SENTINEL CBER SURVEILLANCE PROTOCOL

EVALUATION OF HPV9 (GARDASIL 9) VACCINE SAFETY SURVEILLANCE USING THE TREESCAN DATA MINING METHOD

Prepared by: W. Katherine Yih, PhD, MPH,[1] Judith C. Maro, PhD, MS,[1] Inna Dashevsky, MS,[1] Steven Anderson, PhD, MPP,[2] Meghan A. Baker, MD, ScD,[1] Adamma Mba-Jonas, MD, MPH,[2] Estelle Russek-Cohen, PhD,[2] Azadeh Shoaibi, PhD, MHS[2] Lihan Yan, PhD,[2] Martin Kulldorff, PhD[3]

Author Affiliations: 1. Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA; 2. Center for Biologics Evaluation and Research, Food and Drug Administration, Silver Spring, MD; 3. Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

June 30, 2016

Sentinel CBER Surveillance Protocol

Evaluation Of HPV9 (Gardasil 9) Vaccine Safety Surveillance Using The TreeScan Data Mining Method

Table of Contents

I. INTRODUCTION

As part of FDA's ongoing vaccine safety surveillance activities, this study seeks to provide postmarketing safety data for Gardasil 9 (9-valent human papillomavirus vaccine (types 6, 11, 16, 18, 31, 33, 45, 52, and 58), Merck & Co.) in Sentinel's Postlicensure Rapid Immunization Safety Monitoring (PRISM) system. The study represents the first component of a general approach to postmarketing vaccine safety surveillance taken by PRISM/Sentinel, which includes (1) surveillance for unexpected adverse events, (2) surveillance for suspected or biologically plausible pre-specified adverse events, and (3) highly customized protocol-based assessment of the association between a vaccine and a specified adverse event where, based on findings of prior studies, other information sources, or surveillance, concern about an association exists. The purpose of the first two components is to identify signs of possible safety concerns and then determine if an association is sufficiently strong as to merit a formal in-depth evaluation (the third component) or possibly some form of regulatory action. This general approach provides a framework for monitoring vaccine safety that can be applied to any new vaccine after licensure, although the framework will necessarily evolve as new methods and other features become available for use in PRISM/Sentinel.

Gardasil 9 was approved by FDA in December 2014 and covers five more HPV types than Merck's previously approved quadrivalent HPV vaccine, Gardasil. To avoid ambiguity, the terms "HPV9" and "HPV4" will be used to refer to these two Merck vaccines, respectively, for the remainder of this protocol; "HPV2" will refer to GlaxoSmithKline's Cervarix. In girls and women 9 through 26 years of age, HPV9 is indicated for the prevention of cancerous lesions of the cervix, vulva, vagina, and anus caused by HPV types 16, 18, 31, 33, 45, 52, and 58; certain precancerous lesions of the cervix, vulva, vagina, and anus caused by HPV types 6, 11, 16, 18, 31, 33, 45, 52, and 58; and genital warts caused by HPV types 6 and 11. In boys and men 9 through 26 years of age, HPV9 is indicated for the prevention of anal cancers caused by HPV types 16, 18, 31, 33, 45, 52, and 58; anal precancerous lesions caused by HPV types 6, 11, 16, 18, 31, 33, 45, 52, and 58; as well as genital warts caused by HPV types 6 and 11. It is administered in a 3-dose schedule, with the second dose administered 2 months after the first, and the third dose 6 months after the first. ACIP recommends routine vaccination of females and males aged 11–12 years against HPV with HPV9, HPV4, or HPV2 vaccine. Vaccination is recommended for females ages 13 through 26 years and for males ages 13 through 21 years who have not been previously vaccinated; the ACIP also recommends HPV vaccination through age 26 for men who have sex with men and for immunocompromised persons who have not been vaccinated previously.[1]

FDA approved HPV9 after reviewing results on safety, efficacy, and immunogenicity. HPV9's safety was evaluated in 7 clinical studies involving approximately 23,000 males and females ages 9 through 26 years, including approximately 15,700 subjects who received HPV9. These clinical studies found injection-site reactions to be higher among HPV9 recipients than among HPV4 recipients.[2] However, rates of systemic reactions, new-onset medical conditions, serious adverse events, and deaths following vaccination were comparable between HPV9 and HPV4 recipients. No serious safety issues were identified in prelicensure studies of HPV9.

With this project, we propose to add tree-based scan statistics to FDA's postmarketing safety monitoring methods for HPV9 to address the first component of the general safety surveillance approach outlined in the first paragraph. Tree-based scan statistics are a data mining method that can be used for vaccine and drug safety surveillance to look for any of a wide range of unsuspected but potential adverse reactions.[3-5] This method, which for brevity we will refer to by the name of the free

software product that employs it (www.treescan.org), "TreeScan," is typically used to simultaneously evaluate thousands of different outcomes, casting a wide net.[6] The main advantage is that otherwise unknown adverse reactions may be found. The main disadvantage is that it is not possible to adjust for all possible confounders, as they vary by outcome. This means that if there is a statistical alert generated by the TreeScan method that appears to be of public health concern, it must be carefully evaluated using other epidemiological study designs. No conclusion about causality should be based on the TreeScan analyses alone. In effect, the TreeScan method serves as a tool for identifying possible adverse reactions that merit further careful pharmacoepidemiological investigation.

In a pilot project, the TreeScan method, specifically the self-controlled tree-temporal scan statistic, was applied to PRISM/Sentinel's automated electronic health care data in order to evaluate whether there were any short term adverse events after HPV4 vaccination, without pre-specifying the type of adverse events or risk window *a priori*.[7] The tree-temporal method not only simultaneously evaluates several thousand health outcomes and groups of related health outcomes but also simultaneously evaluates a large number of *potential risk windows*, adjusting for the multiple testing inherent in the many health outcomes and risk windows examined. The main goal of the earlier pilot was to determine the suitability of the method for monitoring the safety of HPV9 in the current study and of vaccines administered to adolescents, in general. In the pilot, which used the conditional self-controlled tree-temporal scan statistic as the primary analysis, just two classes of alerts appeared. One was in the category of "cellulitis and abscess of the arm." Cellulitis is a known adverse event and is listed in the HPV4 package insert; thus, this was not investigated further. The other signal was in the more general category of "other complications of surgical and medical procedures." Approximately 90% of the 36 cases contributing to that signal appeared to have either conditions already identified as possible vaccine-associated adverse events or (in 3 cases) no specified symptoms but also no further medical visits until at least 60 days after the visit in which the incident diagnosis was identified. The other ~10% (4) of the cases had diverse symptoms, different in each case. The fact that only easily interpretable, not unexpected alerts were found provided reassurance not only about vaccine safety but also about the TreeScan conditional temporal-tree scan method.[8]

With the HPV4 pilot successfully concluded, we will use the same method for HPV9, running an analysis for the 18-month product review[i] and reviewing lists of diagnoses and procedures in claims data for case-patients contributing to any unanticipated alerts emerging from the analysis.

II.   SPECIFIC AIMS

1) To assess the safety of HPV9 vaccine in females and males ages 9 through 26 years, using the conditional self-controlled tree-temporal scan statistic, evaluating thousands of potential adverse events (Section IX)

---

[i] § 915 of *Food and Drug Administration Amendments Act of 2007, Public Law 110-85*..."[The Secretary via the FDA shall provide drug safety information to patients and prescribers by] preparing, by 18 months after approval of a drug or after use of the drug by 10,000 individuals, whichever is later, a summary analysis of the adverse drug reaction reports received for the drug, including identification of any new risks not previously identified, potential new risks, or known risks reported in unusual number;". Codified at 21 U.S.C. § 355(r)(2)(D).

2) To conduct an initial evaluation of any alerts detected, using clinical judgment to review summary statistics and electronically generated claims profiles of the patients contributing to the alert (Section XI)

III.   STUDY POPULATION AND FOLLOW-UP PERIOD

Data in the Sentinel Distributed Database will be obtained from four Sentinel sites: Aetna, HealthCore, Humana, and Optum. Each site will contribute data from 1/1/2006, or the earliest date that data are available thereafter, through the latest date of complete data available. (This early data start-date will allow ascertainment of prior vaccination with HPV4 and HPV2, needed for the secondary analyses (see Section IV).) For some Data Partners, there are restrictions on the use of patient-level data for certain members, even if the data are de-identified, and these members will be excluded.

As this is a self-controlled study, only vaccinated individuals will be included. We will analyze health outcomes of interest (HOIs) captured in the Days 1-56 follow-up period after HPV9 doses received by females and males on or after their 9th birthday and before their 27th birthday between the licensure date of 12/10/2014 and the latest date of complete data available, inclusive. (For simplicity, we will refer to including or excluding HPV9 "doses" from analysis, although what is meant is "HOIs in the follow-up period after HPV9 vaccination.") To be included in analysis, doses must also be preceded by at least 183 days of enrolled time (to allow determination of incidence—see Section VIII) and followed by at least 56 days of enrolled time. Enrollment gaps of 45 days or less will be bridged, i.e., treated as continuously enrolled time.[ii]

The minimum enrollment periods and the exposure and outcome incidence assessment windows discussed in later sections are illustrated in Figure 1:

[ii] Apparent gaps in enrollment can occur due to administrative glitches during annual renewals or switches from one plan to another within the same health insurance company. They may or may not reflect true lapses in coverage and may or may not cause health events during the gaps to be missed. The allowance of apparent gaps in enrollment of up to 45 days is standard practice in Sentinel studies and is applied uniformly regardless of whether the gap is in a pre-exposure or a post-exposure period. While it is possible that apparent gaps in enrollment might lead to health events (such as health outcomes after HPV9) not being ascertained, we think the possibility of the TreeScan results being noticeably affected by allowing apparent enrollment gaps of up to 45 days is remote. This sense is reinforced by the fact that most HPV vaccination occurs in August and most reenrollment in health insurance companies occurs several months later, in December-January. In sum, we think that the sample size gain is worth the likely minimal data errors that could be introduced by allowing apparent enrollment gaps of ≤ 45 days, especially considering that the aim is surveillance, not hypothesis-testing or precise risk estimation.
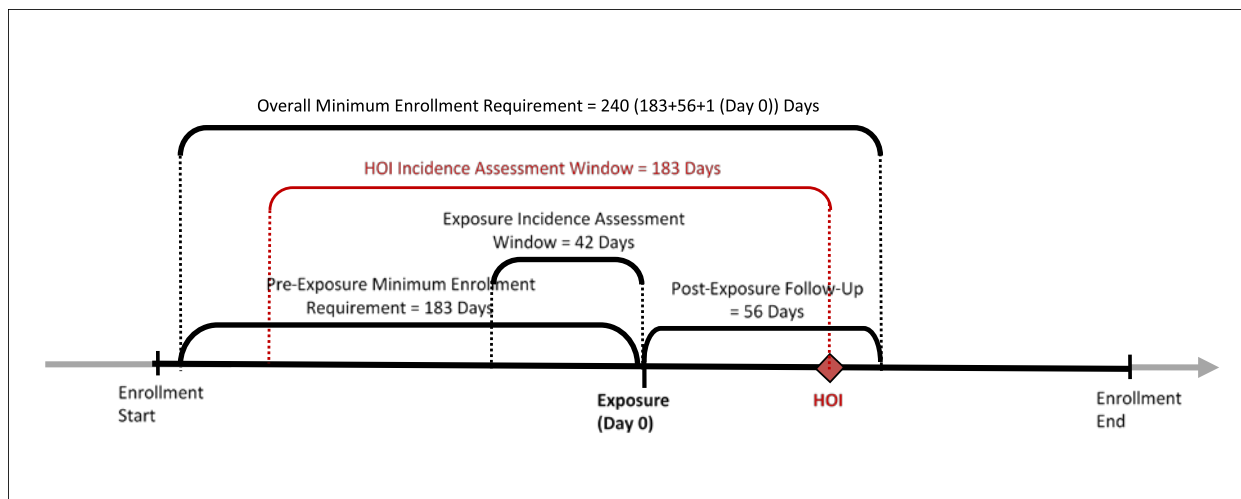
Figure 1.

Descriptive statistics from the PRISM Gardasil (HPV4)-Venous Thromboembolism study[9] indicate that only 5.7% of Dose 2s were given within 56 days of Dose 1. We do not expect dose-spacing patterns to be significantly different for HPV9. Hence, we do not expect overlap (within patients) of 56-day post-vaccination follow-up periods to be common or to affect the results in a major way.

IV. PRIMARY AND SECONDARY ANALYSES

The primary analysis will consider *all* doses of HPV9 without distinguishing among them. Doses of HPV9 occurring within 42 days of a previous dose of HPV9 will be excluded. This will avoid overlapping risk windows and is expected to have minimal impact on the number of doses included, considering that the descriptive statistics for the earlier HPV4 study mentioned above show that only 1.5% of Dose 2s were given within 42 days of Dose 1.

There will be two secondary analyses:

a) The first will consider *all* doses of HPV9 that were *not preceded by either HPV4 or HPV2* (Cervarix). As in the primary analysis, we will not distinguish among Doses 1, 2, or 3 of HPV9, and doses of HPV9 occurring within 42 days of a previous dose of HPV9 will be excluded.

b) The second will consider only *first* doses of HPV9 that were not preceded by either HPV4 or HPV2 (Cervarix).

To check for prior HPV vaccination in both the secondary analyses, we will look back from the HPV9 dose in question through the maximum amount of available enrolled time.

V. STUDY VACCINE

HPV9 vaccination will be identified by means of CPT code 90651. For the secondary analyses, we will also make use of the CPT codes 90649 (HPV4) and 90650 (HPV2), in order to identify and exclude HPV9 doses that were preceded by these other kinds of HPV vaccine. Because vaccinations are sometimes

entered into claims data as NDC codes, we will use those codes as well. Both the CPT and NDC codes for the three HPV vaccines are shown in Table 1, below:

Table 1.

Vaccine code	Code type	Code description	Manufacturer
90651	CPT	HPV9 (Gardasil 9)	Merck
00006411903	NDC	HUMAN PAPILLOMAVIRUS VACCINE, 9-VALENT/PF	Merck
00006412102	NDC	HUMAN PAPILLOMAVIRUS VACCINE, 9-VALENT/PF	Merck
00006411901	NDC	HUMAN PAPILLOMAVIRUS VACCINE, 9-VALENT/PF	Merck
00006411902	NDC	HUMAN PAPILLOMAVIRUS VACCINE, 9-VALENT/PF	Merck
00006412101	NDC	HUMAN PAPILLOMAVIRUS VACCINE, 9-VALENT/PF	Merck
90649	CPT	HPV4 (Gardasil)	Merck
00006404500	NDC	HUMAN PAPILLOMAVIRUS VACCINE, QUADRIVALENT/PF	Merck
00006404501	NDC	HUMAN PAPILLOMAVIRUS VACCINE, QUADRIVALENT/PF	Merck
00006404541	NDC	HUMAN PAPILLOMAVIRUS VACCINE, QUADRIVALENT/PF	Merck
00006410901	NDC	HUMAN PAPILLOMAVIRUS VACCINE, QUADRIVALENT/PF	Merck
00006410902	NDC	HUMAN PAPILLOMAVIRUS VACCINE, QUADRIVALENT/PF	Merck
00006410906	NDC	HUMAN PAPILLOMAVIRUS VACCINE, QUADRIVALENT/PF	Merck
00006410909	NDC	HUMAN PAPILLOMAVIRUS VACCINE, QUADRIVALENT/PF	Merck
00006410931	NDC	HUMAN PAPILLOMAVIRUS VACCINE, QUADRIVALENT/PF	Merck
54569582201	NDC	HUMAN PAPILLOMAVIRUS VACCINE, QUADRIVALENT/PF	Merck
90650	CPT	HPV2 (Cervarix)	GlaxoSmithKline
58160083011	NDC	HUMAN PAPILLOMAVIRUS VACCINE, BIVALENT/PF	GlaxoSmithKline
58160083032	NDC	HUMAN PAPILLOMAVIRUS VACCINE, BIVALENT/PF	GlaxoSmithKline
58160083043	NDC	HUMAN PAPILLOMAVIRUS VACCINE, BIVALENT/PF	GlaxoSmithKline
58160083046	NDC	HUMAN PAPILLOMAVIRUS VACCINE, BIVALENT/PF	GlaxoSmithKline
58160083052	NDC	HUMAN PAPILLOMAVIRUS VACCINE, BIVALENT/PF	GlaxoSmithKline
58160083034	ND	HUMAN PAPILLOMAVIRUS VACCINE, BIVALENT/PF	GlaxoSmithKline
58160083005	ND	HUMAN PAPILLOMAVIRUS VACCINE, BIVALENT/PF	GlaxoSmithKline

VI.   RISK AND COMPARISON WINDOWS

As mentioned in Section III, we will evaluate health outcomes occurring 1-56 days after vaccination only. The day of vaccination (Day 0) is not included since (1) a preventive care visit at which vaccines are given could generate diagnosis codes (outcomes) unrelated to vaccination, such as problems found during an eye examination, and (2) HPV9 may be given during a health care visit that happened due to an illness or other health concern. We will evaluate all temporal risk windows that are at least 2 days long, are at most 28 days long, start between 1 and 28 days after vaccination, and end between 2 and 42 days after vaccination. The comparison or control period consists of those days within the Days 1-56 follow-up period that are not in the risk window being evaluated.

VII.   HIERARCHICAL DIAGNOSIS TREE AND MAPPING ICD-10 TO ICD-9 CODES

Outcomes will be identified using ICD-9 and ICD-10 codes. For the TreeScan analysis, ICD-10 codes will be mapped to ICD-9 codes, as described at the end of this section.

All ICD-9 diagnoses are classified into a hierarchical tree structure defined by the Multi-Level Clinical Classification Software (MLCCS). The MLCCS is a product of the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project (http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp). It is a hierarchical system with four diagnosis levels, although on some branches there may only be two or three levels. The first and broadest level identifies 18 body systems, while the entries at the finest level contain one or multiple ICD-9 codes. For example, convulsions is a third level classification without a fourth level and for which there are five different ICD-9 codes, as shown in Table 2:

Table 2.

Node	Description
06	Diseases Of The Nervous System And Sense Organs
06.04	..Epilepsy; convulsions
06.04.02	....Convulsions
780.3	……..Convulsions
780.31	……..Febrile convulsions
780.32	……..Complex febrile convulsions
780.33	……..Post traumatic seizures
780.39	……..Other convulsions

In the hierarchical tree we will use, we have filled in all levels out to the fifth level for all outcomes, with the fifth level being the ICD-9 code. For example, for ICD-9 code 729.5, "pain in soft tissues of the limb," the first level of the tree 13 is "diseases of the musculoskeletal system and connective tissue" and the second level 13.08 is "other connective tissue disease." There is no finer differentiation of this outcome, but we created additional levels for it by adding zeroes—the third level is 13.08.00 and the fourth level is 13.08.00.00, and the fifth level is the ICD-9 code 729.5.

Using the same MLCCS tree as was used for the HPV4 pilot, some ICD-9 codes will be excluded from the tree and therefore from the analysis, for example, those representing:

- Outcomes that are very unlikely to be caused by vaccination, such as well-care visits, delivery of a baby, vitamin deficiencies, or fracture of a lower limb
- Some conditions unlikely to manifest themselves within the short follow-up time we are dealing with, such as cancer
- Most infectious diseases with an identified organism (e.g., typhoid fever, tuberculosis, shigella)
- Congenital conditions (e.g., sickle cell disease, congenital heart disease)
- Outcomes that are common and of an unspecific or less serious nature, such as fever, croup, and acute pharyngitis.

Because the surveillance period spans the ICD-9 and ICD-10 eras, we will map ICD-10 codes to ICD-9 codes, using the Centers for Medicare & Medicaid Services' General Equivalence Mappings (GEMs) and the rules below:

1. If an ICD-9 code was excluded from the tree used for the HPV4 pilot, exclude it from our current tree as well.
2. If an ICD-10 code maps to none of the included ICD-9 codes, exclude the ICD-10 code.

3. If an ICD-10 code maps to exactly 1 ICD-9 code and that ICD-9 code is included in the tree, include the ICD-10 code and map it to the ICD-9 code.

4. If an ICD-10 code maps to > 1 included ICD-9 code and to 0 excluded ICD-9 codes, include the ICD-10 code and map it to the most frequently used included ICD-9 code, according to the code-use frequencies in Harvard Pilgrim data. (Rationale: The most frequently used are usually the most general, e.g., "unspecified," and it seems more correct to map the ICD-10 code to an "unspecified" version of the outcome than to a specific one.)

5. If an ICD-10 code maps to ≥ 1 included ICD-9 code and to ≥ 1 excluded ICD-9 code, use clinical judgment to determine (a) whether or not to include the ICD-10 code at all, and (b) if it is to be included, to which included ICD-9 code it is to be mapped. (There are 844 such ICD-10 codes, out of 21,479 codes that map to any ICD-9 code(s) in our tree.)

The mapping scheme will be finalized prior to extracting the data and conducting the analyses.

VIII.   INCIDENT DIAGNOSES OF INTEREST

The study will focus on incident diagnoses observed during the 56-day follow-up period, since a repeat diagnosis may be due to a follow-up visit for an earlier episode of illness and is less likely to be due to the vaccine. A diagnosis is an incident diagnosis if it was observed in the inpatient or emergency department setting during the follow-up period and if there was no other diagnosis for that patient in the same third-level branch of the MLCCS diagnosis tree in any setting during the prior 183 days. This means that, even if there is a never-before-seen ICD-9 code, it is not counted if a different ICD-9 code belonging to the same third level branch was observed during the prior 183 days. Based on results of testing, the third level was chosen for determining incidence in order to avoid double-counting and overestimation of incidence which may occur when physicians classify the same episode of illness in two slightly different ways (e.g., convulsions and febrile convulsions) in separate patient visits. In synthesis, we consider the third level neither too coarse nor too fine.

Each patient can contribute multiple incident diagnoses during his/her follow-up period, as long as they are not part of the same third-level branch of the MLCCS tree. In the unusual situation were a patient has multiple incident diagnoses on the same third-level branch on the same day, the program will select the rarest incident outcome, using an outcome frequency list based on emergency department and inpatient data for 9-26.99 year olds in Harvard Pilgrim Health Care.

Appendix 1 illustrates these rules by means of two fictional example patients.

IX.   TREE-TEMPORAL SCAN STATISTIC

With the tree-temporal scan statistic, we are performing multiple temporal scan statistics, one for each of the many overlapping branches of the tree, adjusting for the multiple testing stemming both from the many branches and from the many time intervals evaluated. Each time interval is evaluated on each of the branches, so with our approximately 7300 nodes (i.e., outcome categories, whether first, second, third, fourth, or fifth level, which include, for example, the codes listed in Table 2 in Section VII) on the tree and our 665 potential time intervals, there would be more than 4.8 million potential clusters to evaluate and for which we would need to adjust for multiple testing. If these were 4.8 million independent tests with non-overlapping data, there would be a huge loss in power when

adjusting for all the multiple testing. With scan statistics, such a large loss in power does not happen, since many of the 4.8 million potential clusters are highly overlapping with each other. Hence, the penalty for adjusting for the multiple testing is more modest.

Considering the thousands of overlapping disease outcomes evaluated, adjustment for multiple testing is critical. This is accomplished through the simulation component of the method. The likelihood ratio test statistic from the most likely cut in the real dataset is compared with the likelihood ratio test statistics from the most likely cuts in each of, say, 999 random datasets, and we note its rank. For example, if it has the fifth highest test statistic, its rank is 5. Note that the most likely cut will be on a different branch in the different datasets, so we are not comparing the likelihood ratios for the same cut, but rather, comparing the maxima of the likelihood ratios obtained over all possible cuts. Since the random datasets were all generated under the null hypothesis, if the null hypothesis is true in the real dataset, then the test statistics come from exactly the same probability distribution. This means that, if the null hypothesis is true, the rank test statistic from the real dataset will range uniformly from 1 to 1000, and the probability of having a rank in the top 5% is exactly 5%. If the test statistic from the real dataset is in the top 5%, we will reject the null hypothesis; we have a 5% probability of falsely rejecting the null.

The tree-temporal scan statistic conditions the analysis on the number of cases observed in each node of the tree. This means that, unlike the standard tree-based scan statistic, there is no probability distribution to model the number of cases in each node, but rather, it is deterministic. What is probabilistic is the timing of each case. In an unconditional analysis, under the null hypothesis, the cases are assumed to be uniform across the follow-up period, but we will be using a conditional version where the analysis is also conditioned on the total number of cases occurring on the first day after vaccination, on the second day after vaccination, etc. This adjusts for the type of temporal confounding that would occur if there were some temporal differences in the general healthcare-seeking behavior shortly after compared to longer after the vaccination date. It also adjusts for day-of-week effects that are common to all outcomes. Under the alternative hypothesis, there is at least one branch of the tree for which there is a temporal cluster of cases during some time interval.

For each tree node and time interval, we calculate a Poisson generalized log likelihood ratio test statistic. Let n be the number of cases in the node, let c be the number of those node cases that are also in the time interval, let z be the number of cases in the time interval summed over the whole tree, and let C be the total number of cases in the tree. The number of cases in the cluster, c, is then contrasted with the expected number of cases in the cluster under the null hypothesis, which is $u=nz/C$. The test statistic is then

$$T = c \ln[c/u] + (C-c) \ln[(C-c)/(C-u)] \, I(c>u)$$

where $I()$ is the indication function. $I()$ is 1 when there are more cases than expected in the cluster, and 0 otherwise, and it is included to ensure that we are looking for an excess risk of having the outcome rather than a protective decreased risk.

For each node on the tree, the test statistic is calculated for each time interval under consideration. The node-interval combination with the maximum test statistic is the most likely cluster of cases, that is, the cluster that is least likely to have occurred by chance. Regardless of the data, there is always a most likely cluster, so that in itself does not mean that there is a true cluster.

Branches with zero events do not contribute to the analysis. Also, if there is only 1 case at the fifth level, that is, only one case with a specific ICD-9 code, no signal is possible for that specific ICD-9 code, although that case *can* contribute to a signal on one of the higher level branches.

The distribution of the test statistic is not known analytically, so there is no simple mathematical formula that can be used to obtain a p-value for the detected cluster. To evaluate whether the most likely cluster is statistically significant, after adjusting for the multiple testing inherent in the many node-interval combinations considered, Monte Carlo hypothesis testing is used in a one-sided test that only looks for excess risk at the alpha=0.05 level. We do this by generating 99,999 random replicates of the data. In each random data set, each node has exactly the same number of cases as the real data set, and each day after vaccination has the same number of cases when summed over all nodes. The only thing that varies is the pairing of the nodes and times, which is randomized using a permutation approach. For each random data set, generated under the null hypothesis, we find the most likely cluster in the same way as we did for the real data set, and we note the maximum test statistic of that data set. Note that the node and time interval for the most likely cluster will typically be different in each of the random data sets and also different from the real data. If the null hypothesis is true, then the maximum test statistic from the real data set has a 5% chance of being among the 5000 highest maximum test statistics from the one real and the 99,999 random data sets, so if that is the case, we can reject the null hypothesis at the alpha=0.05 level. If R is the rank of the maximum test statistic from the real data set so that there are exactly R-1 random data sets with a higher maximum, the Monte Carlo based p-value is R/(S+1), where S is the number random data sets used. Adjustment for multiple testing is assured since we are comparing the maximum from the real data set with the maxima from the random data sets.

The tree-temporal scan statistic can be applied with various analysis parameter settings, such as the risk window parameters discussed in Section VI. On the tree we will evaluate all nodes at the ICD-9 level, which is the fifth level of the tree, as well as nodes at the third and fourth level. We will not evaluate the first and second levels, as they are very broad categories.

X.   DATA FORMATS

A.   DATA FROM DATA PARTNER SITES

Separate data sets will be extracted from each Data Partner for the primary and secondary analyses. The data provided by the sites will consist of strata that include:

- ICD-9 or ICD-10 code entered by clinician or coder
- Diagnosis code type (i.e., ICD-9 or ICD-10)
- Number of days between exposure and outcome (all values will be in the range of 1-56)
- Number of cases

The rules for determining incident diagnoses will be applied at the sites, using the MLCCS tree to be provided, as this process requires access to patient-level data.

In addition, we will obtain the total number of HPV9 doses given in the age groups of interest. These are needed to calculate attributable risks.

B. DATA ANALYSIS FOR TREESCAN ANALYSIS

Any ICD-10 codes will be mapped to ICD-9 codes centrally, and a dataset for the tree-temporal scan analysis will be created, with strata that include the following three variables:
- ICD-9 code
- Number of days between exposure and outcome (all values will be in the range of 1-56)
- Number of cases

XI. STATISTICAL ALERT FOLLOW-UP

After the new programming and quality checks are complete, we will (1) execute the TreeExtraction program at the Data Partners, (2) run the TreeScan analysis, and (3) run the DataFreeze program at the Data Partners. The latter will freeze data behind the fire wall for patients that contributed adverse events in a node for which a statistically significant alert was generated.

All statistical alerts, if any arise, will undergo an initial assessment to categorize the alert as known, likely due to confounding or bias, or uncategorized (i.e., a candidate for further assessment). This initial assessment will be undertaken by the TreeScan work group, involving additional clinical expertise from CBER and/or the Sentinel Operations Center as needed. Any "uncategorized" alerts will be explored by generating claims profiles for the patients involved. The claims profile will contain the original diagnosis codes used, including ICD-10 codes. The existing Patient Episode Profile Retrieval (PEPR) system will be employed for this purpose, and the information will be displayed as a TreeScan Vaccine Episode Report (TVER). Members of the work group with clinical expertise will review and interpret the TVER(s).

These procedures are summarized in Figure 2, from the Sentinel report "Infrastructure for Evaluation of Statistical Alerts Arising from Vaccine Safety Data Mining Activities in Mini-Sentinel."[10]
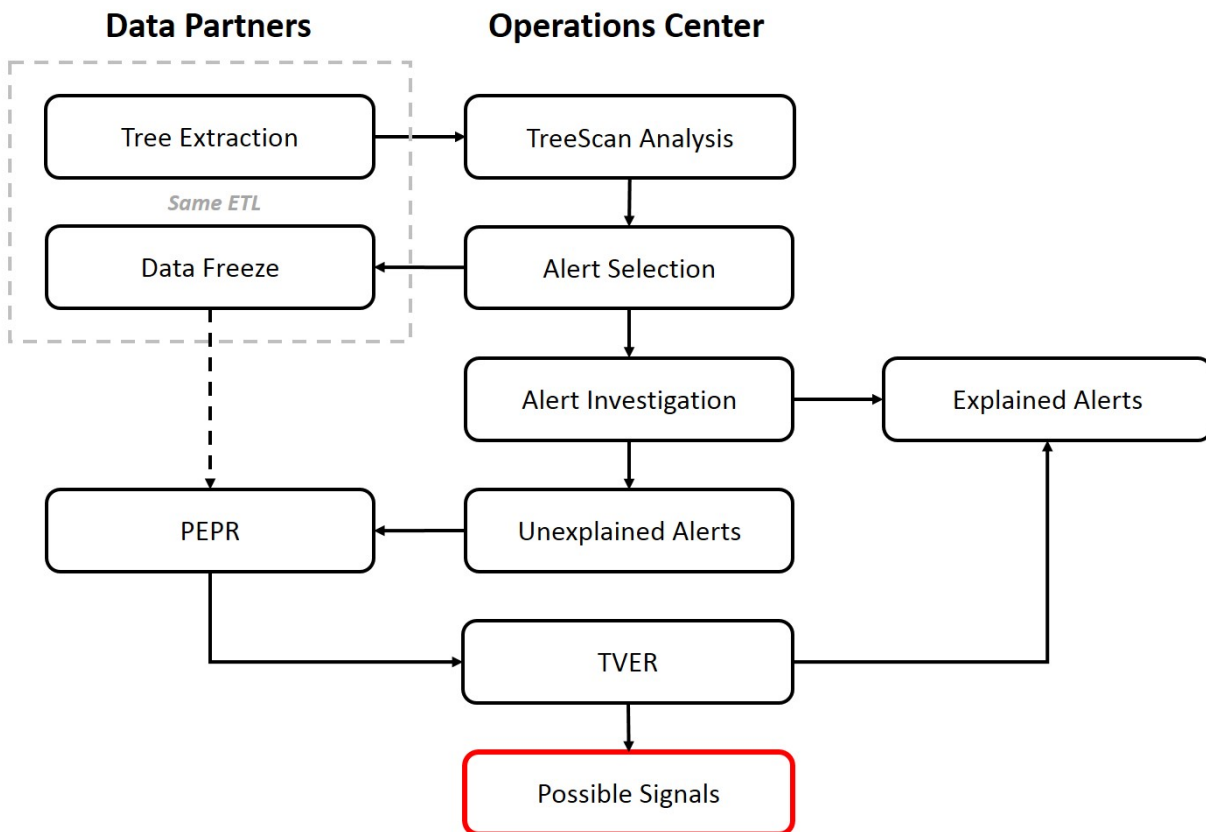
Data Partners   **Operations Center**

Figure 2.

As mentioned in the Introduction (Section I), statistical alerts do not necessarily imply a causal relationship between the exposure and outcome. Rather, a statistical alert could be the result of confounding, other sources of bias, or unexpected interactions. Potential sources of confounding and bias are presented in the Limitations (Section XII) below.

XII.   LIMITATIONS AND POTENTIAL SOURCES OF BIAS

There are some limitations of the proposed methods, which are either inherent to the tree-based scan statistic or related to the way in which it is to be implemented.

We are only considering risk windows that begin between 1 and 28 days post vaccination and end between 2 and 42 days post vaccination. This means that we are purposely not evaluating the risk of adverse reactions (e.g., anaphylaxis) occurring on the actual day of vaccination, Day 0, the reason being that diagnosis codes for conditions that preceded the vaccination visit are often entered on the day of the visit and can produce spurious alerts. (Presumably, adverse events occurring on Day 0 could be detected through the Vaccine Adverse Event Reporting System (VAERS), especially as same-day events are particularly likely to be identified as possibly vaccine-associated. Alternatively, existing Sentinel tools can be employed to examine specific HOIs of concern after HPV9.) It also means that we can only detect adverse reactions that manifest themselves fairly soon after vaccination, i.e., relatively acute outcomes. In order to evaluate adverse reactions that occur several months or years after

vaccination, it would be necessary to use a longer follow-up period. While the tree-temporal scan statistic can be used for longer follow-up periods, it is as yet untested for such applications.

To capture only the most serious types of outcomes, we are using outcomes recorded at emergency department visits and inpatient hospital stays. Outcomes in outpatient settings will be excluded. This could be either an advantage (e.g., in avoiding a potentially large number of statistical alerts for minor conditions) or a disadvantage. A potential disadvantage is reduced power to detect adverse reactions that are treated primarily (or seen initially) in an outpatient setting. It is possible, for example, that patients with conditions such as new-onset autoimmune diseases, complex regional pain syndrome (CRPS), or postural orthostatic tachycardia syndrome (POTS) would be seen first in an outpatient setting.

The analysis will be done using a single tree, which includes over 6,000 ICD-9 codes. The method can be used with other tree definitions as well as with multiple trees used simultaneously. While we expect that most trees developed with clinical expertise will generate similar results, some trees could potentially miss alerts generated by other trees.

The self-control tree-based scan statistic automatically adjusts for all non-time varying confounders, but it does not adjust for time-varying confounders. Examples of potential time-varying confounders are:

- *Seasonality:* HPV vaccine uptake has a demonstrated pattern of seasonality, with the greatest uptake occurring in August prior to the start of the school year. Thus, seasonality is a potential source of confounding in the case of outcomes that also are seasonal in nature. The resulting bias could go in either direction.
- *Concomitant exposures:* HPV vaccine is recommended to be administered as part of the "adolescent" set of vaccines at 11-12 years old. This set also includes meningococcal and tetanus-diphtheria-acellular pertussis vaccines. Given the frequent co-administration of these vaccines, a statistical alert that appears in HPV safety monitoring may in fact be associated with one or more concomitant vaccines instead.
- *Concomitant routine screening:* HPV vaccine administration is likely to occur with routine screenings for participation in school sports and recommended screenings for sexually transmitted diseases and mental health conditions among portions of this population. The medical evaluation (follow-up) that results from these routine screenings may occur during the surveillance period and appear to be attributable to the vaccine. For example, some girls may start oral contraceptive use at the same time as they get the HPV vaccine, and an adverse reaction to the oral contraceptive could then show up as a statistical alert after HPV vaccination.
- *Concomitant condition evaluated at the medical visit:* We will exclude Day 0 in the analysis to prevent capturing antecedent conditions that were present at the time of vaccination. However, it is still possible that an individual may be administered the vaccination coincident with evaluation of unrelated symptoms or a medical condition. Follow-up evaluation or treatment may then appear to be associated with the vaccine.
- *Latent or unobservable illness at the time of exposure:* The vaccination might be associated with Outcome A, which may or may not produce an alert, but which leads to a follow-up test or other medical workup procedure. That workup identifies Outcome B, which produces an alert. That is, in the absence of the vaccine-associated Outcome A, Outcome B would not likely have been detected, and Outcome B is not a true vaccine-associated outcome.

- *Day-of-the-week effects:* Since the HPV vaccine is mostly given on weekdays, between Monday and Friday, there may be confounding if and only if the outcome also has a weekly pattern. The *overall* day-of-week effect, as generated by all outcomes combined, is adjusted for by the conditional tree-temporal scan statistic. However, there can be residual day-of-week patterns that differ for different outcomes. The TreeScan method can adjust for such confounding as well, but in the TreeScan HPV4 pilot study, adjustment for this day-of-week by outcome interaction effect made very little difference, indicating that any potential bias from this effect regarding HPV vaccines and outcomes in emergency department or inpatient settings is small.

When simultaneously evaluating thousands of outcomes as potential adverse reactions, it is impossible to carefully adjust for all possible confounders and data dependencies. That is, what we gain in ability to simultaneously evaluate thousands of potential adverse reactions, we lose in ability to carefully consider clinical and epidemiological knowledge about all those outcomes. It must be kept in mind that the purpose of TreeScan analysis is to serve as a screening tool that is used to identify *potential* concerns that require further attention. Once an alert is generated, the attention needed could be anything from a quick recognition of an obvious source of confounding to the launching of a careful and detailed pharmacoepidemiological investigation. We reiterate that TreeScan results should not by themselves be viewed as evidence of a causal relationship between a vaccine and an outcome.

Relying on electronic healthcare databases has key advantages, including representativeness of routine clinical practice and efficient capture of the healthcare experiences of a large patient population. However, there are fundamental limitations to using administrative claims data for vaccine safety surveillance that will not be automatically overcome with TreeScan (no matter what its inherent strengths), such as variability in coding practices across the full range of outcomes evaluated and the fact that a diagnosis date is not necessarily the same as the date of onset of symptoms.

XIII. ACKNOWLEDGMENTS

We would like to acknowledge the participation of our fellow workgroup members: Carolyn Balsbaugh, Jeffrey Brown, David Cole, Rositsa Dimova, Lisa Ortendahl, Megan Reidy, John Scott, and Jawahar Tiwari.

XIV. REFERENCES

1.	Petrosky E, Bocchini JA, Jr., Hariri S, et al. Use of 9-valent human papillomavirus (HPV) vaccine: updated HPV vaccination recommendations of the advisory committee on immunization practices. MMWR Morbidity and mortality weekly report 2015;64:300-4.

2.	Merck & CO. Inc. Gardasil 9 [package insert]. Whitehouse Station, NJ. 2015. Available at http://www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/ucm426445.htm

3.	Brown JS, Petronis KR, Bate A, et al. Drug Adverse Event Detection in Health Plan Data Using the Gamma Poisson Shrinker and Comparison to the Tree-based Scan Statistic. Pharmaceutics 2013;5:179-200.

4.	Kulldorff M, Dashevsky I, Avery TR, et al. Drug safety data mining with a tree-based scan statistic. Pharmacoepidemiology and drug safety 2013;22:517-23.

5.	Kulldorff M, Fang Z, Walsh SJ. A tree-based scan statistic for database disease surveillance. Biometrics 2003;59:323-31.

6.	Kulldorff M. TreeScan User Guide [software manual]. Version 1.1. 2015. Available at http://www.treescan.org

7.	Yih WK, Nguyen M, Maro JC, et al. Pilot of self-controlled tree-temporal scan analysis for Gardasil vaccine. A Mini-Sentinel CBER/PRISM methods protocol. Posted March 30, 2015. Available at http://mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Pilot-Self-Controlled-Tree-Temporal-Scan-Analysis-Gardasil-Vaccine-Protocol.pdf

8.	Yih WK, Maro JC, Nguyen M, et al. Pilot of Self-Controlled Tree-Temporal Scan Analysis for Gardasil Vaccine. Draft of final report to FDA. Proposed publication 2016.

9.	Yih WK, Greene SK, Zichittella L, et al. Evaluation of the risk of venous thromboembolism after quadrivalent human papillomavirus vaccination among US females. Vaccine 2016;34:172-8.

10.	Cole DV, Kulldorff M, Baker MA, et al. Infrastructure for evaluation of statistical alerts arising from vaccine safety data mining activities in Mini-Sentinel. Final Report to FDA. Pending publication 2016.

Fictional Patient A
18 year old with depression and pelvic inflammatory disease

Look back period to define "new onset condition"

Scanning Risk Window

Observation Window

Vaccination Date

ED Visit #1
599.0 / 10.01.04.03
Urinary tract infection

276.51 / 03.08.02
Dehydration

789.00 / 17.01.07
Abdominal pain NOS

784.0/06.05.02
Headache

Outpatient Visit #3
311 / 05.08.02
Depressive disorder

614.9/ 10.03.02.03
Female Pelvic
Inflam Disorder NOS

789.09/ 17.01.07
Abdominal pain,
site NEC

ED Visit #2
V71.09 / 05.14.01
Observation for mental
condition

296.2 / 05.08.02
Major depressive disorder

614.9/ 10.03.02.03
Female Pelvic Inflam
Disorder NOS

616.0 / 10.03.02.02
Cervicitis

V62.84 / 05.13
Suicidal Ideation

789.09/ 17.01.07
Abdominal pain, site NEC

Outpatient Visit #1
490 / 08.02.04
Bronchitis

786.50 / 07.02.05
Chest pain

786.2 / 08.08.03
Cough

Outpatient Visit #2
V20.2 / 17.02.02
Routine child health check

311 / 05.08.02
Depressive disorder

V70.2 / 05.14.01
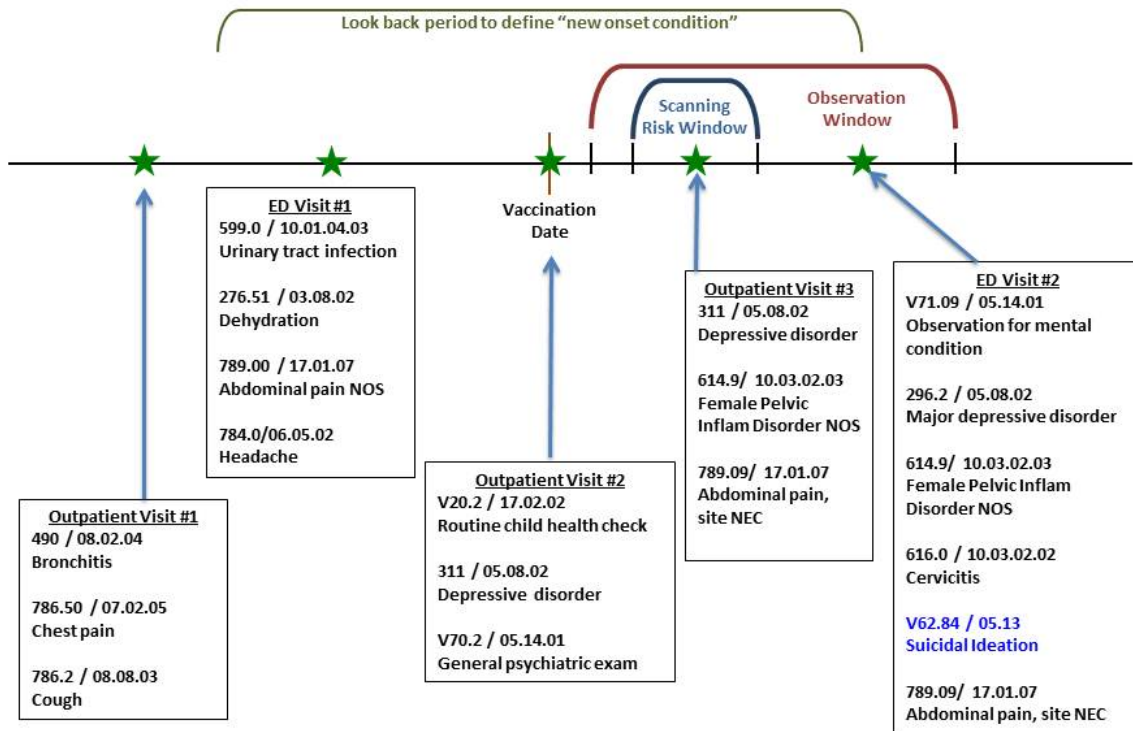General psychiatric exam

Figure 3.

This patient has 5 total healthcare visits in the period of analysis, 2 of which occur after vaccination (Outpatient visit #3 and ED visit #2). Of the 9 potential diagnosis codes post-vaccination, **only 1 will enter the Tree-Scan analysis (suicidal ideation from ED2).** All 3 diagnoses in OV3 are excluded because the analysis is restricted to the inpatient or ED settings. The 5 diagnoses from ED2 were excluded based upon several criteria designed to distinguish new onset conditions from pre-existing conditions, or acute exacerbations of pre-existing conditions.

Table 3.

Excluded Condition	Shares 3rd Level MLCCS with prior diagnosis in	Prior Diagnosis in Look back Period in
Observation for mental condition	OV2	
Major depressive disorder	OV3, OV2	
Female pelvic inflammatory disorder		OV3
Cervicitis	OV3	
Abdominal pain	ED1	OV3

Fictional Patient B
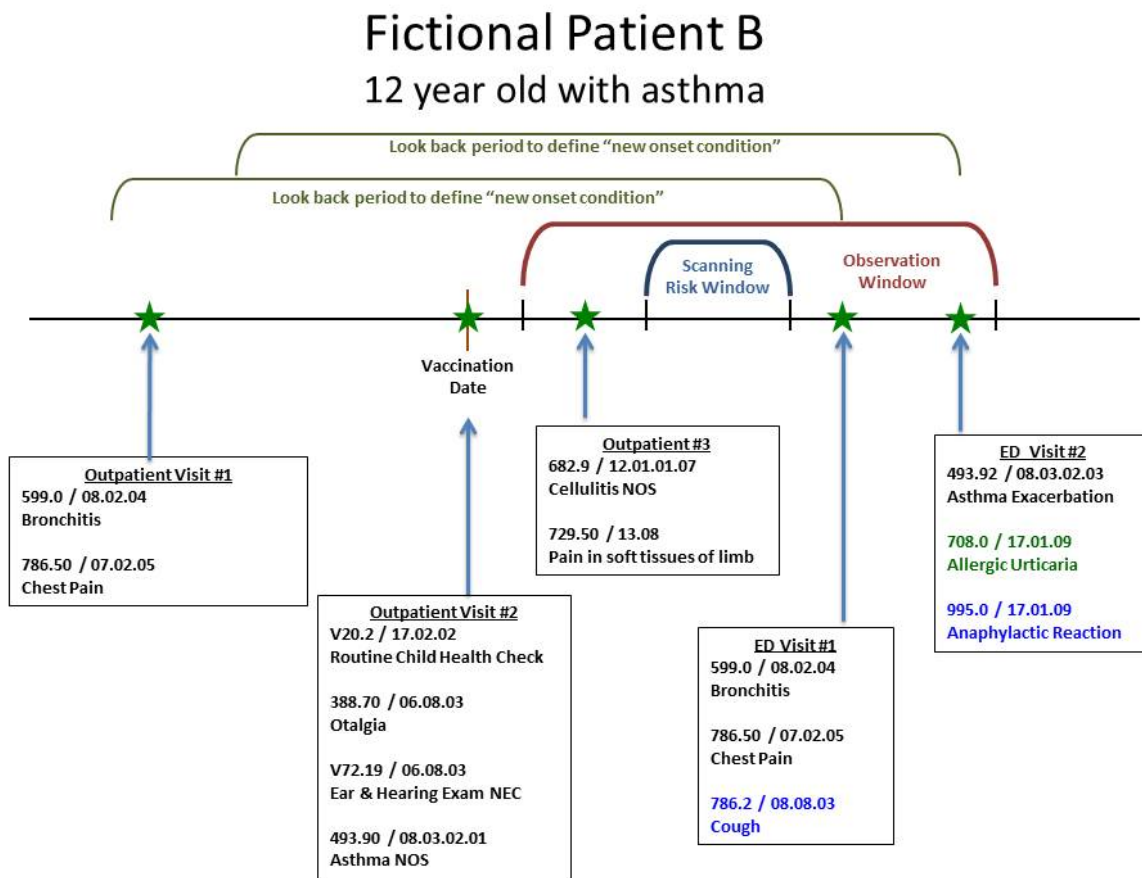**12 year old with asthma**

Figure 4.

This patient has 5 total healthcare visits in the period of analysis, 3 of which occur post-vaccination (OV3, ED1, ED2). Of the 8 potential diagnosis codes post-vaccination, **only 2 will enter the TreeScan analysis (cough from ED1 and anaphylactic reaction from ED2).** Both diagnoses in OV3 are excluded because the analysis is restricted to the inpatient or ED settings, even though "cellulitis" and "pain in soft tissues of limb" may represent actual vaccine-related events. The other 4 diagnoses were excluded as detailed below. Recall that each patient can contribute multiple diagnosis codes as long as they are on different 3rd level MLCCS branches. If two or more codes share the same 3rd level branch on the same visit, the program will select the rarest incident outcome. On ED2, "allergic urticaria" and

"anaphylactic reaction" occupy the same 3rd level. Only anaphylactic reaction is selected because it is the rarer of these two diagnoses.

Table 4.

Excluded Condition	Shares 3rd Level MLCCS with prior diagnosis in	Prior Diagnosis in Look back Period in
Observation for mental condition	OV2	
Major depressive disorder	OV3, OV2	
Female pelvic inflammatory disorder		OV3
Cervicitis	OV3	
Abdominal pain	ED1	OV3