

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

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

**DRAFT vs 3.0: 7/1/2020**



# Table of Contents

|            |  |    |
|------------|--|----|
| <b>1.0</b> | <b>Study Summary</b> .....   | 4  |
| 1.1        | Title of Study .....   | 4  |
| 1.2        | Investigators .....  | 4  |
| 1.3        | Protocol Rationale.....  | 5  |
| 1.4        | Protocol Objectives .....  | 6  |
| 1.5        | Methods Summary .....  | 7  |
| <b>2.0</b> | <b>Abbreviations</b> .....   | 9  |
| <b>3.0</b> | <b>Background Information</b> .....                                  | 9  |
| 3.1        | Scientific Background.....   | 10 |
| 3.2        | Rationale and Justification.....                                     | 12 |
| 3.3        | Expected Benefits .....  | 13 |
| <b>4.0</b> | <b>Study Objectives</b> .....  | 14 |
| 4.1        | Primary Objectives .....   | 14 |
| 4.2        | Secondary Objectives.....  | 17 |
| <b>5.0</b> | <b>Methods</b> .....   | 18 |
| 5.1        | Study Design .....   | 18 |
| 5.2        | Participant Eligibility .....  | 20 |
| 5.3        | Observation Time during Influenza Season .....                       | 22 |
| 5.4        | Exposures of Interest .....  | 22 |
| 5.5        | Outcomes of Interest.....  | 27 |
| 5.6        | Summary of Data Elements for Primary Objectives .....                | 28 |
| 5.7        | Summary of Methods for Secondary Objectives.....                     | 31 |
| <b>6.0</b> | <b>Statistical Analysis</b> .....                                    | 31 |
| 6.1        | Analysis Plan .....  | 31 |
| 6.2        | Sample Size.....   | 38 |
| 6.3        | Enhanced Data Collection Elements to Meet Secondary Objectives ..... | 39 |
| <b>7.0</b> | <b>Data Sources and Management</b> .....                             | 40 |
| 7.1        | Data Sources .....   | 40 |
| 7.2        | Variables .....  | 41 |
| 7.3        | Data Management.....   | 42 |
| 7.4        | Data Validation .....  | 43 |

|            |   |    |
|------------|---|----|
| <b>8.0</b> | <b>Ethical Considerations for Protection of Human Research Subjects</b> .....   | 45 |
| 8.1        | Institutional Review Board Review .....   | 45 |
| 8.2        | Patient Confidentiality.....  | 46 |
| 8.3        | Request for Wavier of Informed Consent.....   | 47 |
| 8.4        | Benefits to Participants.....   | 48 |
| 8.5        | Data Records Lifecycle and Destruction.....   | 48 |
| 8.6        | Guidance for Decision Making .....  | 49 |
|            | <b>Appendix A: Outcomes, Exposures, and Modifiers for Study Objectives</b> .....  | 50 |
|            | <b>Appendix B: Specification of Data Elements</b> .....   | 54 |
|            | <b>Appendix C: Assessment of inclusion criteria</b> .....   | 55 |
|            | <b>Appendix D: Diagnosis Codes and Other Variables Used to Define for High Risk Conditions, ARI Hospitalizations, and Study Outcome</b> ..... | 57 |
|            | <b>Appendix E: Proposed Data Structure to be Shared by Sites with Westat</b> .....  | 62 |
|            | <b>Appendix F: Data Management Plan</b> .....   | 64 |
|            | <b>Appendix G: Suggestions for Site Data Validations using Primary Health Records</b> .....   | 65 |
|            | <b>References</b> .....   | 68 |

## **1.0 Study Summary**

### **1.1 Title of Study**

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

### **1.2 Investigators**

*Centers for Disease Control and Prevention*

- Melissa Rolfes
- Shikha Garg
- Carrie Reed
- Andrea Steffens
- Eduardo Azziz-Baumgartner
- Lenee Blanton
- Catherine Bozio
- Manish Patel
- Mark Thompson
- Alicia Fry
- Nia Mims

*Westat*

- Sarah Ball
- Patricia Shifflett
- Howard Fishbein
- Rebecca Birch

- Rebecca Fink
- Saloni Sapru
- Beth Mittl
- Keith Umbel
- Sarah Bevan
- Carly Hallowell

*Participating sites*

- HealthPartners Institute
- Kaiser Permanente Northwest
- Universities of California at Irvine, Los Angeles, and San Diego, San Francisco (COVID addendum only), and Davis (COVID addendum only)
- University of Colorado, Denver

### **1.3 Protocol Rationale**

This project will create new collaborations among multiple health systems to pilot and understand the use and limitations of existing medical, laboratory, pharmaceutical, and vaccination records for conducting research on influenza. Once established, the network of collaborating sites can address gaps in our understanding of the risk of severe influenza virus infection, its complications, and the effectiveness of influenza vaccination. Furthermore, the joint information captured across the participating sites will enable us to explore and understand how the risk of influenza and the effectiveness of influenza vaccination may differ among people of different ages and among those with underlying medical conditions.

The COVID-19 associated addendum to this protocol (see Appendix G) will extend the current network and associated research data platform (RDP) to capture the data needed to both

retrospectively and prospectively examine some of the most difficult-to-study populations, who may experience some of the most extreme outcomes associated with COVID-19. This addendum will apply our established “virtual network” approach and refine our current methodology to address multiple research questions regarding severe SARS-CoV-2-associated outcomes among high-risk groups to better understand the disease burden of SARS-COV-2 infection, its complications and the trajectory of illness in those with severe COVID-19 illness. Furthermore, building on the current influenza platform, the joint information captured across the participating sites will enable us to explore and understand the impact that circulating SARS-CoV-2 might have on influenza illness and associated vaccination coverage.

#### **1.4 Protocol Objectives**

##### *Network objectives*

We aim to develop a multi-site collaboration to better utilize existing medical and public health records to address key research questions for influenza prevention and control. We will build retrospective cohorts of individuals, identify individuals within the cohorts that are at high risk of influenza complications, determine whether individuals within the cohorts have been vaccinated for influenza, and enumerate laboratory-confirmed influenza-associated events. This effort consists of primary objectives centered on assessing the appropriateness of cohort definitions, assessing validity and biases present in extracted data, and analyzing the data to answer key initial research questions. After these primary goals are met, a set of secondary objectives will be addressed with a deeper level of data collection or analysis. Funding of data collection or analysis for these secondary objectives will depend on resources and may involve single study sites or combinations of some but not necessarily all participating sites.

For the COVID-19 related study addendum, we will leverage the existing network and data collection platform to extract existing medical and public health records to address key research questions about the incidence of COVID illness and its complications among high risk individuals, utilizing medical, laboratory, pharmacy, and vaccination records. We will build prospective cohorts of individuals, identify individuals within the cohorts that are at high risk of complications related to SARS-CoV-2 infection, and enumerate clinical signs and symptoms, severe outcomes, and the care trajectory associated with COVID-19 among those individuals most at risk for severe outcomes.

### ***Analytic objectives***

We aim to determine the validity of and biases generated from different definitions used to construct retrospective cohorts from EHR data, determine the validity of these data on influenza vaccination for the current season and prior seasons, and explore the frequency and patterns of testing for influenza and other respiratory viruses among inpatients and outpatients, and evaluate biases that may come from using information from clinician-driven testing in epidemiologic studies. We will then build upon the information gained in the first objectives to measure incidence of health outcomes related to influenza and estimate the effectiveness of current season influenza vaccination in preventing severe influenza-associated outcomes (e.g. hospitalization, intensive care unit admission) by age groups, influenza type/subtype, and high-risk conditions.

## **1.5 Methods Summary**

To fulfill the network aims each site will construct patient cohorts based on care seeking and if possible health insurance enrollment criteria and extract select demographic, health system utilization, medical, laboratory, pharmaceutical, and vaccination records from electronic health

records for each retrospective study period. Study periods are defined by influenza seasons. Each site will determine the number of years for which they will contribute data, or the period of time in which records are robust enough for cohort identification, beginning as early as the 2010/11 influenza season.

Retrospective cohorts will be defined among individuals within the health system who would visit an outpatient provider or would be hospitalized within the health system if s/he had an influenza virus infection that required hospitalization. A cohort will be defined for each influenza season during the retrospective surveillance period. Influenza season will be broadly defined from September 1 of a year through May 31 of the following year. For analysis, influenza seasons can be more narrowly defined by the frequency of influenza positive test results within the health system. Separate pediatric and adult cohorts will be defined.

For each individual in each of the cohorts, information will be extracted from medical and public health records. Each site will contribute individual-level, but limited, data for the primary analytic objectives. For assessing risk of influenza-associated outcomes, we expect to calculate rates of influenza-positive respiratory hospitalizations among persons at risk of influenza during the season. We intend to calculate rates stratified by age group, high risk underlying conditions, current season vaccination status, current and prior season vaccination status, and other stratifications as suggested by the collaborating network. We will further estimate influenza vaccine effectiveness against influenza-associated respiratory hospitalization and other associated outcomes, both overall and stratified by high-risk conditions. Possible further analytic objectives may be pursued given resources and successful completion of the primary objectives.

## **2.0 Abbreviations**

ARI – Acute respiratory illness

CDC – Centers for Disease Control and Prevention

CPT – Common procedural terminology

EHR – Electronic health record

HIPAA - Health Insurance Portability and Accountability Act

ICD – International Classification of Diseases

ICU – Intensive care unit

RSV – Respiratory syncytial virus

RT-PCR – Reverse transcription polymerase chain reaction

VE – Vaccine effectiveness

## **3.0 Background Information**

Annual influenza vaccination remains the best way to protect against influenza virus infection and associated complications. Each year, the Centers for Disease Control and Prevention (CDC) engages in surveillance and research activities dedicated to estimating the burden of influenza and the effectiveness of seasonal influenza vaccination. While the existing surveillance and research networks are large, encompassing catchment areas of millions or enrolling many thousands of patients with respiratory illness each year, they are not designed to capture the incidence of influenza or the effect of influenza vaccination among people at high-risk of complications nor for rare, yet severe outcomes. Other respiratory viruses also circulate in the United States during the influenza season and sporadically cause widespread illness, contributing to the number of hospitalizations and intensive care unit (ICU) admissions during the winter months in the United States.

This network of health systems is being developed to pilot and understand limitations in the use of existing medical, laboratory, pharmaceutical, and vaccination records for conducting research on complications of influenza and other respiratory virus infections and influenza vaccine effectiveness (VE) against complications of influenza. Ultimately, the network will be able to estimate and further understand the risk of severe outcomes associated with influenza and other respiratory viruses in the general population and among those at high risk of complications. Furthermore, this network can be leveraged to help us understand how seasonal influenza vaccination mitigates several influenza complications.

### **3.1 Scientific Background**

#### ***Burden of influenza***

Influenza continues to be of global public health importance due to the burden of influenza disease across age groups and among individuals with chronic conditions who are at high risk for severe influenza-related outcomes. Annually in the United States, influenza viruses result in millions of illnesses, hundreds of thousands of hospitalizations, and tens of thousands of deaths.<sup>1 2</sup>

While all people are susceptible to influenza virus infection, young children, older adults, and people with underlying medical conditions are particularly vulnerable and account for the majority of influenza-associated morbidity and mortality each year.<sup>3 4</sup> Routine surveillance for hospitalizations with laboratory-confirmed influenza conducted in the United States show that 82–94% of adults and 34–59% of children hospitalized with influenza each season have an underlying medical condition.<sup>5</sup> In fact, across the seasons, people with underlying conditions account for >90% of flu-hospitalizations.<sup>6</sup> A recent population-based study in Auckland, New Zealand, found increased influenza-related hospitalizations among individuals with chronic

medical conditions including congestive heart failure (CHF), chronic obstructive pulmonary disease (COPD), coronary artery disease (CAD), and end-stage renal disease (ESRD) among others.<sup>7 8</sup> Aside from this study in New Zealand, there are limited data on the actual incidence of severe influenza-related events among those at highest risk.

### ***Influenza vaccine effectiveness to prevent severe influenza***

Influenza vaccination is the most important public health strategy to prevent influenza illness and vaccination is recommended annually for all people aged 6 months or older in the United States.<sup>9</sup> However, influenza VE varies from season to season, and across different age and risk groups.<sup>10</sup> A growing body of research points to the effectiveness of influenza vaccination in preventing influenza-associated hospitalizations and other severe outcomes, such as admission to the ICU and in-hospital death.<sup>11 12 13 14 15</sup> While these studies have begun to explore the benefits of vaccination in reducing the risk of influenza-related severe outcomes, with some studies suggesting that the vaccine is more effective in reducing severe disease than reducing infection<sup>16</sup>, the data are largely inconclusive as most are insufficiently powered or not designed to examine and stratify VE by high-risk groups, including children, older adults and those with chronic conditions.

Of the recent studies, Thompson, et al., examined hospitalized adults over multiple influenza seasons found that influenza vaccination decreased influenza-related hospitalizations by 37%, and reduced ICU admission for influenza-related complications by 82%.<sup>17</sup> Analysis of influenza-positive hospitalized adults in the United States also suggests 37% reduction in odds of ICU admission among those vaccinated for influenza.<sup>18</sup> Similarly, studies have shown that influenza vaccination can decrease children's risk of admission to the pediatric intensive care unit by as much as 74%<sup>19</sup> and influenza-associated mortality by 65%.<sup>20</sup> Studies of influenza VE

among those at high risk for influenza-associated complications are rare, with the exception being pregnant women for whom several studies have demonstrated effectiveness of influenza vaccination against influenza infection during pregnancy and in the neonate.<sup>21 22 23</sup>

### ***Burden of other respiratory viruses***

Respiratory viruses, such as respiratory syncytial virus (RSV), adenoviruses, rhinoviruses, enteroviruses, parainfluenza viruses, and coronaviruses, commonly circulate in the United States and contribute to the, so-called, “cold and flu” season. RSV is a common cause of lower respiratory tract infection in young children and the elderly, resulting in substantial numbers of hospitalizations in the United States and globally.<sup>24</sup> More recently in late 2019 and early 2020, the emergence of a novel coronavirus (SARS-CoV-2) resulted in a pandemic. This collaborative network is uniquely poised to address questions about the occurrence and severity of numerous respiratory viruses, including newly emerging viruses.

## **3.2 Rationale and Justification**

Each year CDC communicates with the public about the burden of influenza and the importance of influenza vaccination and other prevention and control measures. People get vaccinated for influenza in part because they perceive the risk of influenza to the health of themselves and their family to be high and they perceive that the influenza vaccine is effective. Part of CDC’s mission is, then, to empower people by using information and data to protect themselves and their family from influenza.

This mission drives the need to measure the incidence of severe health outcomes related to influenza in a variety of groups, including among pediatric and adult age groups, and by high-risk conditions. Furthermore, it is imperative to understand the effectiveness of influenza vaccination against rare, but severe, complications of influenza in the general population and

also among people with underlying medical conditions. These data are essential to supporting the communication messages that CDC and other public health entities rely on to keep people safe from influenza each year.

To effectively estimate the full burden of severe influenza and address these needs, annual retrospective cohorts of individuals will be constructed from a network of large health systems. Within each cohort, individuals can be linked with their vaccination records and their clinical data can be queried to estimate the incidence of outcomes of interest, including influenza-related hospitalizations and outpatient visits. Building each cohort from clinically available data within health systems that integrate, or link, medical records across an individual's medical encounters will enable the analysis to be stratified by underlying medical conditions. Furthermore, joining multiple health systems together will provide the statistical power to investigate less common underlying conditions or clinical outcomes, like ICU admission.

### **3.3 Expected Benefits**

There are no direct benefits to 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 influenza. For example, information from this study may allow for the examination of less common underlying conditions, which to date, have been difficult to examine with statistical precision. These data may influence vaccination strategies for high risk groups which may improve future outcomes for children, elderly adults, and others with high risk conditions. In addition, understanding factors that put individuals at higher risk for influenza-associated hospitalization may help in developing and improving prevention and treatment guidelines.

## 4.0 Study Objectives

We aim to develop a multi-site collaboration to better utilize existing medical and public health records to address key research questions for influenza prevention and control. We will build retrospective cohorts of individuals, identify individuals within the cohorts that are at high risk of influenza complications, determine whether individuals within the cohorts have been vaccinated for influenza, and enumerate laboratory-confirmed influenza-associated events. This effort consists of primary objectives centered on methodological goals and initial research questions.

### 4.1 Primary Objectives

1. Assess the validity and biases associated with defining the source population based on care-seeking.
  - Compare characteristics of the different cohorts, as defined above, assessing differences in the demographics, the frequency of underlying high-risk conditions, and the frequency of respiratory hospitalizations during influenza season.
2. Assess internal and external validity of influenza vaccination data for current and prior seasons.
  - Internal validity:
    - Assess the accuracy of the derived vaccine status variable in relation to a manual review of records on a subset of the cohort.
    - Describe the contributions to the overall conclusion about vaccination status by various sources from which vaccination data were obtained.
    - Describe the consistency of vaccination data that are available by data source, including date of vaccination, vaccine type, vaccine route, vaccine

lot number, setting in which vaccine was administered, number of doses of vaccine (children <9 years).

- Describe completeness of vaccination data (e.g. date of vaccination, type of vaccine, lot number) by data source for the current season and prior seasons.

- External validity:

- Compare vaccine coverage and data completeness to values reported from other studies with active enrollment / prospective data collection.

3. Assess biases associated with clinical testing for influenza and other respiratory viruses and whether they differ in the inpatient versus outpatient setting:

- Assess the accuracy of the extracted test dates, types, and results in relation to a manual review of records on a subset of the cohort.
- Estimate proportion of cohort enrollees tested for influenza and other respiratory viruses during the influenza season.
- Describe types of influenza and other respiratory virus tests performed on cohort enrollees by pediatric and adult age groups and how test types vary by setting.
- Estimate the proportion of respiratory hospitalizations tested for influenza during the influenza season and factors associated with influenza testing in the inpatient setting including age, high-risk conditions, current season influenza vaccination status, ICU stay, timing of encounter within season, diagnoses codes.
- Estimate the proportion of outpatient visits for acute respiratory illness (ARI) that were tested for influenza and factors associated with influenza testing in the outpatient setting including age, high-risk conditions, current season influenza

vaccination status, visit type, timing of encounter within influenza season, diagnoses codes.

4. Measure the incidence of health outcomes related to influenza by pediatric and adult age groups as outlined below. The incidence of these outcomes will be assessed by age groups, demographic characteristics, influenza type/subtype, and high-risk conditions:
  - Laboratory confirmed influenza-associated respiratory hospitalizations.
  - ICU admission among persons with laboratory-confirmed influenza-associated respiratory hospitalizations.
  - Invasive mechanical ventilation use among persons with laboratory-confirmed influenza-associated respiratory hospitalizations.
  - In-hospital death among persons with laboratory-confirmed influenza-associated respiratory hospitalizations.
  - Laboratory-confirmed influenza among persons with outpatient visits for acute respiratory illness.
  
5. Estimate the effectiveness of current season influenza vaccine in preventing the below outcomes by age groups, influenza type/subtype, and high-risk conditions:
  - Laboratory confirmed influenza-associated respiratory hospitalizations (all ages).
  - ICU admission among persons with laboratory-confirmed influenza associated respiratory hospitalizations (all ages).
  - Laboratory-confirmed influenza among persons with outpatient visits for acute respiratory illness (pediatrics; adults if outpatient influenza testing permits).

Appendix A presents the main outcomes, exposures, and modifiers of interest for the primary and secondary study objectives. Appendix B provides the data dictionary of analytic variables needed to meet these objectives.

## **4.2 Secondary Objectives**

After these primary goals are met, a set of secondary objectives will be addressed with a deeper level of data collection or analysis. Funding of data collection or analysis for these secondary objectives will depend on resources and may involve single study sites or combinations of some but not necessarily all participating sites.

- Assess feasibility of defining underlying high risk conditions using a combination of ICD codes, medications, and laboratory data.
- Assess whether specific forms of immunosuppression or use of immunosuppressive medications may increase the risk of severe influenza-associated outcomes and/or reduce influenza VE.
- Among children, examine the association between the number and timing of priming doses with the VE against outcomes of interest in subsequent seasons.
- Examine the association between prior season vaccinations on current season VE.
- Determine measures of disease severity at hospital admission using data on signs and symptoms, vital signs and lab findings and assess their impact on measures of VE.
- Measure the incidence of secondary outcomes following influenza-related hospitalization, including hospital readmissions, rehabilitation and long-term care, post-discharge death.
- Collect data prospectively during an ongoing influenza season to produce within-season estimates of influenza VE.

## **5.0 Methods**

### **5.1 Study Design**

All sites will define retrospective seasonal cohorts of adults and children, which are intended to include all individuals within the health system who would be hospitalized or seek outpatient care within that health system if they had an influenza virus infection that required hospitalization. Among all individuals in the retrospective cohort, medical, laboratory, and vaccination records will be extracted to assess the primary study objectives.

The influenza season will be defined as the time between September 1 of a year and May 31 of the subsequent year. For example, the 2015–2016 influenza season begins on September 1, 2015 and ends after May 31, 2016. Each study site will determine the number of previous influenza seasons (considering seasons between 2010/11 and 2018/19) for which their site will contribute data, depending on the availability and quality of electronic health record (EHR) data in their health system. Adult and pediatric cohorts will be analytically defined for each influenza season for which a site contributes data. Cohorts will be followed up for the duration of each influenza season to meet the primary objectives.

Figures 1 and 2 display the high-level study design for the primary hospitalization and outpatient outcomes, respectively.

Figure 1. High-level study design for hospitalization outcomes

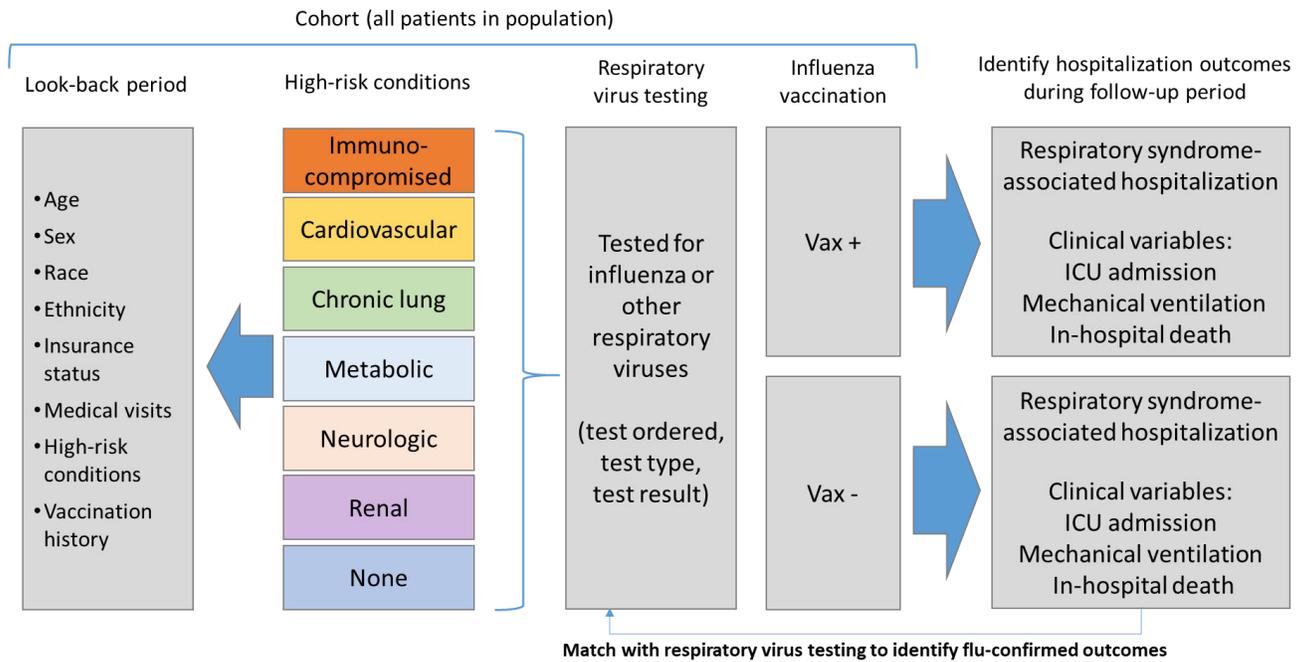
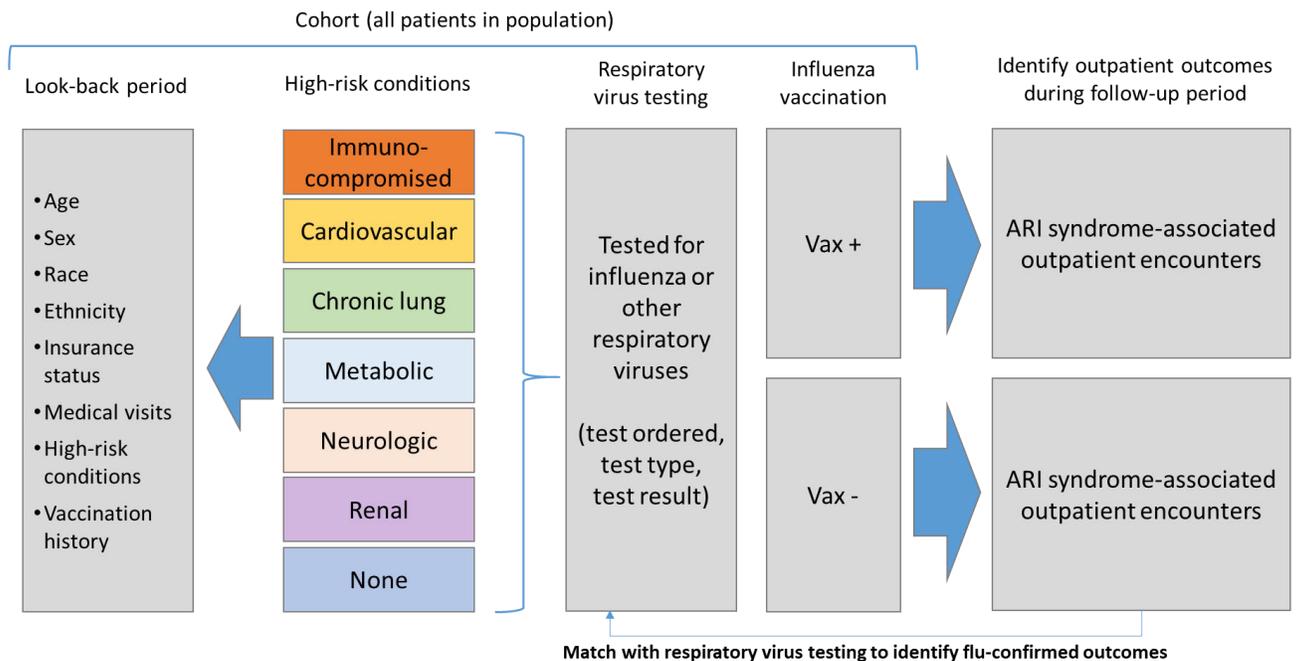


Figure 2. High-level study design for outpatient outcomes



## 5.2 Participant Eligibility

Eligible individuals in each seasonal cohort will be those who can and would seek care for influenza at a healthcare facility within the network. Inclusion in the cohorts will be defined by each site based on enrollment information within the health system, allowing for  $\leq 30$  day administrative gaps, and care being sought within the system in the 12 months prior to the start of the influenza season. Individuals born in the 12 months prior to the start of the influenza season will be eligible for the cohorts as long as other inclusion criteria are met and no exclusion criteria are met.

### *Inclusion criteria*

Several different inclusion criteria are possible and will be assessed in the first primary objective. Individuals will be included if they reside in the catchment area or are active members in the health system's insurance plan and then will be evaluated against each of the following criteria.

- 1) Any individual who is otherwise eligible and had a telephone-based care visit, an ambulatory encounter, an urgent care encounter, an emergency department encounter, or an inpatient encounter with the health system in the 12 months prior to the start of influenza season. Ambulatory encounters include office-based visits to a primary care practice office (including but not limited to general pediatrics, general internal medicine, med-peds, obstetrics/gynecology, or general family medicine), or office-based visits to a medical specialty office (including but not limited to cardiology clinic, diabetes clinic, rehabilitation clinic, gastrointestinal specialist, etc.).

- 2) Any individual who is otherwise eligible and had an ambulatory encounter (to a primary care practice office or a medical specialty office) or an inpatient encounter with the health system in the 12 months prior to the start of influenza season.
- 3) Any individual who is otherwise eligible and had an ambulatory encounter with the health system in the 12 months prior to the start of influenza season.
- 4) Any individual who is otherwise eligible and had an encounter to a primary care practice office with the health system in the 12 months prior to the start of influenza season. General practice encounters include office-based visits to a general pediatrics, general internal medicine, med-peds, obstetrics/gynecology, or general family medicine clinic.
- 5) **[For closed health systems only]** For health systems where individuals can have an active membership in the health system, an additional inclusion criterion will be considered. The inclusion criteria will be any individual who was an active member in the health system for the period of 12 months before the start of influenza season (or was born and became a member during the 12 months prior to the start of or during the influenza season) and were active members until the end of influenza season, death, or disenrollment, whichever occurred first.

For a single season, such as the 2017/18 season, the cohorts defined by the different inclusion criteria will be assessed in terms of the number, the demographic characteristics, the frequency of subsequent medical care, and the frequency of respiratory hospitalizations among cohort members (Appendix C). The definition of “medical care” to use will be determined by joint consensus among the collaborating sites prior to extracting data for the interim and final datasets.

### ***Exclusion criteria***

Individuals will be excluded from the seasonal cohorts if they are not active members of the health system or they resided outside of the catchment area of the health system for the entire 12-month period before the start of influenza season. Active membership can be determined based on enrollment in the health system. Catchment area may be known or approximated based on prior studies or other surveillance activities that the health system is engaged in. Catchment area can also be determined starting first with the counties surrounding the hospitals and then be further refined using information on county of residence, selecting the counties that unique patients most frequently live in.

### **5.3 Observation Time during Influenza Season**

For the primary objective of estimating the incidence of influenza-associated respiratory hospitalizations and ARI outpatient encounters, the person-time at risk for the underlying cohort will be needed. Person-time at risk for influenza will be calculated for all individuals in the seasonal retrospective cohorts. Observation time for a given season will start on the day that influenza season begins, September 1 of each year. For the same season, observation time will end on the day of death, disenrollment, move away from the catchment area, or the end of the influenza season (May 31), whichever occurs first. The total observation time, person-time, is then the difference between the start and end of observation. For analyses, observation time may be restricted to a shorter period of time when influenza viruses are known to be circulating in the communities surrounding the collaborating sites. Final determination for analysis will be made by group consensus.

### **5.4 Exposures of Interest**

#### ***Vaccination status***

Within each annual retrospective cohort, an individual will be considered vaccinated for influenza if there is documented evidence of influenza vaccination and evidence that the vaccination was administered at least two weeks prior to an influenza-associated respiratory hospitalization (or influenza-associated ARI outpatient visit).

In order to assess influenza vaccination status, multiple sources will be queried, as available, including EHR, state/local vaccine registries, all-payer claims or other billing databases. For each available source, data will be provided on date of vaccination, type of vaccine administered, vaccination route, location of vaccination, vaccine lot number and number of vaccine doses (for children <9 years of age). Vaccination records will be considered as evidence of vaccination if there is at least a partial date (month and year). Children (aged <9 years) will be considered fully vaccinated for the current season if they received two doses, at least one month apart, during the current season or if they received one dose of the current season's vaccine and had at least two doses in prior seasons.

Network sites will provide information on number of sources queried for vaccine data and completeness of data by vaccine source. 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 vaccination data extracted from EHRs, sites will provide information on how vaccination data was pulled (for example, through use of CPT or other codes).

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).

### ***Influenza and other respiratory virus testing***

Data on all influenza and other respiratory virus test results will be extracted for all members of the cohort during the surveillance period. 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 conditions***

For the primary objectives, high risk conditions will be defined using ICD diagnosis codes. A patient with at least 1 inpatient encounter or at least 1 outpatient encounter with a high-risk ICD diagnosis code (see Appendix D) recorded in the 12 months prior to the current influenza season will be considered as having a high-risk condition. The following high-risk conditions will be defined:

- Chronic lung disease;
  - Asthma;
  - COPD;
  - Pulmonary tuberculosis
  - Endemic mycoses

- Chronic metabolic disease;
  - Diabetes mellitus;
- Blood disorders;
- Cardiovascular disease;
  - Coronary artery disease;
  - Heart failure;
  - Congenital heart disease;
- Neuromuscular disorder;
- Neurologic disorder;
- Immunocompromised condition;
  - Solid organ malignancy;
  - Hematologic malignancy;
  - Solid organ transplant;
  - Hematopoietic stem cell transplant;
- Chronic renal disease;
- Gastrointestinal/liver disease;
- Rheumatologic/Autoimmune condition.

Additional underlying conditions, including, but not limited to, pregnancy, sickle cell disease, hypertension, obesity, 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 healthcare utilization***

Healthcare utilization is an important data element to consider for the primary objectives. The use of healthcare directly informs the definition of each of the seasonal cohorts and also is

known to be associated with subsequent care for influenza and other respiratory virus infections as well as the propensity for seasonal influenza vaccination. Healthcare utilization will be captured using data on the number of telemedicine encounters, office-based encounters, and inpatient hospital admissions, for any reason, that occurred prior to the start of influenza season, during the influenza, and, for those who remain in the cohort, after the end of influenza season.

To better understand healthcare utilization patterns in the cohort, among vaccinated individuals, and among those with respiratory-associated inpatient or outpatient visits, several different categories of healthcare encounters will be enumerated in the 12 months before, during, and 12 months after each influenza season.

- Primary care encounters include office-based visits that occurred in a general pediatric, general internal medicine, med-peds, general family medicine, or obstetrics/gynecology practice.
- Telemedicine/telehealth encounters include any call for a clinical consultation
- Medical specialty care encounters include office-based visits that occurred in a practice for cardiology clinic, diabetes clinic, rehabilitation clinic, gastrointestinal specialist, etc. Excluded visits are those that are for laboratory only, imaging only, or nurse only.
- Urgent care encounters.
- Emergency department encounters include visits to the emergency department that did not result in an inpatient admission.
- Inpatient hospital admissions include admissions to an inpatient ward or an observation unit.

| <b>Vaccine Type</b> | <b>Age group</b>                     |
|---------------------|--------------------------------------|
| SARS-COV-2          | All ages (as licensed and available) |

***Data on other 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 in the 12 months prior to influenza season (the encounter that included the individual into the cohort)
- Primary, secondary, tertiary and quaternary insurance type

**5.5 Outcomes of Interest**

- Respiratory illness-associated hospitalizations with laboratory-confirmed influenza
  - Respiratory hospitalizations will be defined using ICD-9 and ICD-10 discharge diagnoses codes as listed in Appendix D.
  - Data on all respiratory hospitalizations will be collected for the entirety of the surveillance period (influenza season).
  - The respiratory hospitalizations will be linked with data on influenza testing to determine if the hospitalization was associated with laboratory-confirmed influenza.
    - Laboratory-confirmed influenza associated respiratory hospitalization will be defined by a positive influenza test within 14 days prior to hospitalization or no more than 3 days after hospital admission.

- Hospitalization will be defined as admission to an inpatient ward or an observation unit. Time spent only in the emergency department will not count as a hospitalization.
- Re-admissions that occur within 30 days of the index hospitalization will be counted as part of the index hospitalization. Readmissions that occur >30 days after the index hospitalization will be considered new hospitalization events.
- ICU admission among persons hospitalized with laboratory-confirmed influenza-associated respiratory hospitalization
- Invasive mechanical ventilation among persons hospitalized with laboratory-confirmed influenza-associated respiratory hospitalization
- In-hospital death among persons hospitalized with laboratory-confirmed influenza-associated respiratory hospitalization
- Acute respiratory illness-associated outpatient visits with laboratory-confirmed influenza
  - Acute respiratory illness-associated outpatient visit will be defined using the ICD-9 and ICD-10 discharge codes listed in Appendix D.
  - Data on all acute respiratory illness-associated outpatient visits will be collected for the entirety of the surveillance period (influenza season).
  - The acute respiratory illness outpatient visit will be linked with data on influenza testing to determine if the visit was associated with a positive influenza laboratory test.

Appendix D provides a more detailed list of the variables used to define outcomes of interest.

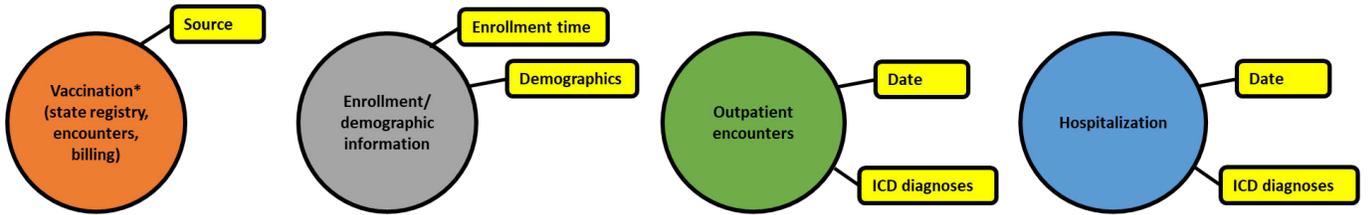
## **5.6 Summary of Data Elements for Primary Objectives**

Please see Appendix B for more detailed list of data elements

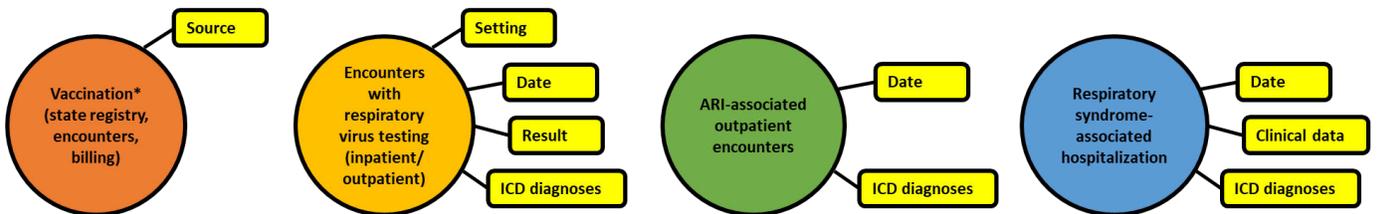
- For the entire cohort:

- Demographic information (age, sex, race, ethnicity)
- Date of enrollment in health plan and/or date of first encounter in the health system in the 12 months before the start of influenza season (as applicable)
- Healthcare utilization (primary/secondary insurance type, number of medical encounters before and during influenza season). This is further defined to be the monthly number of healthcare encounters by setting (general practice, medical specialty, emergency department, urgent care, telehealth, or inpatient).
- Underlying high risk medical conditions (derived from ICD codes)
- Vaccination status (influenza and other vaccinations received dating back to 2010)
- All hospitalizations and outpatient encounters with respiratory virus testing during influenza season of interest:
  - Date of encounter (or date of admission)
  - Date of testing
  - Clinical setting
  - Influenza and other respiratory virus test type and result
  - Diagnosis codes from the encounter/admission
- Respiratory syndrome-associated hospitalizations during influenza season of interest:
  - Hospital admission and discharge dates
  - Diagnosis codes
  - Severe outcomes (ICU admission, mechanical ventilation, in-hospital mortality)
- ARI syndrome-associated outpatient encounters during influenza season of interest
  - Encounter date
  - Diagnosis codes

Data streams for entire cohort during to-be-determined lead-in period:

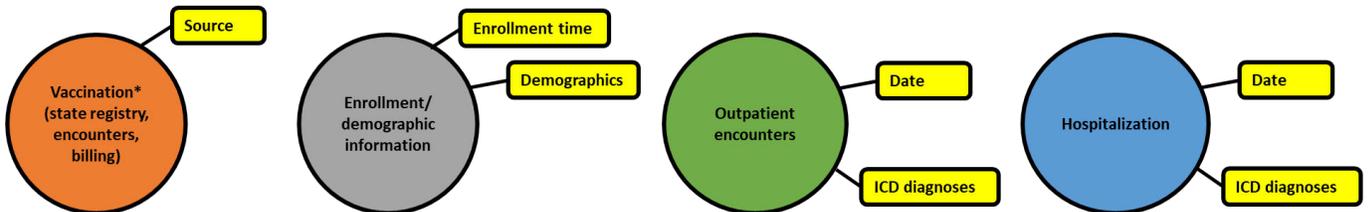


Data streams for influenza season of interest:

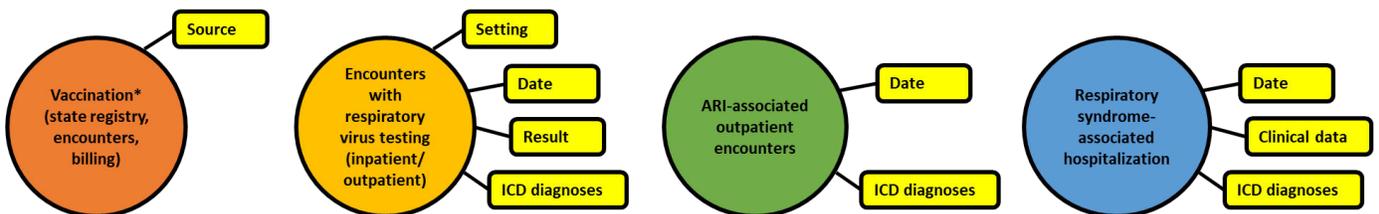


All sources should be linked by subject ID  
 \*For influenza and other vaccinations received dating back to 2010, where possible.

Data streams for entire cohort during to-be-determined lead-in period:



Data streams for influenza season of interest:



All sources should be linked by subject ID  
 \*For influenza and other vaccinations received dating back to 2010, where possible.

## **5.7 Summary of Methods for Secondary Objectives**

After the primary objectives are met, a set of secondary objectives will be addressed with a deeper level of data collection or analysis. Funding of data collection or analysis for these secondary objectives will depend on resources and success of the primary objectives. Secondary objectives may involve some but not necessarily all participating sites.

The secondary objectives include capturing and assessing the validity of additional data elements (such as laboratory values, vital signs, and medications) to better characterize underlying medical conditions and severity of influenza virus infections. It is anticipated that the methods for assessing the validity of the additional data will be similar to the methods used to validate data for the primary objectives, including comparison of extracted data to data in the primary data source, searching for deviations from expected values, and comparing findings to external sources.

In addition to expanding the data elements being captured, sites participating in the secondary objectives may be asked to extract data from the 2019–2020 influenza season, to submit “prospective”, or near real-time, data from the 2020–2021 influenza season, or to expand outcome ascertainment beyond the pre-defined influenza season.

As more details of the methods for the secondary objectives are discussed and consensus reached with the participating sites, the methods and protocol will be amended.

## **6.0 Statistical Analysis**

### **6.1 Analysis Plan**

***Objective 1: Assess the validity and biases associated with defining the source population based on care-seeking***

Correctly capturing the health system's denominator is a critical component of the study objectives, thus all health systems will be asked to validate the eligibility definitions using their information from their membership or source population.

- Compare the demographics of participants defined using the five inclusion criteria, using appropriate bivariate statistics as appropriate.
- Compare, using appropriate bivariate statistics, the healthcare utilization between all known members and those participants defined using the five inclusion criteria.
- Compare the frequency of respiratory-associated hospitalization between the cohorts as defined using the five inclusion criteria.

***Objective 2: Validation of influenza vaccination data***

- Describe the frequency of influenza vaccination among cohort members and vaccine type received. Descriptions will include the full cohort and will be stratified by individual characteristics such as age group, high-risk condition, and prior healthcare utilization.
  - Frequency of vaccination will be compared across strata using appropriate bivariate statistics (e.g. chi-square tests or Fisher's exact tests).
- Among those determined to be vaccinated for influenza
  - Assess the proportion that were identified through various data sources (e.g., procedure codes from a medical visit, other notation in the medical chart, vaccine registry, billing data, etc.). Assess whether the source of vaccination data varies by individual characteristics, including age group, high-risk conditions, and prior healthcare utilization.

- Compare the completeness and consistency of data on date of vaccination and vaccine type between data sources and by patient characteristics such as age group, high-risk conditions, and prior healthcare utilization.
- Comparisons of data by source, for completeness, and for consistency and how these differ by patient characteristics will be conducted using appropriate bivariate tests of association and/or multivariate regression. As an example, regression analyses could include data source as the outcome of interest with patient characteristics as explanatory variables.
- Describe the timing of vaccine receipt among cohort members by age group, high risk condition, prior healthcare utilization, and vaccine type received.

This description will be explored using a variety of methods, including time to vaccination, median week of vaccination, or frequency of vaccination at least two weeks prior to the start of influenza circulation. Comparison of timing by patient characteristics will be conducted using methods appropriate to the outcome type; for example, proportional hazards modeling for time to vaccination.

***Objective 3: Biases associated with clinical testing***

- Describe the frequency of clinical testing for influenza and other respiratory viruses in the cohort during an influenza season by setting (inpatient vs. outpatient), age group, high-risk status, influenza vaccination, prior healthcare utilization, and timing of encounter within the season.
- Describe the diagnoses associated with visits where clinical testing for influenza and/or other viruses was performed.

- Among hospitalizations with respiratory illness during the influenza season, describe the frequency and types of clinical testing used.
  - Characteristics and predictors of receipt of diagnostic test for influenza will be assessed using bivariate tests of association and/or multivariate regression with receipt of test as outcome of interest and patient and clinical characteristics (including age, vaccination status, high-risk conditions, timing of the clinical encounter within the season, prior healthcare utilization, diagnoses, etc.)
- Among outpatient visits for ARI during the influenza season, we will describe the frequency and types of clinical testing used.
  - Characteristics and predictors of receipt of diagnostic test for influenza will be assessed using bivariate tests of association and/or multivariate regression with receipt of test as outcome of interest and patient and clinical characteristics (including age, vaccination status, high-risk conditions, timing of the clinical encounter within the season, prior healthcare utilization, diagnoses, etc.)

***Objective 4: Measure incidence of health outcomes related to influenza***

- Incidence of laboratory-confirmed influenza-associated respiratory hospitalizations will be calculated by dividing the number of respiratory hospitalizations associated with a positive influenza laboratory test observed during each study season by the person-time of the cohort at risk during the corresponding season.
  - This will be calculated for the following risk groups: age group, high-risk conditions, vaccinated for influenza, unvaccinated, race/ethnicity, socio-economic status, etc.
  - Incidence will be compared between risk groups using rate differences or relative rates

- Incidence of ICU admission, mechanical ventilation, and in-hospital death associated with influenza will be similarly calculated, with the respective numerators restricted to those laboratory-confirmed influenza-associated respiratory hospitalizations with the relevant outcomes identified.
  - This will be calculated for the following risk groups where sample size permits: age group, high-risk conditions, vaccinated for influenza, unvaccinated, race/ethnicity, socio-economic status, etc.
  - Incidence will be compared between risk groups using rate differences or relative rates
- Incidence of laboratory-confirmed influenza-associated acute respiratory illness (ARI)-associated outpatient visits will be calculated by dividing the number of ARI visits with a laboratory test positive for influenza observed during each study season by the person-time of the cohort at risk during the corresponding season.
  - This will be calculated for the following risk groups where sample size permits: age group, high-risk conditions, vaccinated for influenza, unvaccinated, race/ethnicity, socio-economic status, etc.
  - Incidence will be compared between risk groups using rate differences or relative rates

As multiple events are possible during a season, incidence estimates throughout this objective may be calculated in two ways: 1) counting only the first occurrence, and 2) counting the cumulative number across the season.

***Objective 5: Assessment of vaccine effectiveness***

- We will estimate the effect of influenza vaccination on risk of lab-confirmed influenza-associated respiratory hospitalization by age group, influenza type/subtype, and high-risk conditions
  - The risk of influenza-associated respiratory hospitalization among cohort members who were or were not vaccinated with the current season influenza will be compared using appropriate methods for longitudinal data with time varying exposure, including Poisson regression or proportional hazards modeling.
  - VE will be estimated as  $(1 - \text{adjusted relative risk}) \times 100$ .
  - We will assess and adjust for confounders as appropriate. These may include characteristics such as age, sex, site, race/ethnicity, high-risk health conditions, prior healthcare utilization, time of event during the influenza season, etc.
  - Data collected for this study are intended to be pooled across sites. VE will first be estimated for each site and heterogeneity in estimated VE across sites examined.
- We will estimate the effect of influenza vaccination on risk of lab-confirmed influenza-associated ARI outpatient visit among children (adults only if influenza testing allows for this) by age group, influenza type/subtype, and high-risk conditions
  - The risk of influenza-associated outpatient visits among cohort members who were or were not vaccinated with the current season influenza will be compared using appropriate methods for longitudinal data with time varying exposure, including Poisson regression or proportional hazards modeling.
  - VE will be estimated as  $(1 - \text{adjusted relative risk}) \times 100$ .

- We will assess and adjust for confounders as appropriate. This may include characteristics such as age group, sex, site, race/ethnicity, high-risk health conditions, prior healthcare utilization, time of event during the influenza season, etc.
- Data collected for this study are intended to be pooled across sites. VE will first be estimated for each site and heterogeneity in estimated VE across sites examined.
- VE analyses against influenza-associated respiratory hospitalizations will also be conducted using a test-negative case-control design
  - This analysis will be conducted among individuals who were hospitalized for acute respiratory illness during the influenza season and who were tested for influenza.
  - Cases will be defined as hospitalized individuals who tested positive for influenza.
  - Controls will be defined as hospitalized individuals who tested negative for influenza.
  - Influenza vaccination status will be determined for all cases and controls and the odds of vaccination will be compared using logistic regression.
  - VE will be estimated as  $(1 - \text{adjusted odds ratio}) \times 100$ .
  - We will assess and adjust for confounders as appropriate. This may include characteristics such as age group, sex, site, race/ethnicity, high-risk health conditions, prior healthcare utilization, time of event during the influenza season, etc.
  - Data collected for this study are intended to be pooled across sites. VE will first be estimated for each site and heterogeneity in estimated VE across sites will be examined.

Beyond the primary objectives listed here, additional analyses may be pursued and will be considered by the joint steering committee.

## **6.2 Sample Size**

Table 1 presents sample size requirements for the primary objective of VE against severe outcomes. Existing data from hospital-based surveillance was used to estimate the reported incidence of laboratory-confirmed hospitalizations, ICU admissions, and deaths.<sup>25</sup> Data from the Behavioral Risk Factor Surveillance System (BRFSS) and National Health Interview Survey (NHIS) was used to estimate vaccine coverage.<sup>26</sup>

For adults, we assumed vaccine coverage of 30% in adults aged 18-49 years, 42% in adults aged 50-64 years, and 65% in adults 65 years and older. We assumed a population made up of 33% of adults in each of the three age groups.

For children, we assumed vaccine coverage of 70% in children aged 6 months to 4 years and 50% in children aged 5 to 17 years. We assumed a population made up of 50% of children in each of the two age groups.

Table 1 Sample Size Estimations

|                              | Minimum Detectable VE | Person-years needed |             |
|------------------------------|-----------------------|---------------------|-------------|
|                              |                       | Adults              | Pediatrics  |
| VE against hospitalization   | 40%                   |                     | 398,319     |
|                              | 35%                   | 197,593             | 569,374     |
|                              | 30%                   | 292,719             | 842,278     |
|                              | 25%                   | 456,014             | 1.3 million |
|                              | 20%                   | 766,799             |             |
| VE against ICU admission     | 40%                   |                     | 1.01 m      |
|                              | 35%                   | 1.2 million         | 1.4 m       |
|                              | 30%                   | 1.8 m               | 1.9 m       |
|                              | 25%                   | 2.8 m               | 2.8 m       |
|                              | 20%                   | 4.8 m               |             |
| VE against in-hospital death | 40%                   |                     | 28.1 m      |
|                              | 35%                   | 5.4 m               | 39.4 m      |
|                              | 30%                   | 7.9 m               | 52.8 m      |
|                              | 25%                   | 12.4 m              | 73.8 m      |
|                              | 20%                   | 20.8 m              |             |

### 6.3 Enhanced Data Collection Elements to Meet Secondary Objectives

For a more comprehensive list of data elements needed to meet the secondary objectives, see Appendix B.

- All inpatient/outpatient encounters for to-be-determined lead-in period (to refine high-risk conditions)
  - Medication prescribing/administration (e.g. asthma or immunosuppressive medications)
  - Laboratory testing/results (e.g. for diabetes, hemoglobin A1C tests)
- Respiratory syndrome-associated hospitalizations during influenza season of interest (all patients; discharge codes listed in Appendix D)

- Medication orders/administration (e.g. influenza antivirals, antibiotics, steroids, vasopressors)
- Expanded laboratory testing/results (e.g. CBC, CMP)
- Vital signs (e.g. heart rate, respiratory rate, blood pressure, temperature, O2 saturation, Glasgow Coma Scale)
- Non-invasive mechanical ventilation (e.g. CPAP, BiPAP, high flow nasal cannula)
- Radiography (e.g. chest x-ray)
- ARI syndrome-associated outpatient encounters during influenza season of interest (all patients; ICD codes TBD)
  - Medication orders/administration (e.g. influenza antivirals, antibiotics)
  - Expanded laboratory testing/results (e.g. CBC, CMP)
  - Vital signs (e.g. heart rate, respiratory rate, blood pressure, temperature, O2 saturation, Glasgow Coma)
  - Radiography (e.g. chest x-ray)

## **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. As the data coordinating center, Westat will receive data from the sites and derive variables for analysis. Alternatively, 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. Appendix E displays an example of the types of data shared by the sites with Westat for processing.

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**

A listing of the analytic variables necessary to meet the primary project objectives are shown in Appendix A. Depending on funding to address secondary objectives, 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), radiography, and other laboratory tests [including diabetes, hemoglobin A1C, complete blood count (CBC), comprehensive metabolic panel (CMP), etc.]. A listing of the analytic variables necessary to meet the primary

and secondary project objectives is provided in Appendix B. Additionally, Appendix D contains the ICD codes and other variables used to define high-risk conditions, ARI hospitalizations, and outcomes.

### **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. Information for extraction includes data elements described in Section 7.2 and Appendices B and D. 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 (hospital, medical office, urgent care, ER, etc.), county, and state. 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 E, 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. Westat will derive the analytic

variables for sites sending data matching descriptions 1 and 2 above and 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 F, 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 influenza 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. 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 influenza testing on 50% of its pediatric 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 then 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. Appendix G includes some suggestions for data validation steps using primary medical or health records.

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

- 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 F).

Westat's Data Management Plan (Appendix F) 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 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 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: Outcomes, Exposures, and Modifiers for Study Objectives

| Objective   | Outcome Variables  | Exposure Variables   | Stratification Variables   | Confounders/Modifiers                              |
|---|--|--|--|--|
| <i>Primary Objectives</i>   |  |  |  |  |
| Objective 1:<br><i>Assess the validity and biases associated with defining the source population based on care-seeking.</i> | Respiratory hospitalizations   | Cohort definitions<br>Healthcare utilization<br>Healthcare setting | High Risk Conditions<br>Insurance  | Age<br>Sex<br>Race<br>Ethnicity                    |
| Objective 2:<br><i>Determine the internal and external validity of data on influenza vaccination</i>                        | Internal validity<br>Accuracy of derived vaccine status variable in relation to manual review of records<br>Contributions to conclusion about vaccination status by source<br>Consistency and completeness of vaccination data available by source<br><br>External validity<br>Comparison of vaccine coverage and data completeness to values from other studies | Data source<br><br>EHR<br><br>Registry<br>Billing/Admin. records   | Vaccination date<br><br>Vaccine type<br><br>Vaccine route<br><br>Vaccine lot number<br><br>Setting of administration<br><br>Num. of doses (<9 years) |  |
| Objective 3:<br><i>Assess biases associated with clinical testing for influenza and</i>                                     | Proportion of cohort enrollees tested for influenza and other respiratory viruses<br>Types of influenza and other virus tests performed<br>Proportion of respiratory hospitalizations tested<br>Proportion of ARI outpatient visits tested   |  | Age<br><br>Healthcare setting<br>High-risk conditions <sup>1</sup><br>Vaccination status   | Sex<br>Healthcare utilization<br>Race<br>Ethnicity |

| Objective   | Outcome Variables  | Exposure Variables | Stratification Variables   | Confounders/Modifiers   |
|---|--|--------------------|--|---|
| <i>other respiratory viruses</i> <sup>2</sup>   |  |                    | Testing date<br>Timing of encounter<br>Discharge diagnosis codes   | Insurance coverage<br>Hospital length-of-stay   |
| Objective 4:<br><i>Measure incidence of health outcomes related to influenza by pediatric and adult age groups, influenza type/subtype, and high-risk conditions</i>                | Laboratory confirmed influenza-associated respiratory hospitalizations<br>ICU admission<br>Invasive mechanical ventilation<br>In-hospital death<br><br>Outpatient/Urgent care visits |                    | Age<br>Influenza type/subtype<br>High-risk conditions <sup>1</sup> | Sex<br>Healthcare<br>Race<br>Ethnicity<br>Insurance coverage<br><br>Hospital length-of-stay     |
| Objective 5:<br><i>Estimate the effectiveness of current season influenza vaccine in preventing severe outcomes by age groups, influenza type/subtype, and high-risk conditions</i> | Laboratory confirmed influenza-associated respiratory hospitalizations<br>ICU admissions<br>Outpatient/Urgent care visits (pediatrics)   | Influenza          | Age<br>Influenza type/subtype<br>High-risk conditions <sup>1</sup> | Sex<br>Healthcare<br>Race<br>Ethnicity<br><br>Insurance coverage<br><br>Hospital length-of-stay |
| <i>Secondary Objectives</i>   |  |                    |  |   |

| Objective  | Outcome Variables | Exposure Variables | Stratification Variables | Confounders/Modifiers |
|--|-------------------|--------------------|--------------------------|-----------------------|
| <p>Objective 1:<br/><i>Assess feasibility of defining underlying conditions using combination of ICD, medication, and lab data</i></p>                             |                   |                    |                          |                       |
| <p>Objective 2:<br/><i>Assess whether immunosuppression or use of immunosuppressive medications increases risk of severe outcomes or reduces VE</i></p>            |                   |                    |                          |                       |
| <p>Objective 3:<br/><i>Examine association between number and timing of priming doses with VE against outcomes in subsequent seasons (among children only)</i></p> |                   |                    |                          |                       |

| Objective  | Outcome Variables | Exposure Variables | Stratification Variables | Confounders/Modifiers |
|--|-------------------|--------------------|--------------------------|-----------------------|
| Objective 4:<br><i>Examine association between prior season vaccination and current season VE (among adults only)</i>  |                   |                    |                          |                       |
| Objective 5:<br><i>Determine measures of disease severity at hospital admission and assess impact on measures of VE</i>  |                   |                    |                          |                       |
| Objective 6:<br><i>Measure incidence of secondary outcomes following influenza-related hospitalization</i>   |                   |                    |                          |                       |
| <sup>1</sup> High-Risk Conditions: chronic lung disease, asthma, COPD, chronic metabolic disease, diabetes mellitus, blood disorders/hemoglobinopathy, cardiovascular disease, coronary artery disease, heart failure, congenital heart disease, neuromuscular disorder, neurologic disorder, immunocompromised condition, solid organ malignancy, hematologic malignancy, solid organ transplant, hematopoietic stem cell transplant, chronic renal disease, gastrointestinal/liver disease, rheumatologic/autoimmune condition, pregnancy (in woman of childbearing age), prematurity (in children <2 years of age). |                   |                    |                          |                       |
| <sup>2</sup> Influenza (all types/subtypes), respiratory syncytial virus, adenovirus, parainfluenza (1-4), human metapneumovirus, rhinoviruses/enteroviruses, coronaviruses, other novel viruses.  |                   |                    |                          |                       |

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

See accompanying Excel file “VISIONDataElementsCodebook.”

## Appendix C: Assessment of inclusion criteria

Retrospective cohorts for this protocol will be defined based on care-seeking behavior; however, deriving a cohort based on seeking care could introduce bias to estimates of incidence of influenza-associated outcomes and the effect influenza vaccination on reducing the incidence of such outcomes. To assess the validity and biases associated with the definition of the cohort, the first primary objective of the protocol will assess and compare the demographics, frequency of health-care utilization, and frequency of influenza-related outcomes between cohorts defined in different ways.

Following the protocol, five different inclusion criteria are being considered for defining retrospective cohorts for each influenza season. For the 2017/18 influenza season, or another season as agreed upon by the collaborating sites, each collaborating site will define a retrospective cohort using the definitions outlined in the [Participant Eligibility](#) section. **Of note:** only collaborating sites who have access to and can define a source population based on individual membership in a health system will describe characteristics of a cohort as defined using active participant membership.

The following table is an example of what will be filled out to describe the aggregate characteristics of the defined cohort. Of note, the influenza season is defined from September 1 through May 31 of the following year.

|  | <b>In 12 months prior to start of influenza season, individuals who had:</b> |   |                              |  |                                |
|--|--|---|------------------------------|--|--------------------------------|
|  | ≥1 care encounter <sup>a</sup> or ≥1 inpatient admission                     | ≥1 ambulatory care encounter <sup>b</sup> or ≥1 inpatient admission | ≥1 ambulatory care encounter | ≥1 general practice encounter <sup>c</sup> | Active membership <sup>d</sup> |
| <b>Number of unique patients in the cohort</b> |  |   |                              |  |                                |
| <b>Demographics</b>                            |  |   |                              |  |                                |
| Age, in years (n, %)                           |  |   |                              |  |                                |
| <2   |  |   |                              |  |                                |
| 2–4  |  |   |                              |  |                                |
| 5–17   |  |   |                              |  |                                |
| 18–49  |  |   |                              |  |                                |
| 50–64  |  |   |                              |  |                                |
| 65–74  |  |   |                              |  |                                |
| 75–84  |  |   |                              |  |                                |
| ≥85  |  |   |                              |  |                                |
| Sex (n, %)                                     |  |   |                              |  |                                |
| Male   |  |   |                              |  |                                |
| Female   |  |   |                              |  |                                |
| Other  |  |   |                              |  |                                |
| Race (n, %)                                    |  |   |                              |  |                                |
| White  |  |   |                              |  |                                |
| Black  |  |   |                              |  |                                |
| Other  |  |   |                              |  |                                |
| Hispanic (n, %)                                |  |   |                              |  |                                |
|  |  |   |                              |  |                                |

| <b>Healthcare utilization</b>  |  |  |  |  |  |
|--|--|--|--|--|--|
| <b><i>In 12 months before start of influenza season</i></b>              |  |  |  |  |  |
| General practice office visits, median (IQR) <sup>e</sup>                |  |  |  |  |  |
| Medical specialty office visits, median (IQR)                            |  |  |  |  |  |
| Urgent care visits   |  |  |  |  |  |
| Emergency department visits, median (IQR)                                |  |  |  |  |  |
| Inpatient admissions, median (IQR)                                       |  |  |  |  |  |
| <b><i>During influenza season</i></b>                                    |  |  |  |  |  |
| General practice office visits, median (IQR)                             |  |  |  |  |  |
| Medical specialty office visits, median (IQR)                            |  |  |  |  |  |
| Urgent care visits   |  |  |  |  |  |
| Emergency department visits, median (IQR)                                |  |  |  |  |  |
| Inpatient admissions, median (IQR)                                       |  |  |  |  |  |
|  |  |  |  |  |  |
| <b>Influenza-related outcomes during the influenza season</b>            |  |  |  |  |  |
| Frequency of respiratory-associated hospitalizations (n, %) <sup>f</sup> |  |  |  |  |  |

<sup>a</sup> Care encounters include telephone-based care visits, an office-based ambulatory encounter, an urgent care encounter, or an emergency department encounter with the health system.

<sup>b</sup> Ambulatory care encounters include office-based visits to a general practice office (including but not limited to general pediatrics, general internal medicine, obstetrics/gynecology, or general family medicine), or office-based visits to a medical specialty office (including but not limited to cardiology clinic, diabetes clinic, rehabilitation clinic, gastrointestinal specialist, etc.).

<sup>c</sup> General practice encounters include office-based visits to a general practice office (including but not limited to general pediatrics, general internal medicine, obstetrics/gynecology, or general family medicine).

<sup>d</sup> Only applicable to “closed” health systems where individuals enroll as members and enrollment is tracked.

<sup>e</sup> Median number of visits per individual, with interquartile range (25<sup>th</sup>–75<sup>th</sup> percentiles).

<sup>f</sup> Respiratory-associated hospitalizations should include any qualifying admissions to an inpatient ward for at least 24 hours, or admissions where the discharge date is at least one calendar day after the admission date, if time in hours is not available. Qualifying respiratory admissions include those with a discharge code, in any positions, of 460-519 (for ICD-9 codes) or J00-J99 (for ICD-10 codes).

## Appendix D: Diagnosis Codes and Other Variables Used to Define for High Risk Conditions, ARI Hospitalizations, and Study Outcome

This appendix includes example conditions and codes used to define high-risk conditions and outcomes. These codes may be modified to include additional conditions or diagnoses as objectives are refined. This list of codes can also be found in the data dictionary referenced in Appendix B.

### High Risk Conditions

| Description   | ICD-9 Code   | ICD-10 Code  |
|---|--|--|
| Blood disorders   | 282.*, 283.*, 284*   | D55.*, D56.0, D56.1, D56.2, D56.4, D56.5, D56.9, D57.0, D57.1, D57.2, D57.4, D57.8, D58.*, D59.*, D60.*, D61.*, D64.0, D64.1, D64.2, D64.3, D64.4, D64.8, D65.*, D66.*, D67.*, D68.*   |
| Cerebrovascular disease                                       | 430, 431.*, 432.*, 433.*, 434.*, 435, 436, 437.*, 438.*, 348   | I60.*, I61.*, I62.*, I63.*, I68.*, I69.*   |
| Chronic lung disease  | 031.0*, 135.*, 277.0*, 273.4, 277.6*, 494.*, 495.*, 500.*, 501.*, 502.*, 503.*, 504.*, 505.*, 506.*, 508.*, 510, 513, 514.*, 515.*, 516.*, 517.*, 518.1*, 518.2*, 518.3*, 518.6*, 518.83, 519.9*, 519.0*, 748.4*, 748.5*, 748.6*, 759.3*, 770.2*, 770.7*, V42.6*, 714.81, 769, 770 | D86.0, E84.*, E88.01, J47.*, J60.*, J61.*, J62.*, J63.*, J64.*, J65.*, J66.*, J67.*, J68.*, J69.*, J70.*, J81.*, J82.*, J84.*, J95.0, J96.*, J98.1, J99.*, P25.*, P26.*, P27.*, P28.*, Q33.*, T86.3, T86.8, Z94.2  |
| Chronic lung disease (asthma)                                 | 493.*  | J45.*  |
| Chronic lung diseases (chronic obstructive pulmonary disease) | 491.*, 492.*, 496.*  | J40.*, J41.*, J42.*, J43.*, J44.*  |
| Chronic lung disease (endemic mycoses)                        | 115.*, 116, 116.1, 484.6, 117.3, 518.6, 117.5, 321, 117.7  | B39.*, B40.*, B41.*, B44.*, B45.*, B46.0   |
| Chronic lung disease (pulmonary tuberculosis)                 | 011.*, 012.*   | A15.*, A31.0   |
| Heart disease   | 393.*, 394.*, 395.*, 396.*, 397.*, 398.*, 402.*, 404.*, 415, 416.*, 417.*, 423.*, 424.*, 425.*, 426.*, 427.1*, 427.2*, 427.3*, 427.4*, 427.5*, 427.6*, 427.8*, 429.*, 440.*, 446.*, V42.1*, V45.0*, V45.81, V45.82, 9971, 441.*, 442.*, 443.*, 444.*, 447.*, 4461                  | I01.*, I02.*, I05.*, I06.*, I07.*, I08.*, I09.*, I11.*, I13.*, I26.*, I27.*, I28.*, I31.*, I34.*, I35.*, I36.*, I37.*, I41.*, I42.*, I43.*, I44.*, I46.*, I48.*, I51.*, I52.*, I97.0, I97.1, M31.*, Z94.1, Z95.*, Z98.61, I71.*, I72.*, I73.*, I74.*, I75.*, I79.* |
| Heart disease (congenital heart disease)                      | 745.*, 746.*, 747.*, 759.3   | Q20.*, Q21.*, Q22.*, Q23.*, Q24.*, Q25.*, Q26.*, Q27.0, Q27.3, Q27.4, Q27.8, Q27.9, Q28.*, Q89.3   |
| Heart disease (heart failure)                                 | 428.*  | I50.*  |

| <b>Description</b>                     | <b>ICD-9 Code</b>  | <b>ICD-10 Code</b>  |
|--|--|---|
| Heart disease (ischemic heart disease) | 410.*, 411.*, 412.*, 414.*   | I21.*, I22.*, I23.*, I24.*, I25.*   |
| Immunosuppression                      | 042.*, 43, 44, 136.3*, 279.*, 288.0*, 288.1*, 288.2*, 288.4*, 446.*, 710.0*, 710.1*, 710.2*, 710.3*, 710.4*, 714.*, V08.*, V42.2*, V42.7*, V42.8*, V42.9*, V42.1*, V42.6*, V42.0*, V424, V58.0*, V58.1*, 996.8, 710, 963.1   | B20.*, B59.*, B97.3, D47.Z1, D70.*, D71.*, D72.*, D73.*, D76.*, D80.*, D81.*, D82.*, D83.*, D84.*, D89.*, M05.*, M06.*, M07.*, M08.*, M30.*, M31.*, M32.*, M33.*, M34.*, M35.0, M35.9, Q89.0, Z21*, Z48.2, Z51.0, Z51.1, Z94.*, Z79.5, Z79.82   |
| Immunosuppression (malignancy)         | 140.*, 141.*, 142.*, 143.*, 144.*, 145.*, 146.*, 147.*, 148.*, 149.*, 150.*, 151.*, 152.*, 153.*, 154.*, 155.*, 156.*, 157.*, 158.*, 159.*, 160.*, 161.*, 162.*, 163.*, 164.*, 165.*, 166.*, 167.*, 168.*, 169.*, 170.*, 171.*, 172.*, 174.*, 175.*, 176.*, 177.*, 178.*, 179.*, 180.*, 181.*, 182.*, 183.*, 184.*, 185.*, 186.*, 187.*, 188.*, 189.*, 190.*, 191.*, 192.*, 193.*, 194.*, 195.*, 196.*, 197.*, 198.*, 199.*, 200.*, 201.*, 202.*, 203.*, 204.*, 205.*, 206.*, 207.*, 208.* | C00.*, C01.*, C02.*, C03.*, C04.*, C05.*, C06.*, C07.*, C08.*, C09.*, C10.*, C11.*, C12.*, C13.*, C14.*, C15.*, C16.*, C17.*, C18.*, C19.*, C20.*, C21.*, C22.*, C23.*, C24.*, C25.*, C26.*, C30.*, C31.*, C32.*, C33.*, C34.*, C37.*, C38.*, C39.*, C40.*, C41.*, C43.*, C44.*, C45.*, C46.*, C47.*, C48.*, C49.*, C4A.*, C50.*, C51.*, C52.*, C53.*, C54.*, C55.*, C56.*, C57.*, C58.*, C60.*, C61.*, C62.*, C63.*, C64.*, C65.*, C66.*, C67.*, C68.*, C69.*, C70.*, C71.*, C72.*, C73.*, C74.*, C75.*, C76.*, C77.*, C78.*, C79.*, C7A.*, C7B.*, C80.*, C81.*, C82.*, C83.*, C84.*, C85.*, C86.*, C88.*, C90.*, C91.*, C92.*, C93.*, C94.*, C95.*, C96.*, D03.*, D46.* |
| Liver disease                          | 571.*, 572.1*, 572.2*, 572.3*, 572.4*, 572.5*, 572.6*, 572.7*, 572.8*  | B18.*, I81.*, I85.*, K70.*, K71.*, K72.*, K73.*, K74.*, K75.*, K76.*, K77.*   |
| Metabolic disease                      | 255.*, 270.*, 271.*, 272.*, 277.*, 278.00, 278.01  | E00.*, E01.*, E03.*, E05.*, E06.*, E15.*, E16.*, E20.*, E21.*, E22.*, E23.*, E24.*, E25.*, E26.*, E27.*, E28.*, E29.*, E31.*, E32.*, E34.*, E66.01, E66.2, Z68.4, E70.*, E71.*, E72.*, E74.*, E75.2, E76.*, E77.*, E78.*, E79.*, E80.*, E83.*, E85.*, E88.*   |
| Metabolic disease (diabetes mellitus)  | 250*, 251.* 357.2*, 362.0*, 362.11, 366.41   | E08.*, E10.*, E11.*, E13.*  |

| Description                            | ICD-9 Code   | ICD-10 Code   |
|--|--|---|
| Neurological/Musculoskeletal disorders | 290.*, 294.1*, 318.1*, 318.2*, 330.*, 331.*, 332.*, 333.0*, 333.4*, 333.5*, 333.6*, 333.7*, 333.8*, 333.9*, 334.*, 335.*, 336.*, 340.*, 341.*, 342.*, 343.*, 345, 344.0*, 358.0*, 358.1*, 359.*, 756.1*, 756.3*, 756.4*, 756.5*, 756.6*, 768.5*, 780.3, 780.72 | H49.81, M12.0, M36.0, E75.02, E75.19, E75.4, F01.*, F02.*, F03.*, F71.*, F72.*, F73.*, F84.2, G10.*, G11.*, G12.*, G13.*, G14.*, G20.*, G21.*, G23.*, G24.*, G25.*, G26.*, G30.*, G31.*, G32.*, G35.*, G36.*, G37.*, G40.*, G45.*, G46.*, G60.*, G61.*, G62.*, G63.*, G64.*, G70.*, G71.*, G73.*, G80.*, G81.*, G82.*, G83.*, G90.3, G91.*, G93.*, G94.*, G95.*, G99.2, P91.*, Q00.*, Q01.*, Q02.*, Q03.*, Q04.*, Q05.*, Q06.*, Q07.*, Q76.*, Q77.*, Q78.*, Q79.*, Q85.*, Q87.4, Q90.*, Q91.*, Q92.*, Q93.*, Q96.*, R41.*, R53.2, R54.* |
| Renal disease                          | 285.21, 403.*, 581.*, 582.*, 583.*, 585.*, 587.*, 588.0*, 588.1*, 590.0* 593.8, V42.0*, V56.*, V45.1*  | I12.*, N01.*, N02.*, N03.*, N04.*, N05.*, N06.*, N07.*, N08.*, N11.*, N14.*, N15.*, N16.*, N18.*, N25.*, N26.*, N28.*, Q27.1, Q27.2, Q60.*, Z49.*, Z91.15, Z94.0, Z99.2   |

## Respiratory Outpatient Visits and Hospitalizations

Respiratory outpatient visits and hospitalizations will be defined using the following ICD-9 and ICD-10 diagnosis codes. Codes can be in any position. For hospitalizations, only discharge codes should be considered. Encounters occurring in the outpatient clinic, emergency department, urgent care, or via telehealth should additionally include encounters with codes related to signs/symptoms (at the bottom of the table).

| <b>Description</b>   | <b>ICD-10*<br/>code</b> |
|--|-------------------------|
| Pulmonary anthrax  | A22.1*                  |
| Whooping cough   | A37*                    |
| Other specified sepsis   | A41.89                  |
| Legionnaire's Disease  | A48.1*                  |
| Cytomegaloviral pneumonitis  | B25.0*                  |
| Viral infection of unspecified site  | B34 *                   |
| Aspergillosis  | B44 *                   |
| Adenovirus   | B97.0*                  |
| Enterovirus  | B97.1*                  |
| Coronavirus  | B97.2*                  |
| Respiratory syncytial virus as the cause of diseases classified elsewhere                | B97.4*                  |
| Other viral agents as the cause of diseases classified elsewhere                         | B97.8*                  |
| Acute suppurative otitis media   | H66.0*                  |
| Suppurative otitis media, unspecified  | H66.4*                  |
| Otitis media, unspecified  | H66.9*                  |
| Otitis media in diseases classified elsewhere  | H67 *                   |
| Diseases of the respiratory system   | J00-J99 *               |
| Other viral diseases complicating pregnancy, childbirth and the puerperium               | O98.5*                  |
| Diseases of the respiratory system complicating pregnancy, childbirth and the puerperium | O99.5*                  |
| COVID-19   | U00                     |
| COVID-19   | U07.1                   |
| COVID-19   | U49                     |
| Encounter for observation for suspected exposure to biological agents ruled out          | Z03.81 *                |
| Encounter for observation for other suspected diseases and conditions ruled out          | Z03.89 *                |
| Encounter for screening for respiratory tuberculosis                                     | Z11.1 *                 |
| Encounter for screening for other bacterial diseases                                     | Z11.2 *                 |
| Encounter for screening for other viral diseases   | Z11.5 *                 |
| Encounter for screening for other viral diseases   | Z11.59 *                |
| Encounter for screening for other infectious and parasitic diseases                      | Z11.8 *                 |

|  |           |
|--|-----------|
| Encounter for screening for infectious and parasitic diseases, unspecified   | Z11.9 *   |
| Contact with and (suspected) exposure to other viral communicable diseases   | Z20.82 *  |
| Contact with and (suspected) exposure to other viral communicable diseases   | Z20.828 * |
| Contact with and (suspected) exposure to other communicable diseases   | Z20.89 *  |
| Contact with and (suspected) exposure to unspecified communicable disease  | Z20.9 *   |
| Mucocutaneous lymph node syndrome [Kawasaki]   | M30.3 *   |
| <b>Encounters occurring in the outpatient clinic, telehealth, emergency department, or urgent care should additionally capture encounters with these codes</b>           |           |
| Hemorrhage from respiratory passages, cough, abnormalities of breathing, pain in throat and chest, other symptoms/signs involving the circulatory and respiratory system | R04-R09 * |
| Fever  | R50 *     |
| Headache   | R51 *     |
| Malaise and fatigue  | R53 *     |
| Shock, unspecified   | R57.9 *   |
| Symptoms and signs specifically associated with systemic inflammation and infection  | R65 *     |
| Chills without fever   | R68.83 *  |

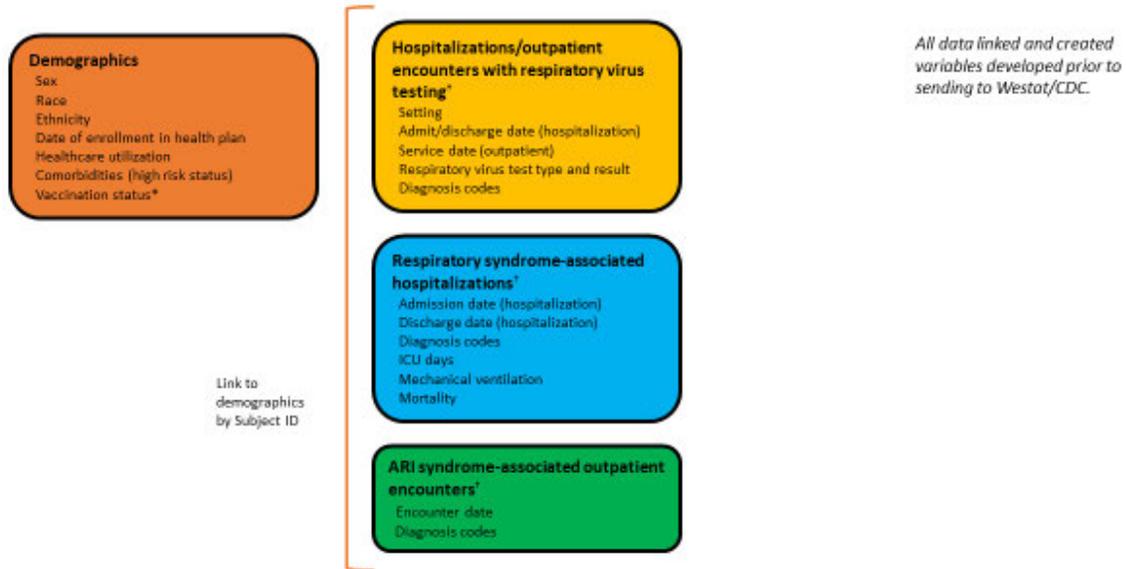
\*For the complete list these of codes, please refer to the data dictionary referenced in Appendix B.

## **Appendix E: 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 for each seasonal cohort and then share the cohort datasets with Westat & CDC. The proposed data structure would be one data package for each influenza season. For the primary objectives, data packages would be created for each season going back as early as the 2010–2011 season through the 2018–2019 influenza season. For secondary objectives, data packages would be created for the 2019–2020 influenza season and contemporary season, depending on whether the prospective data capture objective is exercised.

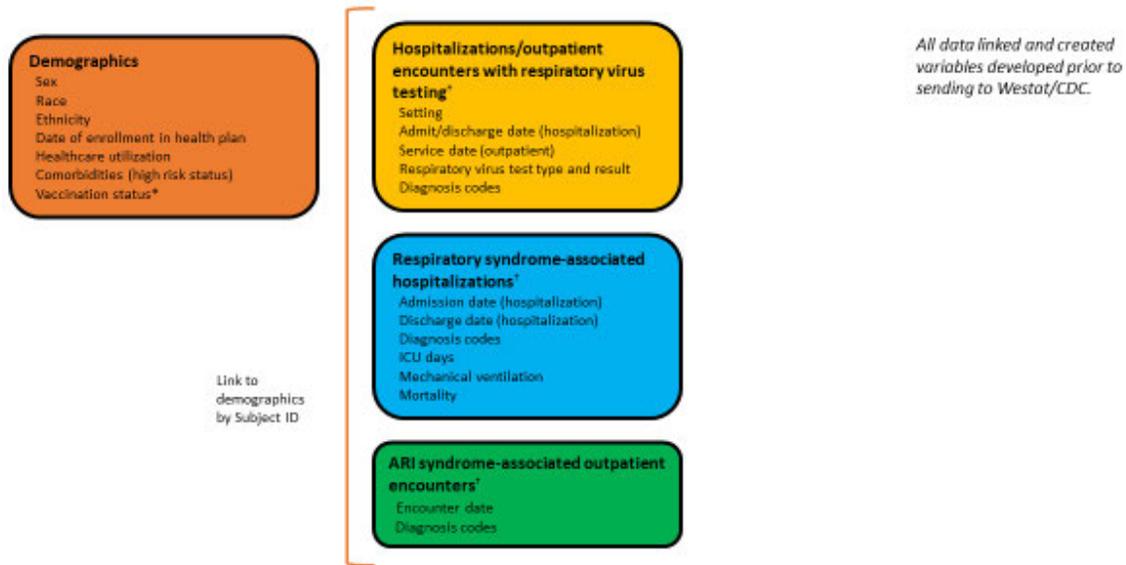
Each seasonal data package would include data as specified in Appendix B. Sites would be responsible for deriving data elements using the definitions from Appendix B and doing so prior to sending data to Westat.

Example 3: Pre-processed (by the study site) individual-level records for cohort are shared with Westat & CDC



\*For influenza and other vaccinations received dating back to 2010, where possible.  
 †For influenza season of interest.

Example 3: Pre-processed (by the study site) individual-level records for cohort are shared with Westat & CDC



\*For influenza and other vaccinations received dating back to 2010, where possible.  
 †For influenza season of interest.

## **Appendix F: Data Management Plan**

See accompanying document “Data Management Plan.”

## Appendix G: Suggestions for Site Data Validations using Primary Health Records

To ensure that extracted data reflect the underlying data from the primary health records, several validation steps are being requested, including review of a limited number of primary records. Below are suggestions for how many health records to review, how to stratify the sample that are reviewed, and how to report the results of the review.

- Perform primary record review on a minimum number of medical charts and compare the information in the primary record to the information captured through data extraction:
  - If data extraction methods or data storage platforms have changed during the study period, perform data validations for records from before and after the date when the change occurred.
  - If multiple health organizations contribute data to the site, perform data validations for each of the contributing organizations with differing data extraction methods or data storage platforms.
- A minimum number of charts should be reviewed for each of the following groups:
  - Unvaccinated for influenza (per registry, EMR or billing data)
  - Vaccinated for influenza (per registry, EMR or billing data)
  - Tested for influenza (per EMR or laboratory data stream)
  - Not tested for influenza (per EMR or laboratory data stream)

Sites who have participated in multi-site vaccination studies and have validation data from those studies can omit the unvaccinated and vaccinated groups from the primary record review.

- At a minimum for each of the 4 strata, above, 20 charts should be reviewed — for a total of 80 charts.
  - If sites are conducting chart reviews from multiple hospitals or health systems or organizations because of differences in data streams, data extraction methods, or data storage platforms, then the minimum number of chart reviews per group can be 10.
  - For example, if the site consists of 3 health systems/organizations, then chart reviews should be done for each of the health systems. If 10 charts are reviewed per group in a health system, then 40 total charts will be reviewed for each health system. Across 3 health systems, the total minimum number of charts would be 120.
- For each group, randomly sample among children and adults (e.g. vaccine records for 10 unvaccinated children, 10 records for unvaccinated adults, 10 records for vaccinated children, and 10 records for vaccinated adults).
- For those tested and not tested for influenza, randomly sample among children and adults with outpatient encounters and children and adults with inpatient admissions (e.g. 5 charts for children seen in the outpatient setting who were tested for influenza, 5 charts for children in the inpatient setting who were tested for influenza, 5 charts for children

seen in the outpatient setting who were not tested for influenza, 5 charts for children in the inpatient setting that were not tested for influenza, and the same pattern among adults).

- Validate data using known cohort members from relevant influenza seasons. We will not require that validations be performed on all study seasons, however sites should perform validation on seasons that are representative of the typical data extraction and data storage structure for the entire study.
- Validate against the following data elements at a minimum (more if needed)
  - Vaccination status
  - Influenza testing done and test results, if testing was done
  - Death
  - ICU admission
  - Mechanical ventilation
- Sites will not be required to review a minimum number of charts per data element or to achieve a certain percentage concordance between data obtained from data pull vs chart review.
- Use your best judgement to decide whether additional charts need to be reviewed or additional review of the EMR data pull programming needs to take place based on results
- Summarize results from validation in tables; examples of Tables 1 and 2 are below. Include additional tables for each season/period or site that undergoes validation.

**TABLE 1. Characteristics of charts pulled for validation.**

| <b>Study season:</b>                                       |                          |
|--|--------------------------|
| <b>Characteristic</b>                                      | <b>Charts pulled (n)</b> |
| Pediatric outpatient encounters                            |                          |
| Adult outpatient encounters                                |                          |
| Pediatric inpatient admission                              |                          |
| Adult inpatient admissions                                 |                          |
| Among charts pulled for inpatient admissions, number with: |                          |
| ICU admission  |                          |
| Mechanical ventilation                                     |                          |
| In-hospital death  |                          |

**TABLE 2. Concordance between chart review and data pull.**

| <b>Data Element</b>    | <b>Cases Reviewed (n)</b> | <b>Percent Agreement with chart (n, %)</b> | <b>Notes</b> |
|------------------------|---------------------------|--|--------------|
| Influenza vaccination  |                           |  |              |
| Influenza testing done |                           |  |              |
| Influenza test result  |                           |  |              |
| ICU admission          |                           |  |              |
| Mechanical ventilation |                           |  |              |
| In-hospital death      |                           |  |              |

**Additional Notes:**

## References

- <sup>1</sup> <https://www.cdc.gov/flu/about/burden/index.htm>
- <sup>2</sup> Rolfes, M. A., Foppa, I. M., Garg, S., Flannery, B., Brammer, L., Singleton, J. A., Burns, E., Jernigan, D., Olsen, S. J., Bresee, J., & Reed, C. (2018). Annual estimates of the burden of seasonal influenza in the United States: A tool for strengthening influenza surveillance and preparedness. *Influenza and other respiratory viruses*, 12(1), 132–137. <https://doi.org/10.1111/irv.12486>
- <sup>3</sup> Nair, H., Brooks, W. A., Katz, M., Roca, A., Berkley, J. A., Madhi, S. A., Simmerman, J. M., Gordon, A., Sato, M., Howie, S., Krishnan, A., Ope, M., Lindblade, K. A., Carosone-Link, P., Lucero, M., Ochieng, W., Kamimoto, L., Dueger, E., Bhat, N., Vong, S., ... Campbell, H. (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)
- <sup>4</sup> <https://www.cdc.gov/flu/about/burden/2018-2019.html>
- <sup>5</sup> <https://gis.cdc.gov/grasp/fluview/FluHospChars.html>
- <sup>6</sup> <https://gis.cdc.gov/grasp/fluview/FluHospChars.html>
- <sup>7</sup> Tiffany A Walker, Ben Waite, Mark G Thompson, Colin McArthur, Conroy Wong, Michael G Baker, Tim Wood, Jennifer Haubrock, Sally Roberts, Diane K Gross, Q Sue Huang, E Claire Newbern, Risk of Severe Influenza Among Adults With Chronic Medical Conditions, *The Journal of Infectious Diseases*, Volume 221, Issue 2, 15 January 2020, Pages 183–190, <https://doi.org/10.1093/infdis/jiz570>
- <sup>8</sup> Jackson, M. L., Neuzil, K. M., Thompson, W. W., Shay, D. K., Yu, O., Hanson, C. A., & Jackson, L. A. (2004). The burden of community-acquired pneumonia in seniors: results of a population-based study. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 39(11), 1642–1650. <https://doi.org/10.1086/425615>
- <sup>9</sup> Grohskopf, L. A., Alyanak, E., Broder, K. R., Walter, E. B., Fry, A. M., & Jernigan, D. B. (2019). Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices - United States, 2019-20 Influenza Season. *MMWR. Recommendations and reports : Morbidity and mortality weekly report. Recommendations and reports*, 68(3), 1–21. <https://doi.org/10.15585/mmwr.rr6803a1>
- <sup>10</sup> <https://www.cdc.gov/flu/professionals/vaccination/effectiveness-studies.htm>
- <sup>11</sup> Sofia Arriola, C., El Omeiri, N., Azziz-Baumgartner, E., Thompson, M. G., Sotomayor-Proschle, V., Fasce, R. A., Von Horoch, M., Enrique Carrizo Olalla, J., Aparecida Ferreira de Almeida, W., Palacios, J., Palekar, R., Couto, P., Descalzo, M., María Roperó-Álvarez, A., & REVELAC-I network participants (2019). Influenza vaccine effectiveness against hospitalizations in children and older adults-Data from South America, 2013-2017. A test negative design. *Vaccine: X*, 3, 100047. <https://doi.org/10.1016/j.jvax.2019.100047>
- <sup>12</sup> Nation, M. L., Moss, R., Spittal, M. J., Kotsimbos, T., Kelly, P. M., & Cheng, A. C. (2020). Influenza vaccine effectiveness against influenza-related mortality in Australian hospitalized patients: a propensity score analysis. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, ciz1238. Advance online publication. <https://doi.org/10.1093/cid/ciz1238>
- <sup>13</sup> Pebody, R. G., Zhao, H., Whitaker, H. J., Ellis, J., Donati, M., Zambon, M., & Andrews, N. (2020). Effectiveness of influenza vaccine in children in preventing influenza associated

---

hospitalisation, 2018/19, England. *Vaccine*, 38(2), 158–164.

<https://doi.org/10.1016/j.vaccine.2019.10.035>

<sup>14</sup> Nichols, M. K., Andrew, M. K., Hatchette, T. F., Ambrose, A., Boivin, G., Bowie, W., Chit, A., Dos Santos, G., ElSherif, M., Green, K., Haguinet, F., Halperin, S. A., Iburguchi, B., Johnstone, J., Katz, K., Lagacé-Wiens, P., Langley, J. M., LeBlanc, J., Loeb, M., MacKinnon-Cameron, D., ... Serious Outcomes Surveillance Network of the Canadian Immunization Research Network (CIRN), the Toronto Invasive Bacterial Diseases Network (TIBDN) (2018). Influenza vaccine effectiveness to prevent influenza-related hospitalizations and serious outcomes in Canadian adults over the 2011/12 through 2013/14 influenza seasons: A pooled analysis from the Canadian Immunization Research Network (CIRN) Serious Outcomes Surveillance (SOS Network). *Vaccine*, 36(16), 2166–2175.

<https://doi.org/10.1016/j.vaccine.2018.02.093>

<sup>15</sup> Thompson, M. G., Pierse, N., Sue Huang, Q., Prasad, N., Duque, J., Claire Newbern, E., Baker, M. G., Turner, N., McArthur, C., & SHIVERS investigation team (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. <https://doi.org/10.1016/j.vaccine.2018.07.028>

<sup>16</sup> Castilla, J., Godoy, P., Domínguez, A., Martínez-Baz, I., Astray, J., Martín, V., Delgado-Rodríguez, M., Baricot, M., Soldevila, N., Mayoral, J. M., Quintana, J. M., Galán, J. C., Castro, A., González-Candelas, F., Garín, O., Saez, M., Tamames, S., Pumarola, T., & CIBERESP Cases and Controls in Influenza Working Group Spain (2013). Influenza vaccine effectiveness in preventing outpatient, inpatient, and severe cases of laboratory-confirmed influenza. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 57(2), 167–175. <https://doi.org/10.1093/cid/cit194>

<sup>17</sup> Thompson, M. G., Pierse, N., Sue Huang, Q., Prasad, N., Duque, J., Claire Newbern, E., Baker, M. G., Turner, N., McArthur, C., & SHIVERS investigation team (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. <https://doi.org/10.1016/j.vaccine.2018.07.028>

<sup>18</sup> Arriola, C., Garg, S., Anderson, E. J., Ryan, P. A., George, A., Zansky, S. M., Bennett, N., Reingold, A., Bargsten, M., Miller, L., Yousey-Hindes, K., Tatham, L., Bohm, S. R., Lynfield, R., Thomas, A., Lindegren, M. L., Schaffner, W., Fry, A. M., & Chaves, S. S. (2017). Influenza Vaccination Modifies Disease Severity Among Community-dwelling Adults Hospitalized With Influenza. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 65(8), 1289–1297. <https://doi.org/10.1093/cid/cix468>

<sup>19</sup> Jill M. Ferdinands, Lauren E. W. Olsho, Anna A. Agan, Niranjana Bhat, Ryan M. Sullivan, Mark Hall, Peter M. Mourani, Mark Thompson, Adrienne G. Randolph, Effectiveness of Influenza Vaccine Against Life-threatening RT-PCR-confirmed Influenza Illness in US Children, 2010–2012, *The Journal of Infectious Diseases*, Volume 210, Issue 5, 1 September 2014, Pages 674–683, <https://doi.org/10.1093/infdis/jiu185>

<sup>20</sup> Flannery, B., Reynolds, S. B., Blanton, L., Santibanez, T. A., O'Halloran, A., Lu, P. J., Chen, J., Foppa, I. M., Gargiullo, P., Bresee, J., Singleton, J. A., & Fry, A. M. (2017). Influenza Vaccine Effectiveness Against Pediatric Deaths: 2010–2014. *Pediatrics*, 139(5), e20164244. <https://doi.org/10.1542/peds.2016-4244>

- 
- <sup>21</sup> Madhi, S. A., Cutland, C. L., Kuwanda, L., Weinberg, A., Hugo, A., Jones, S., Adrian, P. V., van Niekerk, N., Treurnicht, F., Ortiz, J. R., Venter, M., Violari, A., Neuzil, K. M., Simões, E. A., Klugman, K. P., Nunes, M. C., & Maternal Flu Trial (Matflu) Team (2014). Influenza vaccination of pregnant women and protection of their infants. *The New England journal of medicine*, 371(10), 918–931. <https://doi.org/10.1056/NEJMoa1401480>
- <sup>22</sup> Thompson, M. G., Li, D. K., Shifflett, P., Sokolow, L. Z., Ferber, J. R., Kurosky, S., Bozeman, S., Reynolds, S. B., Odouli, R., Henninger, M. L., Kauffman, T. L., Avalos, L. A., Ball, S., Williams, J. L., Irving, S. A., Shay, D. K., Naleway, A. L., & Pregnancy and Influenza Project Workgroup (2014). Effectiveness of seasonal trivalent influenza vaccine for preventing influenza virus illness among pregnant women: a population-based case-control study during the 2010-2011 and 2011-2012 influenza seasons. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 58(4), 449–457. <https://doi.org/10.1093/cid/cit750>
- <sup>23</sup> Thompson, M. G., Kwong, J. C., Regan, A. K., Katz, M. A., Drews, S. J., Azziz-Baumgartner, E., Klein, N. P., Chung, H., Effler, P. V., Feldman, B. S., Simmonds, K., Wyant, B. E., Dawood, F. S., Jackson, M. L., Fell, D. B., Levy, A., Barda, N., Svenson, L. W., Fink, R. V., Ball, S. W., ... PREVENT Workgroup (2019). Influenza Vaccine Effectiveness in Preventing Influenza-associated Hospitalizations During Pregnancy: A Multi-country Retrospective Test Negative Design Study, 2010-2016. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America*, 68(9), 1444–1453. <https://doi.org/10.1093/cid/ciy737>
- <sup>24</sup> Shi, T., McAllister, D. A., O'Brien, K. L., Simoes, E., Madhi, S. A., Gessner, B. D., Polack, F. P., Balsells, E., Acacio, S., Aguayo, C., Alassani, I., Ali, A., Antonio, M., Awasthi, S., Awori, J. O., Azziz-Baumgartner, E., Baggett, H. C., Baillie, V. L., Balmaseda, A., Barahona, A., ... RSV Global Epidemiology Network (2017). Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *Lancet (London, England)*, 390(10098), 946–958. [https://doi.org/10.1016/S0140-6736\(17\)30938-8](https://doi.org/10.1016/S0140-6736(17)30938-8)
- <sup>25</sup> <https://www.cdc.gov/flu/weekly/fluviewinteractive.htm>
- <sup>26</sup> <https://www.cdc.gov/flu/fluview/index.htm>