Rapid Cycle Analysis (RCA) to monitor the safety of COVID-19 vaccines in near real-time within the Vaccine Safety Datalink

PROTOCOL SUMMARY

Version 1.0
February 24, 2021

Principal Investigators: Nicola Klein, MD, PhD[1], James Donahue, DVM, PhD[2], Eric Weintraub, MPH[3]

Project Team: Laurie Aukes[1], Joan Bartlett[1], Ed Belongia[2], Berwick Chan[1], Frank Destefano[3], Bruce Fireman[1], Julianne Gee[3], Kristin Goddard[1], Kayla Hanson[2], Burney Kieke[2], Catherine Lee[1], Ned Lewis[1], Dave McClure[2], Mike McNeil[3], Karen Nunley[1], Pat Ross[1], Erica Scotty[2], Tom Shimabukuro[3], Arnold Yee[1], Ousseny Zerbo[1]

VSD Co-Investigators: Jason Glanz[4], Simon Hambidge[5], Elyse Kharbanda[7], Martin Kulldorff[8], Bruno Lewin[6] Allison Naleway[9], Jen Nelson[10], Josh Williams[5], Stan Xu[6], Katherine Yih[11]

1. Kaiser Permanente Northern California, Vaccine Study Center, Oakland, CA
2. Marshfield Clinic Research Institute, Marshfield, WI
3. U.S. Centers for Disease Control and Prevention, Immunization Safety Office, Atlanta, GA
4. Kaiser Permanente Colorado, Denver, CO
5. Denver Health, Denver, CO
6. Kaiser Permanente Southern California, Los Angeles, CA
7. HealthPartners, Bloomington, MN
8. Brigham and Women's Hospital, Boston, MA
9. Kaiser Permanente Northwest, Portland, OR
10. Kaiser Permanente Washington, Seattle, WA
11. Harvard Pilgrim Health Care Institute, Boston, MA

Vaccine development is well underway for SARS-CoV-2, the virus that causes COVID-19. Multiple vaccines have been made available under emergency use authorization since December 2020 and more are on the way. Protocol development is complete for Rapid Cycle Analysis (RCA) to monitor the safety of COVID-19 vaccines in the Vaccine Safety Datalink (VSD). This document provides an overview of these plans.

Investigators & Participating Sites
The COVID-19 vaccine RCA will be co-led by Kaiser Permanente Northern California (KPNC PI: Nicky Klein) and Marshfield Clinic Research Institute (MCRI PI: Jim Donahue) in collaboration with investigators at the Centers for Disease Control and Prevention (CDC) and Harvard. All VSD infrastructure sites will participate in the COVID-19 vaccine RCA.

Objectives
1. To conduct near-real time safety surveillance for COVID-19 vaccines in the VSD:
    a. Using concurrent comparators (KPNC)
    b. Using historical comparators (MCRI)
2. To describe the uptake of COVID-19 vaccines over time in the VSD (KPNC)
3. To conduct long-term safety surveillance for COVID-19 vaccines in the VSD (MCRI)

Surveillance Population
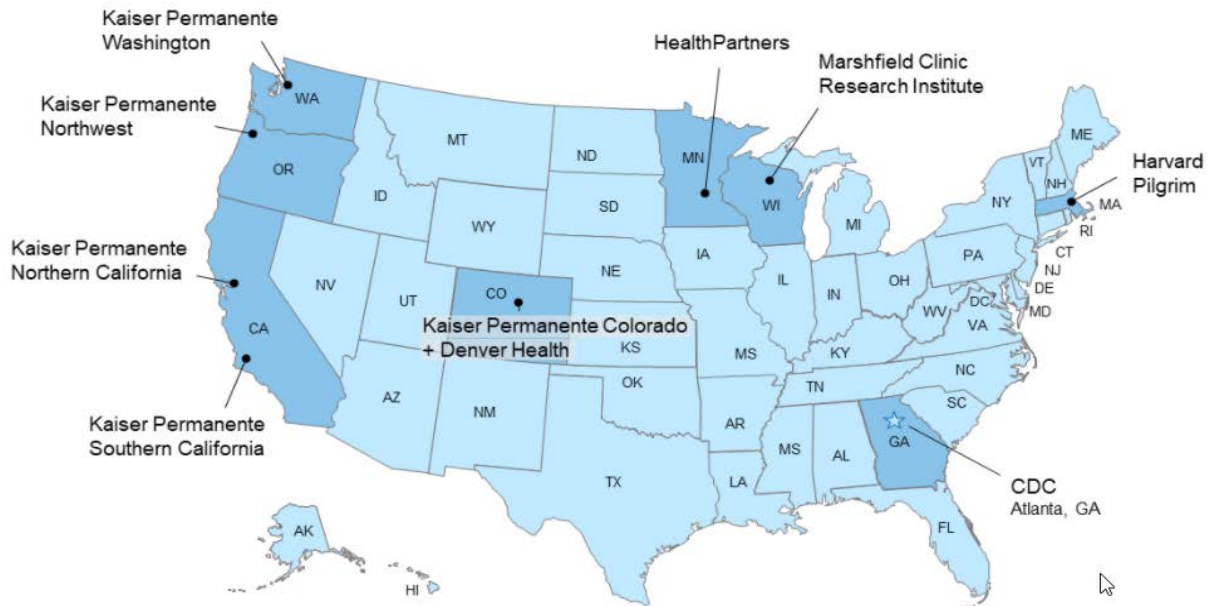
The VSD is a collaboration between the Immunization Safety Office at the CDC and nine integrated healthcare systems across the U.S. (https://www.cdc.gov/vaccinesafety/ensuringsafety/monitoring/vsd/index.html). Healthcare systems contribute data on their members and patients, creating a large population of individuals for whom near complete immunization and healthcare records are available. The VSD population is approximately 12 million people, or 3.6% of the U.S. population. The VSD population includes individuals across the age spectrum; about 20% are children and 16% are 65 years or older.

The VSD RCA surveillance population will include all current VSD members. One of the COVID-19 vaccines has initially been authorized for persons 16 years of age and above. Therefore, the initial surveillance population will consist of individuals who are ≥16 years old. As needed, this age range will be modified to stay consistent with the age groups receiving vaccination.

Exposure Classification
Exposure to COVID-19 vaccines will be identified by CVX codes in VSD vaccine data files. Multiple vaccine products for COVID-19 are available in the U.S., both 2-dose and 1-dose series vaccines (https://www.cdc.gov/vaccines/programs/iis/COVID-19-related-codes.html). We will ascertain vaccination date, product, manufacturer, and dose number for each exposure.

Objective 1: To conduct near-real time safety surveillance for COVID-19 vaccines in the VSD using concurrent and historical comparators.

Duration of Surveillance
Near real-time surveillance will begin in late December following availability of COVID-19 vaccines at VSD sites. When surveillance will end is uncertain at this time, but we currently plan to conduct surveillance for approximately 2 years.

Pre-Specified Outcomes
We identified a priority list of 21 outcomes, developed case ascertainment algorithms, and defined post-vaccination risk and comparison intervals (Table 1). This list was developed and refined in consultation with CDC, the VSD RCA working group, and in coordination with other federal agency stakeholders (i.e., Department of Defense [DoD], Food and Drug Administration [FDA], and Veterans Affairs [VA]). Outcomes that are more serious in nature and would present in the emergency department or inpatient setting are the focus. Case ascertainment criteria for all outcomes were reviewed for accuracy and completeness by relevant subject matter expert clinicians. As further results of phase III clinical trials become available or we become aware of additional potential vaccine safety concerns (e.g., from the Vaccine Adverse Event Reporting System [VAERS]), the outcome list will be updated as appropriate in consultation with investigators at CDC, FDA, and other federal partners. The sequential analytic approach is designed to be flexible to allow for the inclusion of new outcomes during the surveillance period.

Medically attended outcomes will be identified from International Classification of Disease, 10th Revision (ICD-10), diagnosis codes in VSD data files. Post-vaccination for most outcomes, we will evaluate a primary risk interval of 1–21 days and a secondary risk interval of 1–42 days. We will also monitor day 0 (day of vaccination) outcome events, but these will not be included in any planned comparative analyses. To avoid double-counting events in the same risk interval, we will include only the first event that occurred during the risk interval. However, if different events occurred on the same date, all such events will be counted.

Four outcomes will only be monitored – there will be no comparisons of observed-versus-expected events in the risk interval (with or without sequential testing) for acute respiratory distress syndrome (ARDS), anaphylaxis, narcolepsy, and multisystem inflammatory syndrome in adults and children (MIS-A and MIS-C). For post-vaccination MIS-A/MIS-C, ARDS, and

narcolepsy we will extract all events up to 84 days post-vaccination for monitoring purposes. Anaphylaxis will be monitored in the 0–1 days post vaccination.

Table 1: Outcomes of Interest

Outcomes for comparative analyses [1–3]	ICD-10 code(s)	Setting[4]
Acute disseminated encephalomyelitis (ADEM)	G04.00, G04.02	E, I
Acute myocardial infarction (AMI)	I21.*	E, I
Appendicitis	K35.*, K36, K37, K38.8	E, I
Bell's palsy	G51.0	E, I, O
Convulsions / seizures	R56.0*, R56.9	E, I
Disseminated intravascular coagulation (DIC)	D65	E, I
Encephalitis / myelitis / encephalomyelitis / (not ADEM or TM)	G04.30, G04.32, G04.39, G04.8*, G04.9*, G05.*, G37.4	E, I
Guillain-Barré syndrome (GBS)	G61.0	E, I
Immune thrombocytopenia (ITP)	D69.3	E, I, O
Kawasaki disease (KD)	M30.3	E, I
Myocarditis / pericarditis	B33.22, B33.23, I30.*, I40.*	E, I
Pulmonary embolism (PE)	I26.*	E, I
Stroke, hemorrhagic	I60.*, I61.*, I62.*	E, I
Stroke, ischemic	G45.8, G45.9, I63.*	E, I
Thrombotic thrombocytopenic purpura (TTP)	M31.1	E, I
Transverse myelitis (TM)	G37.3	E, I
Venous thromboembolism (VTE)	I26.*, I82.210, I82.220, I82.290, I82.3, I82.4*, I82.6*, I82.A1*, I82.B1*, I82.C1*, I82.81*, I82.890, I82.90	E, I, O

Outcomes for post-vaccination monitoring [1–3]	ICD-10 code(s)	Setting[4]
Acute respiratory distress syndrome (ARDS)	J80	E, I
Anaphylaxis	T78.2*, T80.52*, T88.6*	E, I
Multisystem inflammatory syndrome in children (MIS-C) / Multisystem inflammatory syndrome in adults (MIS-A)	M35.8 + U07.1 (new code as of 1-Jan-2021, M35.81)	E, I
Narcolepsy / cataplexy	G47.41*	E, I, O

[1] Primary risk interval of 1–21 days and a secondary risk interval of 1–42 days.

[2] "First in what period?" (to identify an incident diagnosis) is first since October 1, 2015 (the start of ICD-10 coding) for all outcomes except for myocarditis/pericarditis which is first in 60 days, and anaphylaxis which is not restricted to a 1st event.

[3] Appropriate exclusions were made for each outcome in consultation with relevant subject matter experts.

[4] E=Emergency Department, I=Inpatient, O=Outpatient.

Sequential Analyses

Sequential analyses will be conducted on an at least weekly basis for all pre-specified outcomes. Analyses will be stratified by several factors that may include product, dose, sex, age,

race/ethnicity and site and analyses will be conducted using both concurrent and historical comparators (Table 2). For all analyses, a statistical signal will not be considered stopping criteria.

Table 2. Current Overview of Planned Analyses[1]

Analytic Methods	Moderna			Pfizer			mRNA-Vaccines Combined		
	Dose 1	Dose 2	Combined	Dose 1	Dose 2	Combined	Dose 1	Dose 2	Combined
Vaccinated Concurrent Comparators	✓	✓	✓	✓	✓	✓	✓	✓	✓
Unvaccinated Concurrent Comparators	✓	✓	✓	✓	✓	✓	✓	✓	✓
Historical Background Rates	✓	✓	✓	✓	✓	✓	✓	✓	✓
Historical Well Care Visits	✓	✓	✓	✓	✓	✓	✓	✓	✓

[1]To be expanded as appropriate to account for new vaccines as approved and used within VSD.

Concurrent Comparator Analyses
KPNC will tabulate the cumulative incidence of the targeted outcomes during pre-specified post-vaccination risk intervals. For each outcome, the cumulative number of events observed in the risk interval will be compared to the number expected. The number expected will be derived from the three types of comparators described below, the first of which will be primary when available:
a) <u>vaccinated concurrent comparators</u> in a comparison interval after COVID-19 vaccination.
   — On each day, outcome incidence in the vaccinees who are in their risk interval will be compared to outcome incidence in <u>vaccinees</u> who are concurrently—on the same calendar date—in their comparison interval.
   — Analyses will be stratified in risk sets anchored to the calendar dates of outcome events.
   — Analyses will be stratified by age group, sex, race/ethnicity, and VSD site, as well as by calendar date.
b) <u>unvaccinated concurrent comparators</u> in a comparison interval after COVID-19 vaccination.
   — On each day, outcome incidence in the vaccinees who are in their risk interval will be compared to outcome incidence in <u>unvaccinated</u> individuals who are concurrently—on the same calendar date—at risk.
   — Analyses will be stratified in risk sets anchored to the calendar dates of outcome events.
   — Analyses will be stratified by age group, sex, race/ethnicity, and VSD site, as well as by calendar date.
c) <u>self-controls</u> in a comparison interval after COVID-19 vaccination.
   — Among the vaccinees who had an outcome event in either the risk interval or the comparison interval, we compare outcome incidence in the risk interval with outcome incidence in the comparison interval.

During the initial days of safety surveillance when vaccinations have only just begun, no vaccinees will yet be in a comparison interval, and so comparator types (a) and (c) will not be available yet. Before (a) becomes available, we plan to use (b) for our primary comparators. When (a) is available, (b) and (c) will be used for supplementary analyses.

Statistical Analysis and Rate Ratio Estimates
Poisson regression will be used to model outcome incidence observed during the risk interval in comparison with incidence expected (under the null hypothesis [H_0]). The log of the expected count of events in a risk or comparison interval in a stratum on a calendar day will be modelled as a function of whether the stratum's vaccinees are in a risk versus a comparison interval on that day. The analysis will be conditioned on age group sex, race/ethnicity, and VSD site, (which define the strata), as well as calendar day. Estimates of the rate ratio (RR) will be reported with nominal 95% confidence intervals (CI) rather than CI that are widened to correspond with the threshold of the sequential tests. That is, weekly CI will include all data to date, but will not account for the multiple chances for a false positive signal during surveillance. Trends in outcome incidence over calendar time and time since vaccination, and heterogeneity across subgroups will be tracked. Supplementary analyses will disaggregate the risk interval (days 1–7, days 8–-15, etc.), comparing risk interval weeks with each other and with the incidence expected (under H_0) from our primary comparator. We will also look at whether incidence varies across the weeks of the comparison interval.

Sequential Tests
Sequential analyses will be conducted as data are updated and analyzed weekly. For each outcome, the primary analysis each week will include a sequential test of the one-sided null hypothesis that the vaccine does not increase risk in the risk interval. The threshold for a signal is pre-specified by an alpha-spending plan that keeps the overall chance of a Type 1 error below 0.05 for 2 years (104 weekly analyses). With a Pocock style plan, the 2-sided p-value required for a signal at a weekly analysis is 0.0096, amounting to a 1-sided p-value of 0.0048. That is, the criteria for a signal at a weekly analysis will be a 1-sided p-value of <0.0048. The criteria for signaling are not criteria for "stopping". After a signal, weekly updates will continue as we evaluate the signal.

The multiplicity of different hypotheses tested will be taken into consideration informally. Our sequential testing adjusts for the multiplicity of weekly looks at each hypothesis, but we will not adjust formally for the multiplicity of hypotheses.

Power
The magnitude of the rate ratio (RR) detectable with 80% power will decrease as the number of outcomes expected in the risk interval (under H_0) increases. If our alpha-spending plan sets the threshold for a signal at 2-sided p = 0.0096 (amounting to 1-sided p = 0.0048) then RR = 2.0 is detectable when 32 outcomes are expected in the risk interval. RRs of 5, 4, 3, and 1.5 are detectable when 4, 6, 11, and 110 AEs, respectively, are expected in risk interval.

Table 3 below reports on the distribution of time-to-signal—the number of weekly analyses until we expect to have a 0.50, 0.80, or 0.95 probability of detecting a rate ratio of 1.5, 2.0, or 3.0.

Table 3. Time-to-Signal by the Rate Ratio and the N of Outcome Events Expected Weekly under H$_0$

		Week when chance of signal passes		
		50%	80%	95%
Rate Ratio	Outcomes/week expected in the risk interval			
1.5	1 (TTP)	55	101	
	20 (seizure)	4	6	9
	100 (AMI)	1	2	2
2.0	1 (TTP)	19	32	49
	20 (seizure)	2	2	3
	100 (AMI)	1	1	1
3.0	1 (TTP)	7	11	17
	20 (seizure)	1	1	1
	100 (AMI)	1	1	1

Historical Comparator Analyses

Historical comparator analyses will be most appropriate for infrequent or rare outcomes as using historical data accumulated over multiple years provides more stable estimates and greater statistical power, which potentially leads to earlier detection of a safety signal.

We will use two historical comparator groups to estimate the association between COVID-19 vaccination and pre-specified outcomes. One historical comparator group is the general VSD population with age and sex distributions that mirror those of COVID-19 vaccine recipients at the time of the analysis; this comparator group is used to estimate general background person-time rates. Expected counts are produced by multiplying the observed number of COVID-19 vaccine doses by the rates prorated to the length of the post-vaccination risk interval. We will examine the patterns of background rates across the historical period to determine if the incidence of specific outcomes has changed over time and if it is appropriate to use aggregate rates for the entire period versus a shorter time period, or if we need to project trends into the surveillance period.

The second historical comparator group will include persons having both 1) a well visit in the historical period (identified by ICD-10 codes Z00.0* and Z00.12*), and 2) an influenza vaccine in the 18 months prior to the well visit (identified by CVX codes). This comparator group addresses the concern that vaccinated and unvaccinated people may have systematic differences that could bias or confound the association between the outcomes and the vaccine. We hypothesize that this comparator group will have greater similarity to recipients of COVID-19 vaccines than the general VSD population. The date of the well visit will serve as the anchor for the post-visit risk interval within which outcome events are tabulated.

Data for the historical comparator groups will be derived from the VSD population during the period October 1, 2015 (the start of ICD-10 coding) through December 31, 2019 (preceding the start of the COVID-19 pandemic). we will also extract historical data for the pre-specified outcomes starting on January 1, 2020 and going forward at least through the end of 2020 to separately examine and describe the background rates and trends over this period. We will evaluate data for these periods and incorporate this assessment in addition to consultation with

our co-investigators when deciding how to characterize the background data in the sequential analyses.

Sequential Analysis

We will estimate relative risks (RR) for pre-specified outcomes analyzed sequentially using the Poisson-based maximized sequential probability ratio test (MaxSPRT) and the Conditional MaxSPRT (CMaxSPRT). We will use a Poisson-based MaxSPRT, developed by VSD researchers to compare observed number of events to expected number based on the historical background rates:

- Expected counts based on the incidence rate expected during the risk window multiplied by the number of vaccines administered.
- Reject H_0 of no excess risk if log-likelihood ratio exceeds a critical value → statistical signal.

Critical values will be based on probability of a false positive (e.g. α=0.05) and planned length of surveillance, defined in terms of expected counts under the null hypothesis. When the number of historical cases is small and background rates are unstable, we will use conditional MaxSPRT (CMaxSPRT).

The sequential methods used in this analysis will permit us to maintain an overall one-sided type I error rate of 0.05 across the multiple tests performed for each outcome, subgroup, and statistical method combination.

Power

The number of persons who will be administered the COVID-19 vaccine in VSD is difficult to predict due to issues related to availability, priority group distribution, and acceptance. The number of doses of COVID-19 vaccine needed to detect a range of relative risks (RR) ranges widely from a low of 10,000 to detect a RR of 5.0 for a comparatively common outcome (AMI) to many millions of doses to detect a RR of 1.5 for a rare outcome (GBS).

Assessment of the Lag in Automated Data Accrual

Data lags are a recognized characteristic of electronic medical record systems in which there is a delay between the time an encounter actually occurred and when information from the encounter is available in the automated data. We will include data accrual lag adjustments in the MaxSPRT and CMaxSPRT analyses that will preclude the need to delay initiation of the analyses and permit more timely signal identification. We will use separate data lag adjustment factors by site, setting, age group and week, where the latter corresponds to the number of weeks since the AE occurred (i.e., smaller values for week correspond to lower proportions of the AE appearing in the source data). This general approach to data lag adjustment is used in the ongoing influenza vaccine VSD RCA surveillance.

Covariates

Separate analyses will be conducted within strata defined by age group, sex, vaccine product, and dose. Covariates, in this context additional factors incorporated into the observed vs. comparator group comparisons, include VSD site and race/ethnicity. Comorbid conditions, presence of a high-risk condition, and healthcare utilization measures may be considered with input from CDC and the VSD RCA working group.

Potential Signal Assessment

If the criteria for a signal are met, we will already have the supplemental comparator analyses available for context and interpretation of results. It is assumed that any association that is detected with RCA may require additional evaluation to determine whether the signals are real or spurious.

This additional evaluation may include the following:

- Data quality assessment for errors, anomalies, or unusual patterns.
- Temporal scan statistics to determine if occurrences of the outcome are clustered within segments of the relevant risk interval following vaccination.
    - This will include examination of clusters of outcomes if they appear during brief sub-intervals within the risk interval, or if they appear to cluster in subgroups defined by age, sex, race/ethnicity, site, and known prior COVID-19 disease.
- Site-specific estimates may be computed (if possible) to determine if the association is consistent across the sites. If the association is driven by a particularly strong association at one site, additional analyses will be focused on that site (e.g., quality assessment of the data, chart review, etc.).
- Chart review of outcomes to confirm possible cases and collect additional data.
- Evaluation of the weekly routine supplementary analyses.

Associations that persist after evaluation will be communicated to the relevant stakeholders (e.g., reported to the Advisory Committee on Immunization Practices [ACIP]) and may be investigated by more standard epidemiological methods such as case-centered, case-control or cohort studies.

Objective 2: To describe the uptake of COVID-19 vaccines over time in the VSD

After a COVID-19 vaccine becomes routinely available in the VSD, we will monitor vaccine coverage overall, and in strata defined by age group, sex, race/ethnicity, and VSD site. For each week during the surveillance period, we will tabulate the number of doses delivered, the cumulative number of doses, and vaccine coverage. We will monitor vaccine coverage separately for each type of vaccine, and for all COVID-19 vaccines combined.

Objective 3: To conduct long-term safety surveillance for COVID-19 vaccines in the VSD

Following completion of near real-time surveillance, long-term surveillance will be conducted to continue monitoring the risk of pre-specified outcomes following COVID-19 vaccination. Aggregate data will continue to be collected on a weekly basis and formal, non-sequential analyses for select outcomes (to be determined) will be conducted with periodicity to be determined. In the event of a signal, individual-level data may be requested for a formal epidemiological analysis.

Data Collection, Quality & Management

Automated Data

Automated data will be extracted from standardized VSD files at each participating site. Sites that participate in VSD produce weekly dynamic data files (DDF) that capture information on demographics, immunizations, and ICD-coded diagnoses assigned by healthcare providers in outpatient, emergency, or hospital encounters. The DDF will be used as the primary data source for both Aims 1 and 2, including identifying and following vaccine recipients for outcomes of

interest. Cycle files updated on an annual basis may also be used to extract historical data. We will assess data quality in an ongoing manner to verify case and vaccination status for each outcome.

<u>Chart Review</u>
We will conduct routine chart review for selected rare outcomes - acute disseminated encephalomyelitis (ADEM), anaphylaxis, Guillain-Barré syndrome (GBS), and transverse myelitis (TM) - shortly after a case is detected and prior to analysis. If the criteria for a signal are met, further analyses of the vaccine-outcome association may be undertaken through chart review for any outcome, especially of the outcomes that occur in vaccinees during the risk interval. The goal of the chart reviews will be to verify the automated diagnosis, confirm the outcome as incident, and abstract additional information from the medical record not readily available via automated data such as symptom onset date and results of diagnostic testing.

Project Timeline and Dissemination of Results
This project commenced in fall 2020 and is expected to continue for approximately 3 years. Surveillance began as soon as there was COVID-19 vaccine uptake within the VSD, which occurred in mid-December 2020. Formal sequential analyses will commence when the COVID-19 vaccine is distributed in sufficient volume throughout the VSD population. Data extraction and chart review will occur throughout the surveillance period. Surveillance reports will be shared routinely with CDC and participating sites. Interim results may be presented to the ACIP, other federal partners, and/or published in CDC's Morbidity and Mortality Weekly Report (MMWR). Following completion of surveillance, final results will be presented to the ACIP and a manuscript will be prepared for submission to a peer-reviewed journal. Long-term surveillance is expected to continue for selected rare or high priority outcomes.