to previous years.<sup>9</sup> Influenza testing practices may also change; indeed, influenza testing across the U.S. was higher than normal during April 2020 because of the COVID-19 pandemic.<sup>10</sup> Consequently, data collection on co-circulating respiratory viruses with SARS-CoV-2, especially with respect to testing, care-seeking behavior, and vaccination, will provide important context in interpreting COVID-19 results.

### ***Clinical Features of Severe COVID-19***

Although medically attended COVID-19 typically manifests as an acute respiratory disease, SARS-CoV-2 infection can also cause a variety of other clinical manifestations, including myocardial dysfunction, acute kidney injury and neurologic illness.<sup>11</sup> More information is needed on the different clinical phenotypes of COVID-19. Similar to research on severe influenza disease<sup>12-16</sup>, distinctions can be made between virus disease resulting in acute lung injury and often bacterial superinfection, virus disease associated with extrapulmonary disease, and secondary complications to other organs systems, including neurological impairment. Information is also needed on how these different clinical phenotypes occur within the context of otherwise healthy children and adults versus frail individuals with inadequate physiologic reserve.

Most VE studies view hospitalization as the hallmark of severe disease.<sup>17</sup> Yet, the decision to admit a patient is influenced by many factors (e.g., age, medical history, financial and

family resources, clinical decision-making, resource availability) independent from the signs and symptoms of disease. Therefore, more information is needed to differentiate between the clinical severity COVID-19 versus other host and environment factors (such as the local attack rate and population susceptibility) that may be associated with the clinical threshold for hospital admission and discharge and how this may vary between hospitals, medical systems, and regions.

### ***COVID-19 Vaccine Effectiveness***

Although results on the immunogenicity of COVID-19 vaccines from Phase II trials<sup>18-19</sup> and early reports on the clinical effectiveness of the vaccines from Phase III trials are encouraging, there will soon be an urgent need for data on the field performance of these vaccines and their effectiveness in preventing laboratory-confirmed COVID-19-associated emergency department and urgent care (ED/UC) visits and hospitalizations.

Evaluating the real-world performance of COVID-19 vaccines is important for at least five reasons. First, since immune response to vaccines and their subsequent protection against infection and disease often vary by sex, age, underlying health status, and other host factors, the preventive benefit of new vaccines may differ in the general public compared to participants in randomized clinical trial.

Second, real-world assessments can examine preventive benefits months following vaccination in contrast to the relatively brief surveillance evaluations in Phase III trials. For example, if immune protection following vaccination wanes over time, the impact of waning on VE against medically attended COVID-19 can only be assessed over an extended period of evaluation. Similarly, real-world assessments will also be needed to determine if VE of new vaccines vary depending on the amount and duration of virus exposure and the extent of personal

and community infection control measures. Such effects could result in differences in VE by sex, age, race and ethnicity, occupation, socio-economic status, rural versus urban settings, and other host and environmental factors.

Third, the field performance of vaccines depends in part on the conditions of their administration and adherence to cold-chain requirements. This will be especially challenging for some messenger RNA manufactured COVID-19 vaccines that must be stored at -70° Celsius.<sup>20</sup> Studies of the real-world management and administration of influenza vaccines often identified gaps in cold chain management that may have compromised vaccine immunogenicity.<sup>21</sup> Therefore, it is unclear whether challenges in the administration of new COVID-19 vaccines may result in compromised performance that would only be detected in downstream evaluations of their clinical VE.

Fourth, Phase III trials are focused on symptomatic SARS-CoV-2 infection as the primary outcome, and most of these will be relatively mild illnesses. Therefore, field studies are required to assess VE in preventing less frequent but severe outcomes, including COVID-19 associated hospitalizations and very severe outcomes, including intensive care unit (ICU) admissions, mechanical ventilation, and/or death. VE in preventing secondary complications and sequelae following COVID-19 must also be assessed in large prospective population evaluations. Certainly, information on VE against this continuum of medically attended COVID-19 outcomes by socio-demographic, underlying health, and other risk groups will also require assessments with large populations.

Fifth, since Phase III trials evaluate each vaccine product separately, real-world evaluations are required to compare the field VE of different vaccine products in different population groups. It is likely that multiple vaccine products will be rolled out to different



populations, at different times, and in different geographic regions making such comparisons challenging.

### **3.1 Rationale and Justification**

CDC communicates with medical and public health professionals and the public about the burden of SARS-CoV-2 and the importance of prevention and control measures, which will include COVID-19 vaccines when they are available. Timely information on COVID-19 burden and COVID-19 VE will inform public health models that determine public health policy, guidance, and resource allocations. This information will also be used to inform, educate, and guide the public on ways to protect themselves and their family from this new virus. Given widespread skepticism and hesitancy associated with the new COVID-19 vaccines,<sup>22-23</sup> timely information on the real-world value of COVID-19 vaccines in preventing severe disease and impairment is especially important.

To estimate the burden of COVID-19 and the effectiveness of newly available COVID-19 vaccines in preventing this burden, a prospective assessment of patients receiving care in ED/UC settings and hospitals is required. This effort expands the number of participating health systems and research organizations participating within the existing VISION-COVID network. The network is being expanded to cover more diverse populations in more geographic regions in the US since this increases the capacity to assess different COVID-19 vaccine types and increases the likelihood that the virus will be circulating within the study population. Increasing the total number of ED/UC visits and hospitalizations will also increase CDC's ability to conduct timely estimates of VE among early vaccine target groups and then ultimately among the general population. The combined observations across the network will also facilitate estimates of VE in

relatively small risk groups and against low frequency but severe COVID-19 outcomes and sequelae.

## **4.0 Study Objectives**

We aim to develop a multi-site collaboration to utilize medical, laboratory, vaccination, and other public health records to address key questions about the clinical epidemiology of COVID-19 and COVID-19 VE. Similar questions will be examined for influenza epidemiology and VE during periods of local influenza circulation. These objectives will be accomplished separately by different age strata adults and children. Specifically, we will integrate additional sites into an existing virtual network of healthcare systems, using a common research methodology, data platforms, and data dictionary. The following primary and secondary objectives are planned.

### **4.1 Primary Objectives**

1. Estimate the COVID-19 vaccine effectiveness (VE) in preventing hospitalizations associated with laboratory-confirmed SARS-CoV-2 infection, and do so:
  - a) Using alternative case definitions for COVID-19 disease that will be linked with SARS-CoV-2 testing, including existing COVID-19 medical case definitions and broader syndromic manifestations of disease
  - b) Against severe inpatient outcomes, including ICU admission, invasive mechanical ventilation, prolonged length of stay (LOS), or death
  - c) By socio-demographic and high-risk groups, including age, race and ethnicity, socio-economic status, and underlying medical conditions

- d) By COVID-19 vaccine product, vaccine type (e.g., mRNA-, plasmid DNA-, adenovirus-based), dose (e.g., one dose vs. two dose), and time since vaccination.
2. Among study sites with well-characterized source populations, estimate the rate of hospitalizations associated with laboratory-confirmed SARS-CoV-2 infection since February, 2020 by socio-demographic, underlying health, and other high-risk groups.

## **4.2 Secondary Objectives**

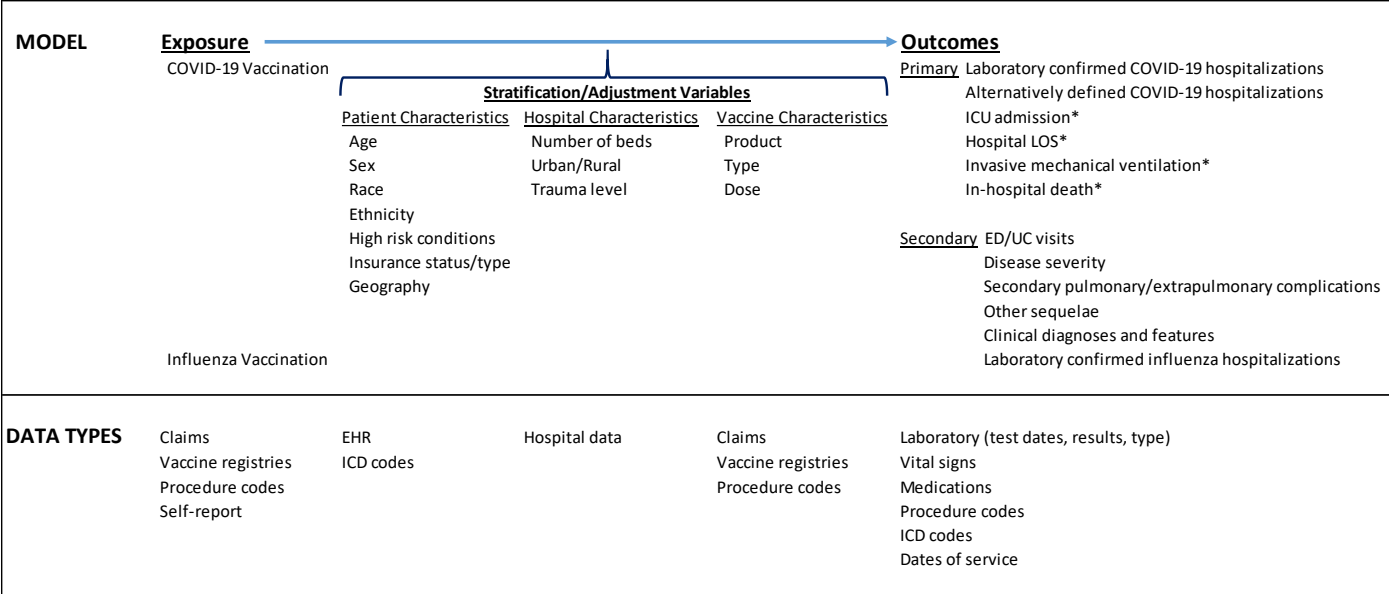
1. Estimate COVID-19 VE in preventing ED/UC visits associated with laboratory-confirmed SARS-CoV-2 infection, and do so by the populations, outcomes, and vaccine types described in Primary Objective 1.
2. Estimate COVID-19 VE in preventing outpatient visits associated with laboratory-confirmed SARS-CoV-2 infection within the pediatric population, and do so by the populations, outcomes, and vaccine types described in Primary Objective 1.
3. Assess whether COVID-19 vaccination attenuates disease severity among those with breakthrough SARS-CoV-2 infections, as indicated by:
  - a. Estimating the odds of severe hospital outcomes (such as ICU admission, mechanical ventilation, prolonged LOS, or death) associated with COVID-19 among vaccinated versus unvaccinated hospitalized patients
  - b. Estimating such effects by age, underlying health status, vaccine type and doses, and other potential effect modifiers.
4. Rule-out whether COVID-19 vaccination is associated with increased likelihood of severe disease (i.e., enhanced disease) among those with breakthrough SARS-CoV-2 infections, as indicated by increased risks of severe outcomes (described in Primary Objective 4) among vaccinated patients.

5. Among study sites that can describe moderate- or long-term outcomes following index hospitalizations, estimate COVID-19 VE against secondary pulmonary and extrapulmonary complications and sequelae following discharge from a laboratory-confirmed COVID-19 hospitalization, and do so by the populations, outcomes and vaccine types described in Primary Objective 1.
6. Describe the clinical testing practices, primary and secondary clinical diagnoses, and other clinical factors and features associated with laboratory-confirmed SARS-CoV-2 infections diagnosed within pediatric outpatient clinics, ED/UC, and hospitals and how these different activities and manifestations vary by medical history and socio-demographic, underlying health, and other risk groups.
7. Estimate the influenza VE in preventing hospitalizations associated with laboratory-confirmed influenza virus infection during periods of local influenza circulation at study sites, and do so by populations, outcomes, and vaccine types similar to those described in Primary Objective 1.
8. For study sites with well-characterized source populations, estimate the rate of hospitalizations associated with laboratory-confirmed influenza virus infection during periods of local influenza circulation at study sites, and do so by socio-demographic, underlying health, and high-risk groups.

## **5.0 Methods**

### **5.1 Study design**

Overall study activities are summarized in the following figure:



\*among COVID-19 hospitalizations

### 5.2 Participant eligibility

The initial analytic population is intentionally broad in order to examine a wide range of COVID-19 disease manifestations and the variety of patients that are tested for SARS-CoV-2 infections. Specifically, individuals of all ages who have an index ED/UC or inpatient encounter associated with any acute illness and/or have respiratory virus testing performed at an index ED/UC or inpatient encounter at a healthcare facility within the network during the study period will be included.

### 5.3 Observation time

SARS-COV-2 is a novel virus and it is unclear whether it will circulate with a defined seasonality. Consequently, the study period will begin on September 1, 2019 and could end on July 30, 2022. If SARS-CoV-2 is later determined to circulate with a known seasonality, analytic cohorts could be created to reflect seasonality a posteriori and objectives could be performed using these analytic cohorts.

## 5.4 Exposures and covariates of interest

### *Vaccination status*

COVID-19 vaccination is the primary exposure of interest for COVID-19 VE objectives. Influenza vaccination is the exposure of interest for secondary influenza VE objectives. The target groups for VE evaluations and the criteria for full versus partial vaccination with either vaccine will follow the age- and risk-group specific and vaccine-specific guidance of the Advisory Committee on Immunization Practices (ACIP).

Documentation of vaccination status will rely on multiple sources of information. Electronic documentation by health records, state/local vaccine registries, all-payer claims or other billing databases is expected to be the primary method. However, self-report of vaccination status from ED/UC or hospital records may also be considered, especially for patient groups and settings where electronic records may be incomplete. All available vaccine information will be extracted from available sources, including date of vaccination, type of vaccine administered, vaccination route, location of vaccination, vaccine lot number and number of vaccine doses, and these variables will be added to the codebook. For vaccination data extracted from EHRs, sites will provide information on how vaccination data was pulled (for example, through use of CPT, CVX, or other internal immunization codes). Self-reports may also be extracted from EHRs, but it is not anticipated that separate chart abstraction or natural language processing will be required.

Network sites will provide information on the types and number of sources queried for vaccine data. Because each site is querying multiple sources, a hierarchy of sources will be established for each site or through joint consensus with collaborating partners.

For a subset of patients (number to be determined by joint consensus by the collaborating partners), accuracy of extracted vaccination data will be verified through manual review of source data (i.e. medical chart review). For study sites that rely in part on self-report vaccination status, a survey of patients will be conducted to verify their self-reported vaccination status similar to previously published methods of validating influenza vaccination documentation.<sup>24</sup>

### ***Clinical Virus Testing Results***

Data on all SARS-CoV-2, influenza virus, and other respiratory virus testing results will be extracted for index ED/UC visits and hospitalizations. For study sites applying a look-back period or building a patient cohort (referenced later in this section within “Population subgroups and covariates of interest”), clinical testing results will also be extracted for all included patients during the surveillance period. In particular, after identifying an index inpatient encounter (and later ED/UC encounter), any available prior clinical testing results will be examined to identify any positive test result for a virus (e.g., SARS-CoV-2) before said encounter. For each test result, sites will provide data on encounter type associated with the lab test, diagnoses (based on ICD code) associated with encounter during which testing occurred, date of specimen collection, type of test performed and test result. Network sites will provide information on how laboratory testing data was pulled (for example, through use of CPT or other procedure codes or from laboratory databases). For a subset of patients (number to be determined by joint consensus by the collaborating partners), accuracy of extracted laboratory data will be verified through manual review of source data (i.e. medical chart review or review of laboratory databases). Respiratory viruses for which data will be collected, in addition to influenza, include respiratory syncytial virus, adenovirus, parainfluenza virus, human metapneumovirus, rhinovirus, enterovirus, coronaviruses, and other viruses, including novel viruses.

### ***High risk underlying medical conditions***

For the common data set focused on patient ED/UC encounters or hospitalizations only, high risk underlying medical conditions will be defined using ICD-10 diagnosis codes (see Appendix A). However, site-specific methods may also be applied that draw on broader data sets (e.g., prescribed medications; claims), registries, and look-back periods to assess the presence of high risk conditions. Methods that allow for common versus site-specific approaches and how they apply to minimally versus fully adjusted VE models are described later in this section “Population subgroups and covariates of interest”. High-risk underlying medical conditions include:

- Chronic lung disease;
  - o Asthma;
  - o COPD;
  - o Pulmonary tuberculosis
  - o Endemic mycoses
- Chronic metabolic disease;
  - o Diabetes mellitus;
- Blood disorders;
- Cardiovascular disease;
  - o Coronary artery disease;
  - o Heart failure;
  - o Congenital heart disease;
- Clinical obesity
- Neuromuscular disorder;



- Neurologic disorder;
- Immunocompromised condition;
  - o Solid organ malignancy;
  - o Hematologic malignancy;
  - o Solid organ transplant;
  - o Hematopoietic stem cell transplant;
- Chronic renal disease;
- Gastrointestinal/liver disease;
- Rheumatologic/Autoimmune condition;
- Prematurity (applicable only to pediatric population);
- Medical complexity (applicable only to pediatric population);
- Congenital heart disease (applicable only to pediatric population).

Additional underlying conditions, including, but not limited to, pregnancy, sickle cell disease, hypertension, and cystic fibrosis have also been associated with increased risk of severe influenza or COVID-19 disease and thus might be further defined and explored in the network.

***Data on other host factors, healthcare context, and exposures of interest***

Individual level data will be extracted and provided on:

- Patient age as of the start date of the observation period
- Sex
- Race and ethnicity
- Date of enrollment in the membership health plan or date of first qualifying healthcare encounter within the look-back period (the encounter that included the individual into the

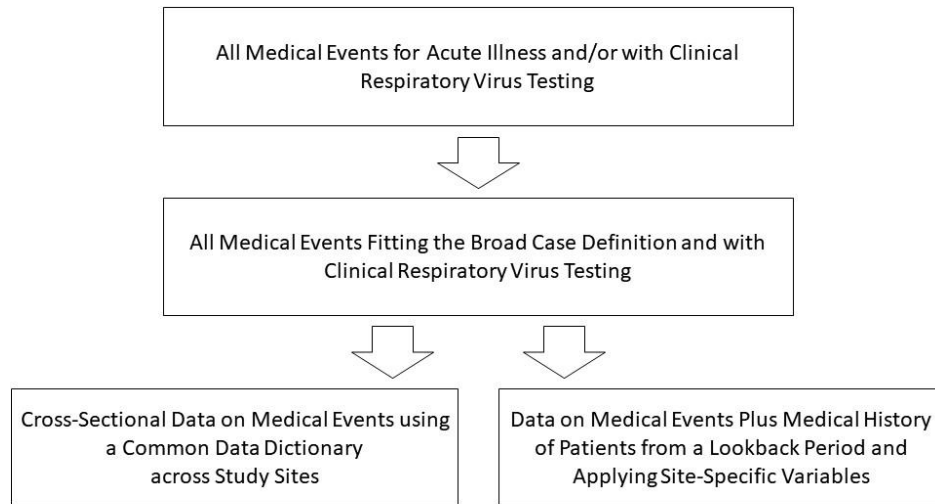
cohort) (only applicable to datasets for the fully adjusted VE model, as referenced later in this section “Population subgroups and covariates of interest”)

- Primary, secondary, tertiary and quaternary insurance type
- Census tract (Algorithms for determining census tract will be determined by group consensus and applied uniformly across the network sites) and similar methods to assess socio-economic status
- Characteristics of the hospitalization (e.g., admission source, discharge disposition, length of stay)
- Variables characterizing the facility (e.g., facility ownership, facility type, urban/rural classification, tertiary and teaching hospitals, geography)
- Variables related to clinical testing (e.g., cycle threshold value, type of PCR test, variable to standardize readings across machines)
- Proxy variables to identify healthcare workers, frontline workers, and essential workers

Additional factors of interest include respiratory support (such as supplemental oxygen non-invasive mechanical ventilation), clinical laboratory values, vital signs, clinical procedures that are indicative of illness severity, or medications used for treatment in the inpatient or outpatient setting. The available data elements and the best algorithms to ascertain these exposure variables will be examined and customized within each study site. As more details on host, environment, and exposure variables that are relevant to the clinical epidemiology of COVID-19 and COVID-19 VE are identified, the methods and protocol will be amended as needed.

***Patient subgroups and covariates of interest for objectives centered on medical events***

Most of the primary and secondary objectives of this effort center on two categories of medical events: ED/UC encounters and hospitalizations. Pediatric outpatient data will be integrated based on sites' feasibility. The following data structure applies to all types of events.



The first level of data centers on all medical events associated with acute illness (using a broad case definition) and/or included respiratory virus testing. This dataset will utilize the cross-sectional data available for that event, including but not limited to diagnostic and syndromic codes, demographic information, laboratory results, and vaccination data. This level of information will allow sites to describe clinical testing practices and reasons for medical care for the full denominator of events that could be considered.

The second level of data narrows this full set of events to those involving a broad case definition for diseases of interest and those for which clinical testing for infection occurred. From these events, a cross-sectional data set will be created using a common data dictionary across study sites. Within each site, a second set of data will be created which adds medical history to each patient event using a lookback period within their electronic records. This data

will consist of common data elements across sites, but the patient inclusion criteria and specific variables may differ by site depending on data availability and structure. This same site-specific data set will also include prospective assessments (look-forward) of outcomes such as re-admissions, complications, and sequelae.

***Patient subgroups and covariates of interest for objectives centered on rates and other population measures***

Most study sites will also create data sets for cohorts based on health plan members and/or medical utilization history. For sites which can characterize source populations in this way, rates of laboratory-confirmed COVID-19 outcomes can be estimated using cohort denominators. Index pediatric outpatient, ED/UC and hospital events can also be examined within the prospective cohort framework in contrast to the cross-sectional approach described above.

## **5.5 Outcomes of interest**

The primary outcome of interest centers on pediatric outpatient visits, ED/UC encounters or hospitalizations associated with an acute respiratory illness or other COVID-19-like illness (ARI/CLI) with laboratory-confirmed SARS-CoV-2 infection. Other outcomes of interest involve alternative case definitions (such as extra-pulmonary disease not included in ARI/CLI), specific types of medical utilization (e.g., intensive care unit admission, invasive mechanical ventilation, and in-hospital death), and secondary events that occur following the index events (such as secondary pulmonary and extrapulmonary complications and sequelae). Additional details in defining some of these outcomes are listed here:

- Hospitalizations, ED/UC encounters, and pediatric outpatient visits associated with an ARI/CLI or other acute illness associated with COVID-19 will be defined using ICD-10 diagnosis codes.
- Laboratory-confirmed infection diagnosis will be determined by examining clinical laboratory testing that was conducted up to 14 days prior to, during, and up to 3 days after the index medical event.
- Re-admissions that occur up to 14 days from hospital discharge may be considered part of the same medical event.
- Severe patient outcomes among persons hospitalized with laboratory-confirmed SARS-CoV-2 and influenza-associated respiratory hospitalization will include ICU admission, invasive mechanical ventilation, in-hospital death.
- Complications that occur during a respiratory hospitalization will initially be defined using ICD coded discharge diagnoses. Diagnoses will be initially aggregated into systems for acute respiratory, acute renal, acute neurological, acute cardiovascular, and acute inflammatory complications. Complications may be further refined using specific ICD codes, with the possibility of also using clinical laboratory data, vital signs, medications, and other interventions. These refinements will be discussed in a working group setting and will be implemented after network consensus.
- Healthcare encounters with ICD codes for outcomes of interest that occur within the 12 months following discharge date of a respiratory hospitalization or any COVID-19 tested hospitalization. The date, setting, and diagnostic codes from these encounters will enable analysis of complications, sequelae, frailty, and increased healthcare utilization after a SARS-CoV-2 infection. The time frame for capturing data on

potential sequelae may be refined and ICD codes for outcomes of interest will be identified, which will be discussed in a working group setting and will be implemented after network consensus.

- The criteria described above can be applied to laboratory-confirmed influenza virus infections.

## **5.6 Summary of data elements for primary and secondary objectives**

In addition to the data elements being captured in the codebook (Appendix A), additional data elements being pursued to meet the primary and secondary objectives are listed below.

Elements with asterisk (\*) will be further defined through working groups and network consensus:

### **1. All respiratory hospitalizations during the observation period:**

- LOS in general hospital ward and LOS in ICU (as determined by time and dates of admission, transfer, and discharge events)
- Characteristics of the facility (e.g., facility type, facility ownership, trauma level, urban vs. rural location, number of beds, teaching vs. not)
- Non-invasive respiratory support\*;
- Where patient is admitted from (home, long-term care facility, etc.)\*;
- Discharge disposition\*;
- Vital signs\*;
- Clinical laboratory values\*;
- Medications\*.
- Geographic clustering

2. All healthcare encounters that occur after respiratory hospital discharge\*:

- Encounter setting;
- Encounter date; and
- Diagnosis codes.

## **6.0 Statistical Analysis**

### **6.1 Analysis Plan**

The proposed analytic plan is subject to review and revision by key stakeholders, including CDC, Westat, and collaborating partners. Methods will be amended as necessary.

#### ***Rates***

Rates will be estimated for:

- Hospitalizations associated with laboratory-confirmed SARS-CoV-2 infection between February 2020 and the end of the observation period
- Hospitalizations associated with laboratory-confirmed influenza virus infection during periods of local influenza circulation

The denominator will be the cumulative at-risk person-time contribution for individuals who meet the cohort definition. Sites with well-characterized source populations are anticipated to enumerate this person-time and estimate these hospitalizations rates, though sites, in consultation with CDC and Westat, may decide whether to contribute to the rate estimation. The operationalized definition of at-risk person-time will be discussed and agreed upon with the sites who elect to estimate rates, CDC, and Westat. Hospitalization rates will be estimated by socio-demographic and high-risk groups, including age, race and ethnicity, and underlying conditions.

These methods for estimating hospitalization rates may be modified to estimate other rates of interest (e.g., rates of ED visits, rates of acute respiratory illness inpatient or outpatient encounters), as feasible.

***Patterns of testing and care-seeking to inform national and local disease burden models***

- Describe the frequency of clinical testing for SARS-CoV-2 (and for influenza separately) among encounters for ARI and other acute illness associated with COVID-19 by setting (e.g., inpatient, ED), age group, high-risk status, COVID-19 vaccination, current seasonal influenza vaccination, encounter outcomes, prior healthcare utilization, and timing of the encounter within the pandemic
  - This will be calculated by socio-demographic and high-risk groups
- Estimate the testing rate for SARS-CoV-2 and separately for influenza (and then also restricted to PCR testing)
  - This will be calculated by socio-demographic and high-risk groups
- Describe the frequency of clinical testing for SARS-CoV-2 and for influenza in the cohort during the observation period by setting, age group, presence of underlying conditions, prior healthcare utilization, and timing of encounter within the pandemic
- Assess patient and clinical characteristics and predictors of SARS-CoV-2 testing and influenza testing using bivariate tests of association and/or multivariate regression
  - Repeat this analysis but focusing on SARS-CoV-2 positivity (compared to SARS-CoV-2 negativity among persons tested) and similarly for influenza positivity
  - Assess separately for hospitalizations and ED encounters associated with an ARI/CLI and other acute illness associated with COVID-19



- Compare the distribution of settings for seeking care (e.g., telemedicine, ambulatory, urgent care, ED) among persons who meet the case definition for ARI/CLI and other acute illness associated with COVID-19 before and during the COVID-19 pandemic.

These analyses will be useful in evaluating potential biases associated with clinical testing for SARS-CoV-2 and influenza, as a component of estimating COVID-19 and influenza vaccine effectiveness.

### ***COVID-19 vaccine effectiveness (VE) sample size considerations***

For sample size or observation number considerations, analysis focused on the primary objective of VE against COVID-19-associated hospitalizations. Sites that will be contributing to this effort include all of the sites mentioned above, in addition to three non-Westat sites. The combined effort will involve up to eleven sites sending electronic medical data from hospital visits of patients with ARI/CLI or other acute illness associated with COVID-19, including COVID-19 vaccination status and laboratory testing results. The sites will be expected to send updated data every two weeks; thus, observation needs were estimated considering this bi-weekly schedule. The aim of this analysis is to determine the number of weeks required to accumulate sufficient data to achieve 80% power to detect a vaccine effectiveness of 60% using a minimally adjusted case control model. The model was simulated by generating data for the average weekly acute respiratory illness (ARI) visits for each site by age group and with assumptions for overall VE and VE by age groups.

The number of subjects is estimated by using the mean/median ARI/CLIs seen weekly over the course of a year as reported by each potential site or platform.

Table. Anticipated sites and number of ARI/CLIs by age group

Site or Platform	Median/ Mean Weekly Hospital Admissions for ARI/CLI Over 1 Year		
	Ages 18-49	Ages 50-64	Ages ≥65
Site 1	55	75	279
Site 2	93	129	378
Site 3	80	92	115
Site 4	7	9	23
Site 5	6	9	37
Site 6	22	28	62
Site 7	171	175	266
Site 8	152	195	346
Site 9	19	25	93
Site 10	81	90	213
Site 11	95	41	88

We assumed that sites will test between 25% and 50% of ARI/CLI patients for COVID-19 by RT-PCR (or by rapid antigen tests confirmed by PCR). We assumed between 4% and 10% will test positive. We simulated scenarios where vaccination status can be confirmed for between one third and two thirds of patients. We assumed vaccine uptake and specifically completion of the 2-dose regimen will vary by age group and increase over-time. We assumed that the analytic period would begin when adults under age 65 years will have 5-8% 2-dose vaccination coverage

and achieve 35-40% vaccination by the end of the year; adults over age 65 were assumed to start at 10-15% 2-dose vaccinated and achieve 65-70% vaccination by the end of the year.

Table. Simulation assumptions for each site

EMR Site	% Tested	% Positive	% Vaccine Status known	Age Group	Vaccine coverage rates by quarter
1	25%	4%	33%	<65	(5%, 10%, 20%, 35%)
				≥65	(10%, 20%, 35%, 65%)
2	50%	4%	66%	<65	(8%, 13%, 25%, 40%)
				≥65	(15%, 25%, 40%, 70%)
3	25%	5%	66%	<65	(8%, 13%, 25%, 40%)
				≥65	(15%, 25%, 40%, 70%)
4	50%	5%	66%	<65	(5%, 10%, 20%, 35%)
				≥65	(10%, 20%, 35%, 65%)
5	25%	6%	33%	<65	(5%, 10%, 20%, 35%)
				≥65	(10%, 20%, 35%, 65%)
6	50%	6%	33%	<65	(8%, 13%, 25%, 40%)
				≥65	(15%, 25%, 40%, 70%)
7	25%	8%	33%	<65	(8%, 13%, 25%, 40%)
				≥65	(15%, 25%, 40%, 70%)
8	50%	8%	33%	<65	(5%, 10%, 20%, 35%)
				≥65	(10%, 20%, 35%, 65%)

9	25%	9%	66%	<65	(5%, 10%, 20%, 35%)
				≥65	(10%, 20%, 35%, 65%)
10	50%	9%	33%	<65	(8%, 13%, 25%, 40%)
				≥65	(15%, 25%, 40%, 70%)
11	25%	10%	66%	<65	(8%, 13%, 25%, 40%)
				≥65	(15%, 25%, 40%, 70%)

Field VE for COVID-19 vaccines are assumed to be between 60% and 75%.

Simulated subjects were created for each site and age group according to the numbers in the above Table, assuming two weeks of accrual, and randomly assigned to a testing status, positivity status, and 2-dose vaccination status based on the assumptions in the above Table. Subjects assigned to be 2-dose vaccinated had their odds of test positivity adjusted by the assumed VE. Odds of being COVID-19 positive was analyzed using a marginal logistic model with vaccination as a main effect and clustering by site.

The simulation was repeated 1000 times and power calculated. This was repeated, accruing an additional two weeks of data each time, until 80% power was achieved. The following VE’s were tested at both 60 and 75%.

1. Overall VE adjusted for group (18-49, 50-64, ≥65)
2. VE by age group, dichotomous (18-64, ≥65)
3. VE by age group, multinomial (18-49, 50-64, ≥65)

The results of the simulation were as follows:

1. To achieve 80% power to detect a VE of 60% or 75% for all adults adjusted by age group, we would need 4-6 weeks of data across sites.
2. To achieve 80% power to detect a VE of 60% to 75% for adults over age 65 years, we would need 6-8 weeks of data. For a VE strata of adults under age 65 years, we would need 14 weeks.
3. To achieve 80% power to detect a VE of 60% to 75% for adults aged 18-49 or 50-64 years, we would need 6 months of data.

### ***COVID-19 VE estimation models***

The primary statistical model for estimating COVID-19 VE will be the test-negative design (TND), whereby VE equals  $1 - \text{odds ratio} [\text{ratio of odds of vaccination among COVID-19-positive cases to the odds of vaccination among COVID-19-negative controls}] \times 100\%$  using logistic regression. The TND has been used extensively to estimate VE against medically attended influenza virus illness and is believed to minimize biases associated with access to vaccines and healthcare seeking.<sup>25-26</sup>

COVID-19 VE will be estimated using multivariate logistic regression with laboratory-confirmed SARS-CoV-2 infection as the outcome and COVID-19 vaccination status as the exposure of interest. A set of covariates will be included in the model to adjust for potential confounding. A minimally adjusted model will include *a priori* determined covariates that are expected to be associated with both the likelihood of COVID-19 vaccination (and complete vs. partial vaccination [based on ACIP criteria]) and with the likelihood of COVID-19 positivity. A fully adjusted model will be estimated using a propensity score modeling approach. VE is calculated as a function of the adjusted odds ratio (aOR) where  $VE=1-aOR$ .

For the fully adjusted model, differences in vaccinated and unvaccinated subjects will be balanced for each site using inverse propensity score weighting. Propensity to vaccinate will be estimated with boosted regression trees and used to calculate average treatment effect (ATE) weights.<sup>27</sup> Variables to be used in predicting vaccination status include demographics (e.g. age, sex, race), high-risk medical conditions (e.g. lung disease, heart disease, immunosuppression), healthcare utilization patterns, vaccination history, and other exposures of potential importance. Propensities and weights will be calculated using R package *twang*<sup>28</sup> or comparable software.

To estimate overall VE using fully adjusted models across sites, heterogeneity across site-specific VEs will be assessed using the Q and I<sup>2</sup> statistics. If heterogeneity is found to be substantial, only site-specific VEs will be reported. Otherwise, all site data will be pooled and an overall VE will be estimated using either a mixed-effects or generalized estimating equation model to account for correlation among observations within sites.

To address study objectives, VE models will be stratified by socio-demographic, health, and other risk groups, as data becomes available. VE models focused on ED or UC outcomes and models focused on children will apply the same methodology.

For the sites with well-characterized membership or source population data is available and thus the study population's person-time at risk can be estimated, COVID-19 VE will be calculated using survival analysis framework. The hazard ratio (HR) will be fit using calendar time to account for the exact calendar date of each COVID-19 case. Individuals may go from unvaccinated to vaccinated, and thus may contribute both unvaccinated and vaccinated person-time at risk. To allow for the vaccination status to be time-varying and to model possible re-infection, the Anderson and Gill (AG) extension of the Cox Proportional Hazard (PH) model will be used to estimate VE. Robust standard errors will be calculated using a sandwich covariance

matrix to predict covariance among observations. COVID-19 VE will be calculated by estimating the hazard ratio of lab confirmed SARS-CoV-19 positivity among vaccinees and non-vaccinees;  $VE(\%)=(1-HR)\times 100$ .

## **7.0 Data Sources and Management**

### **7.1 Data Sources**

Many of the required patient data variables are routinely captured in the EHR. Others may need to be added from additional sources such as participant health plan enrollment or administrative data, administrative claims, or vaccine registries linked to EHR data at the patient level. Reliance on open text fields, such as physician notes, will be kept to a minimum.

Each data element that is extracted will have an operational definition accounting for the coding structure, completeness, and limitations of the data source. This operational definition is particularly important to address situations that may arise when extracting data from health records. For example, a concept (e.g. vaccination status) may be captured by more than one variable in a single provider database (e.g., recorded in vaccination records or in visit notes for self-reported) or in more than one database (e.g., EHR or vaccine registry). Thus the validity of the analyses will depend on consistent operational definitions.

Operational definitions for necessary data elements will be defined collaboratively by CDC, Westat, and network sites to ensure consistency and accuracy of definitions across sites; this approach will also be applied to building patient cohorts using look-back periods or among sites with well-characterize source populations. Sites may choose to derive some or most data elements prior to submitting final datasets to Westat/CDC, based on the agreed-upon operational

definitions. Alternatively, sites may submit their data to Westat, the data coordinating center, and Westat would derive variables for analysis.

As the data coordinating center, Westat will receive data from the sites. At Westat, the study database will be housed in an integrated central research data warehouse (RDW) platform. The RDW will be fully integrated with the data management and tracking systems necessary for carrying out the processes associated with entry/upload, transmission, QA, version control, standardization, storage, and security.

## **7.2 Variables**

The codebook is included as Appendix A. Depending on feasibility and prioritization, additional variables will be added, including medications (antivirals, antibiotics, steroids, vasopressors), non-invasive mechanical ventilation, vital signs (such as heart rate, respiratory rate, blood pressure, temperature, O<sub>2</sub> saturation, Glasgow Coma Scale), and other laboratory tests [including diabetes, hemoglobin A1C, complete blood count (CBC), comprehensive metabolic panel (CMP), etc.].

## **7.3 Data Management**

During the course of the study, research staff at each site will extract information from the EHR of participants, state or local vaccination registries, and billing records, as available. Information for extraction includes data elements described in Section 7.2 and Appendices A and B. Sites will perform data validation on the extracted data as discussed in Section 7.4, below. Each site will create and maintain a database onsite that links demographic and clinical information extracted from the EHR, state or local vaccination registries, and billing records to a coded patient identifier. The key linking the coded identifier to the patient IDs will be kept at the individual sites. Additionally, the sites within a health network will also be given a coded



identifier. Some details of the site will be included in the data, such as facility type and location, but no attempts will be made to identify facilities. Upon execution of an appropriate data use agreement (DUA) or data transfer agreement (DTA), a HIPAA-defined limited data set (LDS) will be forwarded (using a secured data transfer protocol) to the study coordinating center (Westat). Data transferred to Westat will not include identifying elements such as name, medical record number, postal address, or any other elements not allowed in a LDS. These datasets will contain individual-level records of pre-processed variables derived by the study site for the analysis necessary for the primary objectives.

Additional details about site data processing can be found in Appendix B, Proposed Data Structure to be shared by Sites with Westat. Westat will perform additional quality control and validation on the received data, as described in Section 7.4 below, and will concatenate the files to form one database. The database will be transferred to CDC using a secured transfer protocol for analysis (see Data Management Plan, Appendix C, for further detail).

#### **7.4 Data Validation**

Data will be validated at different points throughout the protocol and study period and will be an iterative process. If errors are found, Westat will coordinate with the site and the extraction will be re-programmed and data re-extracted and re-validated. Sites and Westat will both perform various aspects of this quality control and data validation. There are three main types of validation that will be conducted for this study:

1. Basic validation

The basic validation includes data quality checks such as confirmation that values are non-missing, values are of the correct type and length, and values are in the appropriate range. All variables will undergo basic validation. Expected variable type, length, and range will be

included in the codebooks. Range checks will be particularly important for dates and healthcare utilization. Expected percent missing will be determined at the site-level. For example, some sites may do more SARS-CoV-2 testing than others and thus the expected percent missing for those sites would be different.

## 2. Internal crosschecks

The internal crosscheck validation will include two types of checks and may be performed by Westat. The first is comparing calculated proportions of various data elements in the extracted data to the proportions in the source data. For example, if a site generally performs SARS-CoV-2 testing on 50% of its adult patients and the extracted data show only 5% that would indicate an error. Variables that will undergo this check are sex, race, ethnicity, insurance coverage, insurance type, respiratory virus testing and results (percent tested and percent positive), vaccination status (by age and high-risk condition), invasive mechanical ventilation, ICU admission, in-hospital mortality, and high-risk conditions.

The second type of internal crosscheck is comparing related data elements, i.e., a value for one data element is checked against the value for another data element. For example, all patients with a pregnancy-related diagnosis code should have sex recorded as female. Checks of this type could include:

- Number of inpatient visits corresponds to number of admission dates
- Number of outpatient visits corresponds to number of dates of outpatient encounters
- If date of vaccination is non-missing then source of data (EHR, registry, administrative records, etc.) should be non-missing

- If the date of respiratory virus testing is non-missing than the type of test and the test result should be non-missing; likewise, the test result should be consistent with what is detectable by the test type
- A patient with a record for mechanical ventilation should also have an ICU admission date

Additionally, sites could crosscheck the high-risk conditions with other data not extracted for this study. For example, they could look at medication use among those coded with high-risk conditions commonly treated with medication (e.g. asthma, diabetes).

### 3. Comparisons with external data sources

Comparisons with external data sources includes comparing values to those recorded in the primary records as well as comparing rates, proportions, and distributions of variables across participating sites and to national, regional, or state data.

Comparison of extracted data to the primary records will be done on a limited basis, but can be informative for understanding flow of data into the data warehouse and appropriateness of structured data fields to the data of interest for this project. CDC, Westat, and sites will reach consensus regarding appropriate methods and requirements for each site's data validation against primary records.

Variables for which rates, proportions, and distributions can be compared across participating sites and to national, regional, or state data include:

- SARS-CoV-2 and influenza vaccination coverage
- Respiratory virus testing and positivity
- Mechanical ventilation
- ICU admissions

- Proportion of patients with high-risk conditions

Westat will work with each site to review potential errors and decide on corrective action.

In a few cases, sites may need to perform limited review of medical charts to clarify data elements flagged for review.

## **8.0 Ethical Considerations for Protection of Human Research Subjects**

### **8.1 Institutional Review Board Review**

Westat will serve as the single IRB of Record for this study for all participating sites and coordinating center for overseeing protections of human subjects research (45 C.F.R. § 46.114).

The Westat IRB will enter into an IRB Authorization Agreement (IAA) that will include a communication plan with each institution prior to study implementation. IAAs and other documentation necessary in order to document compliance with the single IRB policy are maintained by Westat's IRB. Westat's IRB will use several mechanisms to communicate with sites, including email, phone calls and direct person-to-person communications as needed.

The protocol, data collection instruments, and other documents associated with the protocol shall be approved by Westat's IRB in compliance with all applicable laws, including 45 CFR 46. Subsequently, the protocol and related documents must be re-reviewed at least annually. Westat is responsible for preparation and submission of all documents and periodic reports required by the IRB and may seek input from sites regarding local implementation.

### **8.2 Patient Confidentiality**

All patients in the dataset will be assigned a linkable patient identification code (i.e. study identification code). Sites will be responsible for assigning and maintaining the link between the patient's identifying information and study ID. Documents maintaining this link will never be transferred to the coordinating center or study investigators. Personal identifiers (patient's name,

address, medical record number, and encounter number) will exist at the participating site, as part of the hospital administrative data but will be replaced by a random generated code (linkable patient identification code), which will allow linkage of data without CDC or the coordinating center (Westat) having any access to these personal identifiers. All study data and administrative documentation will be identified by the study identification code only, to maintain participant confidentiality. Limited datasets will be created for the study; the study will comply with each institution's human subjects, privacy, and information security laws, if any. All study data files will be stored separately from any study records that contain names or other personal identifiers. All local databases must be secured with password protected access systems.

Listings that link study (and personal) IDs to other identifying information must be stored in a separate, locked file (or encrypted) in an area with limited access (or maintained in a directory separate from any study specific data files/sets) at each participating facility. Links between the study identification codes and personal identifiers will be destroyed by the participating site after publication of the findings (for additional information please see the Data Management Plan, Appendix C).

Westat's Data Management Plan (Appendix C) details how Westat will protect any identifiers from improper use or disclosure, how Westat will destroy the identifiers after study completion, and how the protected health information will not be reused for other research.

### **8.3 Request for Waiver of Informed Consent**

The study relies on existing data already collected as part of patient's routine care or for billing purposes. No supplemental data collection will be done as part of this study. In addition, this study presents minimal risk to participants because there is no interaction or intervention with patients; therefore, a waiver of informed consent is requested. Minimal risk includes

disclosure of clinical information on the patients' medical condition to persons outside of this protocol's defined study. Although patient information already available in the administrative databases will be collected, only information associated with a HIPAA limited dataset will be collected for the study. There is no risk to the participants' health from participation nor any impact on patients' current health care or therapeutic management plan because patients will not be contacted at any time. Consequently, patients will not be provided information about their participation.

Additionally, it will be impractical to conduct this study without waiving informed consent. By the time access to the datasets is available, most of the patients, if not all, will be out of the hospital (or some may have died during hospitalization), and the vast majority may have been hospitalized many years prior and may no longer live in the area or receive their care at the relevant study site. To contact each patient in this large, retrospective study for informed consent or to notify them of study results would place an insurmountable burden on investigative staff and would prohibit successful completion of the study.

#### **8.4 Benefits to Participants**

There are no direct benefits for patients whose data contribute to this study. There may be future indirect benefits to the populations of the participating health systems, especially those with risk factors for severe illness from SARS-CoV-2 and influenza. For example, information from this study may influence vaccination strategies for high-risk groups that may improve future outcomes for children, elderly adults, and those with high-risk conditions. In addition, understanding factors that make these groups at higher risk for SARS-CoV-2 and influenza-associated hospitalization may help in developing and improving prevention and treatment guidelines.

## **8.5 Data Records Lifecycle and Destruction**

Each participating study institution will store any paper study records in a physically secure location that is only accessible to authorized study staff. At close out, assuming no restrictions on data retention, the de-identified analytic files, documentation and all code used in processing will be archived in a way that would allow replication of the results. Each organization affiliated with the study, through subcontract or otherwise, must destroy data according to contractual specifications and must provide Westat with a certificate of destruction.

A certificate of destruction will be required for electronic and hard copy data and each must detail the type of data destroyed, how and when it was destroyed, and the signature of the authorized data security manager or corporate executive. In addition, any exceptions to the data destruction, e.g., data that must be maintained for internal records, must be identified in the certificate of destruction, along with a detailed rationale for why the data were not/could not be destroyed, at study conclusion.

## **8.6 Guidance for Decision Making**

A project steering committee will provide high-level input into this project, with CDC and Westat having final approval and sign off on any decisions made. The project steering committee will consist of two individuals from each site (decided by the site), a Westat representative, and a CDC representative. Each site will have one vote. The day-to-day overall project management will occur through the Westat study lead who will interface directly with CDC; however, the steering committee will be consulted on over-arching project issues including final protocol decisions, adjudicating any protocol deviations that might occur, reviewing and confirming analysis plans, and making final decisions on analyses, manuscripts, and authorship as needed. Upon the completion of all study deliverables, at a minimum, aggregate tables from publications

of this collaboration will be publicly shared as specified in U.S. Government Data Sharing guidelines. Additional data may be publicly shared to further satisfy the U.S. Government Data Sharing guidelines, as determined by consensus of steering committee members and per site data use agreements.



## **Appendix A: Specification of Data Elements**

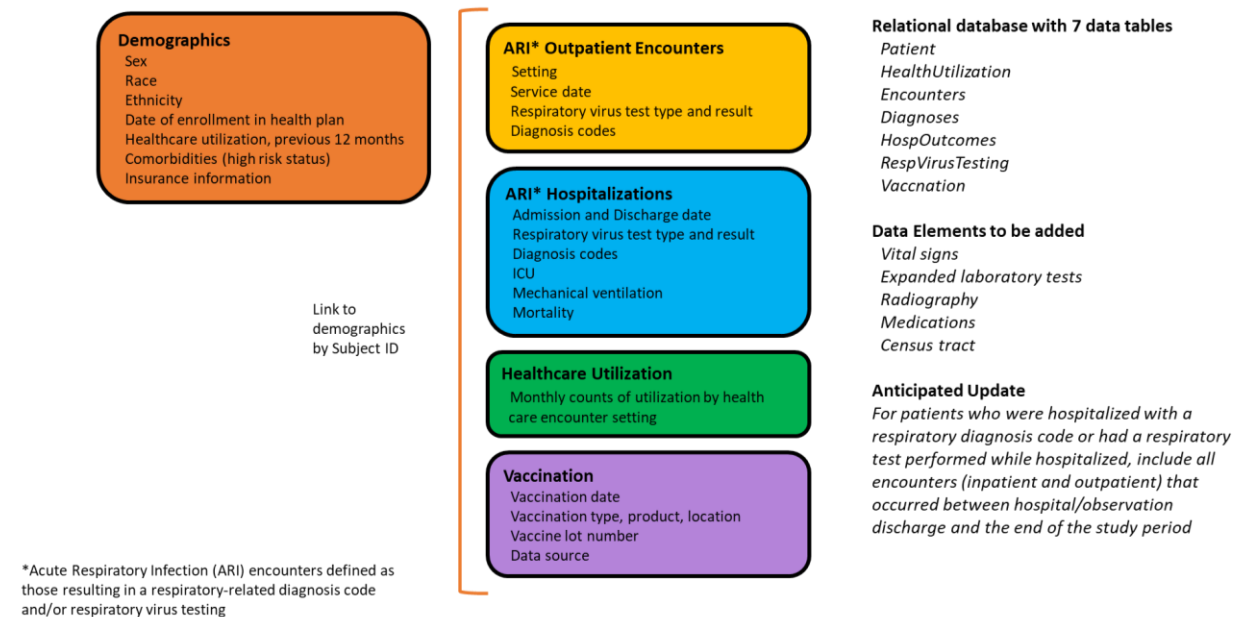
See accompanying Excel file “COVIDDataElementsCodebook.”

## Appendix B: Proposed Data Structure to be Shared by Sites with Westat

Based on conversations with participating collaborating sites, study sites would prefer to pre-process individual-level records related to the COVID network and then share these pre-processed datasets with Westat & CDC. The proposed data structure would be one data package for each COVID data delivery. Initially, sites will be asked to submit data on a monthly basis, until a threshold is reached for 2-dose COVID-19 vaccination, at which points site would increase the frequency of data transmission to a bi-weekly basis.

Each data package would include data as specified in Appendix A. Sites would be responsible for deriving data elements using the definitions from Appendix A and doing so prior to sending data to Westat.

Relational Design: Preprocessed (by the study site) individual-level records shared with Westat and CDC



## **Appendix C: Data Management Plan**

See accompanying document “Data Management Plan.”

## References

- <sup>1</sup> COVID-19 Hospitalizations (cdc.gov), as of November 21, 2020.
- <sup>2</sup> <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/pdf/covidview-05-29-2020.pdf>
- <sup>3</sup> Garg, S., Kim, L., Whitaker, M., O'Halloran, A., et al. (2020). Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. *MMWR. Morbidity and mortality weekly report*, 69(15), 458–464. <https://doi.org/10.15585/mmwr.mm6915e3>
- <sup>4</sup> Azar, K., Shen, Z., Romanelli, R. J., Lockhart, S. H., et al. (2020). Disparities In Outcomes Among COVID-19 Patients In A Large Health Care System In California. *Health affairs (Project Hope)*, 101377hlthaff202000598. Advance online publication. <https://doi.org/10.1377/hlthaff.2020.00598>
- <sup>5</sup> Garg, S., Kim, L., Whitaker, M., O'Halloran, A., et al. (2020). Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 - COVID-NET, 14 States, March 1-30, 2020. *MMWR. Morbidity and mortality weekly report*, 69(15), 458–464. <https://doi.org/10.15585/mmwr.mm6915e3>
- <sup>6</sup> Nair, H., Brooks, W. A., Katz, M., Roca, A., et al. . (2011). Global burden of respiratory infections due to seasonal influenza in young children: a systematic review and meta-analysis. *Lancet (London, England)*, 378(9807), 1917–1930. [https://doi.org/10.1016/S0140-6736\(11\)61051-9](https://doi.org/10.1016/S0140-6736(11)61051-9); <https://www.cdc.gov/flu/about/burden/2018-2019.html>
- <sup>7</sup> Ludvigsson J. F. (2020). Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. *Acta paediatrica (Oslo, Norway : 1992)*, 109(6), 1088–1095. <https://doi.org/10.1111/apa.15270>

- <sup>8</sup> Shekerdemian, L. S., Mahmood, N. R., Wolfe, K. K., et al. (2020). Characteristics and outcomes of children with coronavirus disease 2019 (COVID-19) infection admitted to US and Canadian pediatric intensive care units. *JAMA Pediatrics*. Published online May 11, 2020. doi:10.1001/jamapediatrics.2020.1948.
- <sup>9</sup> Itaya T, Furuse Y, Jindai K. (2020). Does COVID-19 infection impact on the trend of seasonal influenza infection? 11 countries and regions, from 2014 to 2020, *International Journal of Infectious Diseases*. doi: <https://doi.org/10.1016/j.ijid.2020.05.088>
- <sup>10</sup> <https://www.cdc.gov/flu/about/burden/preliminary-in-season-estimates.htm>
- <sup>11</sup> Gupta, A., Madhavan, M.V., Sehgal, K. et al. (2020). Extrapulmonary manifestations of COVID-19. *Nat Med* **26**, 1017–1032. <https://doi.org/10.1038/s41591-020-0968-3>
- <sup>12</sup> Gounder AP, Boon ACM. (2019). Influenza Pathogenesis: The Effect of Host Factors on Severity of Disease. *J Immunol*. 202(2):341-350.
- <sup>13</sup> Herold S, Becker C, Ridge KM, Budinger GR. (2015). Influenza virus-induced lung injury: pathogenesis and implications for treatment. *Eur Respir J*. 45(5):1463-1478
- <sup>14</sup> Short KR, Kroeze E, Fouchier RAM, Kuiken T. (2014). Pathogenesis of influenza-induced acute respiratory distress syndrome. *Lancet Infect Dis*. 14(1):57-69
- <sup>15</sup> Sellers SA, Hagan RS, Hayden FG, Fischer WA, 2nd. (2017). The hidden burden of influenza: A review of the extra-pulmonary complications of influenza infection. *Influenza Other Respir Viruses*. 11(5):372-393.
- <sup>16</sup> Thompson MG, Pierse N, Sue Huang Q, et al. (2018). Influenza vaccine effectiveness in preventing influenza-associated intensive care admissions and attenuating severe disease among adults in New Zealand 2012-2015. *Vaccine*. 36(39):5916-5925.
- <sup>17</sup> Feng S, Cowling BJ, Sullivan SG. (2016), Influenza vaccine effectiveness by test negative

design – comparison of inpatient and outpatient settings. *Vaccine* 34(14):1672–9.

<sup>18</sup> Folegatti PM, Ewer KJ, Aley PK, Angus B, et al. (2020). Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *The Lancet* 396(10249): 467-478

[https://doi.org/10.1016/S0140-6736\(20\)31604-4](https://doi.org/10.1016/S0140-6736(20)31604-4).

<sup>19</sup> Zhu F, Guan X, Li Y, Huang J, et al. (2020). Immunogenicity and safety of a recombinant adenovirus type-5-vectored COVID-19 vaccine in healthy adults aged 18 years or older: a randomised, double-blind, placebo-controlled, phase 2 trial. *The Lancet* 396 (10249): 479-488

[https://doi.org/10.1016/S0140-6736\(20\)31605-6](https://doi.org/10.1016/S0140-6736(20)31605-6)

<sup>20</sup> COVID-19 Vaccine U.S. Distribution Fact Sheet | Pfizer

<sup>21</sup> Department of Health and Human Services, Office of Inspector General. Vaccine for children program: vulnerabilities in vaccine management. June 2012.

Vaccines for Children Program: Vulnerabilities in Vaccine Management (OEI-04-10-00430; 06/12) (hhs.gov)

<sup>22</sup> Dror AA, Eisenbach N, Taiber S, Morozov NG, et al. (2020). Vaccine hesitancy: the next challenge in the fight against COVID-19. *Eur J Epidemiol.* Aug;35(8):775-779. doi: 10.1007/s10654-020-00671-y. Epub 2020 Aug 12. PMID: 32785815.

<sup>23</sup> COCONEL Group. A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. *Lancet Infect Dis.* 2020 Jul;20(7):769-770. doi: 10.1016/S1473-3099(20)30426-6. Epub 2020 May 20. PMID: 32445713; PMCID: PMC7239623.

<sup>24</sup> Irving SA, Donahue JG, Shay DK, et al. (2009). Evaluation of self-reported and registry-based influenza vaccination status in a Wisconsin cohort. *Vaccine.* Nov 5;27(47):6546-9. doi: 10.1016/j.vaccine.2009.08.050. Epub 2009 Sep 1. PMID: 19729083.

<sup>25</sup> Foppa IM, Haber M, Ferdinands JM, Shay DK. (2013). The case test-negative design for studies of the effectiveness of influenza vaccine. *Vaccine*. 31:3104–9.

<sup>26</sup> Jackson ML, Nelson JC. (2013). The test-negative design for estimating influenza vaccine effectiveness. *Vaccine*. 31:2165–8.

<sup>27</sup> McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, 9(4), 403.

<sup>28</sup> Ridgeway, G., McCaffrey, D., Morral, A., Burgette, L., & Griffin, B. A. (2017). *Toolkit for Weighting and Analysis of Nonequivalent Groups: A tutorial for the twang package*. Santa Monica, CA: RAND Corporation.