MINI-SENTINEL CBER/PRISM METHODS PROTOCOL

PILOT OF SELF-CONTROLLED TREE-TEMPORAL SCAN ANALYSIS FOR GARDASIL VACCINE

Version 2.0

Prepared by: W. Katherine Yih, PhD, MPH,[1] Michael Nguyen, MD,[2] Judith C. Maro, PhD,[1] Meghan Baker, MD, PhD,[1] Carolyn Balsbaugh, MPH,[1] Jeffrey Brown, PhD,[1] David Cole, BM,[1] Inna Dashevsky, MS,[1] Martin Kulldorff, PhD[1]

Author Affiliations: 1. Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA; 2. Center for Biologics Evaluation and Research, Food and Drug Administration, Rockville, MD

March 30, 2015

Mini-Sentinel is a pilot project sponsored by the U.S. Food and Drug Administration (FDA) to inform and facilitate development of a fully operational active surveillance system, the Sentinel System, for monitoring the safety of FDA-regulated medical products. Mini-Sentinel is one piece of the Sentinel Initiative, a multi-faceted effort by the FDA to develop a national electronic system that will complement existing methods of safety surveillance. Mini-Sentinel Collaborators include Data and Academic Partners that provide access to health care data and ongoing scientific, technical, methodological, and organizational expertise. The Mini-Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223200910006I.

History of Modifications

Version	Date	Modification	By
V2	3/27/2015	• Made small clarifying edits throughout, including to make terminology (e.g., "outcome" vs. "adverse event") more consistent • Added a paragraph about multiple testing to Section IV • Added a paragraph about the hierarchical tree being filled to the fourth level for all outcomes to Section VIII • Added a sentence about exploring the effect of using different levels for defining incident diagnoses to Section IX • Added an appendix of Frequently Asked Questions and answers about TreeScan	Mini-Sentinel TreeScan Workgroup

MINI-SENTINEL CBER/PRISM METHODS PROTOCOL

Pilot of Self-Controlled Tree-Temporal Scan Analysis for Gardasil Vaccine

Table of Contents

I. STUDY PURPOSE

The TreeScan methodology is a statistical approach for vaccine and drug safety surveillance when looking for any of a wide range of unsuspected but potential adverse reactions. It is typically used to simultaneously evaluate thousands of different outcomes, casting a very wide net. 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 may 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 adverse events that merit a careful pharmacoepidemiological investigation.

The current project is a methods project. Its purpose is to develop and test unconditional and conditional variants of the self-controlled tree-temporal scan statistic, both with and without day-of-week adjustment, for vaccine safety surveillance. Gardasil (human papillomavirus quadrivalent vaccine (types 6, 11, 16, 18), Merck & Co.) will be the test vaccine. We will apply the TreeScan methodology, specifically the self-controlled tree-temporal scan statistic, to automated electronic health care data in order to evaluate whether there are any short term vaccine adverse events after Gardasil vaccination, without pre-specifying the type of adverse events or risk window a priori. The method simultaneously evaluates several thousand potential adverse events and groups of related adverse events, while simultaneously evaluating a large number of potential risk windows, adjusting for the multiple testing inherent in the many adverse events and risk windows evaluated.

This is the third Post-licensure Rapid Immunization Safety Monitoring (PRISM) TreeScan project and a continuation of prior methodological work to develop tree-based scan statistics for post-market vaccine safety surveillance. In two prior PRISM projects, we evaluated conditional and unconditional versions of the Poisson based tree scan statistic for vaccine cohort data, conditional and unconditional versions of the Bernoulli based tree scan statistic for self-control data, and a self-controlled unconditional tree-temporal scan statistic. These methods were evaluated and compared using four different vaccines and a total of 11 different vaccine-age group combinations. Some of these results were presented at the 2013 International Conference on Pharmacovigilance. The four vaccines were Measles, Mumps, and Rubella Virus Vaccine Live (M-M-R II, Merck & Co.) (ages 1-2 and 4-6 years), Measles, Mumps, Rubella and Varicella Virus Vaccine Live (ProQuad, Merck & Co.) (ages 1-2 and 4-6 years), Pneumococcal 7-valent Conjugate Vaccine (Prevnar, Wyeth Pharmaceuticals) (ages around 2, 4, 6, and 12 months) and Pneumococcal Vaccine, Polyvalent (Pneumovax 23, Merck & Co.) (ages 0-18, 18-65 and 65+ years).

For adults, there was an overall excess count of events of around 10% when comparing the 1-28 versus 29-56 day post vaccination windows, across a whole range of different diagnoses. The conditional self-control version adjusts for this, and worked very well, with few statistical signals. The unconditional self-control version, on the other hand, generated several signals for high level branches of the tree with very modest relative risks of around 1.1 or 1.2. We think these statistical alerts were due to temporal confounding, and hence, we do not recommend the unconditional version for adult vaccines. The same phenomenon was not seen for children, so both the conditional and unconditional versions are potential options for analysis of child vaccines. We do not know if there is a big difference in performance for adolescents, and we hope that this methodological project will shed light on this question.

With the Bernoulli model, it is necessary to pre-specify the risk window. The tree-temporal self-control version is attractive in that it does not require the prior specification of the risk window; rather it evaluates many different risk windows while adjusting for the multiple testing. This can help identify signals and periods of increased risk. For example, the Bernoulli self-control version with a 1-28 day risk window generated only a weak signal for MMRV and febrile convulsions (p=0.04), since the risk window was mis-specified, while the tree-temporal scan statistic generated a strong signal (p=0.00001), even though it had to adjust for additional multiple testing due to the many different potential risk windows considered.

The self-control version did not work for 2-month-old Prevnar vaccinees. The reason is that the infant period is especially subject to time-varying confounding. During the first months of life, the risk for many illnesses and conditions changes rapidly over time. A focused pharmacoepidemiology study can use self-controlled methods and adjust for this time-varying confounding when evaluating a single outcome. However, the current version of the self-control TreeScan cannot systematically account for this age-related confounding across the range of all potential outcomes evaluated.

Based on these three observations, we would expect that the hitherto undeveloped conditional tree-temporal scan statistic would perform quite well. In this project, we will develop that variant and compare various versions of the tree-based scan statistic for use in an adolescent population.

The self-control versions automatically adjust for all non-time varying confounders. One potential time-varying confounder is day-of-week effects due to the fact that all vaccines and some outcomes have an uneven weekly pattern, with more observations on weekdays than on weekends. In a Bernoulli based self-control analysis, this is easily overcome by specifying the fixed risk and control intervals with the same ratio of days in each modulus 7, such as for example 1-7 versus 8-21, with a 1-to-2 ratio. That trick does not work for the tree-temporal scan statistic, so the day-of-week effect would have to be accounted for as part of the statistical method.

The various TreeScan methods are compared in summary form in Appendix 1.

This project will serve as a pilot to prepare for the use of the TreeScan methodology for the regulatory review of future vaccines that will occur as part of FDA's 18 month post-licensure safety review. Based on the results of this pilot project, we will know whether the TreeScan methodology is likely to work for human papillomavirus (HPV) vaccines. It will also be invaluable in helping us choose the appropriate version of the TreeScan method, such as the conditional or unconditional tree-temporal scan statistic, as well as the various parameter settings.

II. SPECIFIC AIMS

1) Develop and evaluate a conditional version of the tree-temporal scan statistic

2) Develop and evaluate a day-of-week adjusted version of the tree-temporal scan statistic (for both unconditional and conditional variants)

3) Evaluate the tree-temporal scan statistic for use with adolescent vaccines

4) Provide comparative results in order to inform the choice of what type of tree-temporal scan statistic to use for adolescent vaccines (e.g. conditional vs. unconditional, with vs. without day-of-week adjustment) and what parameter settings to use

5) Enhance the TreeScan software to perform power evaluation for the tree-temporal scan statistic

6) Evaluate power for the tree-temporal scan statistic when used for adolescent vaccines, considering different sample sizes, outcomes, and relative risks

7) Explore and document practices for first-line follow-up of TreeScan-generated statistical alerts

8) Explore and document approaches to assess bias and time varying confounding in TreeScan statistical alerts

III. OVERVIEW OF ANALYSES

The primary analysis will be the conditional self-controlled tree-temporal scan statistic, without day-of-week adjustment. The unconditional version of this method is described in the next section. As secondary analyses, we will include day-of-week adjustment and also conduct unconditional self-controlled tree-temporal analyses, with and without day-of-week adjustment:

Planned tree-temporal analyses. (Post-vaccination observation period = Days 1–56.)

#	1°/ 2°	Conditional/ unconditional	Adjusted for day of week	Alpha level
1	**1°**	Conditional	No	0.05
2	2°	Conditional	Yes	0.01
3	2°	Unconditional	No	0.01
4	2°	Unconditional	Yes	0.01

The method formally adjusts for the multiple testing generated by the many groupings of outcomes and many potential risk windows that are evaluated in a single analysis for a specific exposure. The secondary analyses represent additional multiple testing, which is not formally adjusted for. In the secondary analyses, we will adjust for this additional multiple testing informally by using an alpha level of *0.01* to reject the null hypothesis.

IV. UNCONDITIONAL 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 if there are, for example, 1000 nodes on the tree and 2002 potential time intervals, there would be 2,002,000 potential clusters to evaluate and for which we would need to adjust for multiple testing. If these were 2 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 2 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 outcome definitions 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, which under the null hypothesis is assumed to be uniform across the follow-up period. Under the alternative hypothesis, there is at least one branch for which there is a temporal cluster of cases during some time interval.

With the tree-temporal scan statistic and under the null hypothesis, any outcome is equally likely to occur on any of the days following the initial drug/vaccine exposure. For each tree node and time interval, we calculate the log likelihood ratio (LLR) test statistic:

$$LLR = \ln \frac{(\frac{c}{n})^c (\frac{n-c}{n})^{n-c}}{(\frac{w}{T})^c (\frac{T-w}{T})^{n-c}} \, I(c/n > w/T)$$

where n is the number of cases in the node, c is the number of those node cases that are also in the time interval, w is the length of the time interval, and T is the total length of the follow-up period. I() is the indication function, which is 1 when there are more cases in the time interval than expected under the null, and it is included to ensure that we are looking for an excess risk of having the outcome rather than a protective decreased risk. Note that T is a constant that is the same for every node and every time interval.

For each node on the tree, the LLR is calculated for each time interval under consideration. The node-interval combination with the maximum LLR 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.

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. This is done by generating, say, 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, but the post-exposure timing of those cases varies, with each one generated from a uniform distribution independent of each other case. 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 LLR 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 LLR from the real data set has a 5% chance of being among the 50 highest maximum LLRs from the real and 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 LLR 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. We will only evaluate outcomes occurring 1 to 56 days after vaccination. The day of vaccination is not included since (i) a well care visit at which vaccines are given could generate diagnosis codes (outcomes) unrelated to vaccination, such as problems found during an eye examination, or (ii) vaccines may be given during a health care visit that happened due to a different health concern. In other words, the null hypothesis should be that all diagnosis codes (outcomes) occur with equal probability on each day, and that is clearly not the case for the day of vaccination.

V. POWER EVALUATION

As a component of this study, we propose to empirically assess and compare the statistical power of temporal and non-temporal TreeScan analyses. In allowing TreeScan to detect the appropriate risk window, the tree-temporal analysis involves many more instances of multiple hypothesis-testing than the non-temporal analysis. The tree-temporal analysis might increase power compared to non-temporal analysis if the fixed risk window in the non-temporal analysis is chosen poorly for the outcomes whose risk is elevated, but it may decrease power if the fixed risk window in the non-temporal analysis was well-chosen. We will quantify this trade-off.

We will consider the performance differences in statistical power for the following situations: 1) fixed risk windows when one anticipates the actual risk window perfectly; 2) fixed risk windows when one anticipates the actual risk window imperfectly or not at all, specifying a risk window that is either longer or shorter than the true risk window, and/or in the wrong place; and 3) varying risk windows considered by the tree-temporal scan statistic.

This evaluation will extend existing studies of sample size calculations that were limited to non-temporal tree scan designs. As in those cases, this evaluation will entail intentionally injecting risk into the dataset to assess the detection capabilities of the TreeScan analyses. The key questions we hope to provide evidence for are: “at what point after licensure should TreeScan analyses be conducted?” and “what is the ability to detect a true excess risk for different outcomes and magnitudes of excess risk?”

In order to perform preparatory-to-surveillance sample size calculations, we must specify the background rates for the 6000+ outcomes that make up the Multi-level Clinical Classifications Software (MLCCS) diagnosis tree (see Section VIII for further details). These background rates will allow us to calculate expected outcomes that will occur during the surveillance period. This evaluation will extract age-specific background rates for adolescents from one or more of the Mini-Sentinel databases.

For every scenario that is run, we assign the following characteristics:

1) *Total sample size*, given in number of vaccinees expected at the time of TreeScan deployment, ranging from 100,000 to 5,000,000 vaccinees.

2) *Incidence rate difference of interest*, given in number of excess cases of a specified outcome per *x* vaccine doses, ranging from 1 excess case/1,000,000 vaccine doses up to 5 excess cases/1,000 vaccine doses.

3) *ICD-9 diagnosis*, i.e., node on the diagnostic tree, assigned to a simulated elevated risk. This evaluation will choose diagnosis codes that occur at various frequencies based on the background rates extracted. It will include outcomes that occur as frequently as 1 event/1,000 person-years (i.e., "infrequent" per the Council for International Organizations of Medical Sciences (CIOMS) designation system) up to 1 event/1,000,000 person-years (i.e., "very rare" per CIOMS).

4) *Type of TreeScan analyses,* given as one of these four:

   a. unconditional non-temporal self-controlled

   b. conditional non-temporal self-controlled

   c. unconditional temporal self-controlled

   d. conditional temporal self-controlled

5) *Temporal scanning window parameters,* which are given by a series of parameters that control the number of risk windows considered during any given analysis. These variables will only be relevant in the tree-temporal designs.

6) *Length of the risk and control windows used by TreeScan,* given in days. These variables will only be relevant in non-temporal designs.

7) Length of the true risk window, given in days.

All scenarios will be analyzed with the power evaluation feature of the TreeScan software, which simulates both the null hypothesis and known alternative hypothesis by scenario. This feature is currently available for the unconditional non-temporal design, and as part of this project, we will enhance the software by implementing this power evaluation feature for the other three designs (conditional non-temporal and unconditional and conditional tree-temporal) as well.

We will present the statistical power associated with each scenario, to inform the optimization of parameter settings and/or the timing of the TreeScan analyses relative to sample size accrued. The results will be presented as a set of tables.

VI. STUDY POPULATION AND ENROLLMENT CRITERIA

Data in Mini-Sentinel Common Data Model-format will be obtained from five Sentinel sites: HealthCore, Humana, Aetna, Optum, and Harvard Pilgrim Health Care. Each site will contribute data from 1/1/2006 (or earliest available date thereafter) through the latest date of complete data available. For some Data Partners, there are restrictions on the use of patient-level data for certain members, even if the data are de-identified. These members will be excluded.

As this is a self-control study, only vaccinated individuals will be included. We will include those that received the vaccine on or after their 9th birthday and before their 27th birthday during the available date range. The study will include both female and male vaccinees. In order to be able to define incident diagnoses, only members enrolled in the participating health plans for at least 183 consecutive days prior to vaccination will be included in the base study population. Members must also have been enrolled for at least the full 56-day follow-up period after vaccination in order to be included. Enrollment gaps of 45 days or less will be bridged and treated as continuously enrolled time. Descriptive statistics obtained in September 2014 indicate that there were close to 1.4 million first apparent doses of HPV administered between January 1, 2006 and June 30, 2013, among members satisfying the above criteria.

VII. STUDY VACCINE AND "CONTRIBUTED TIME"

Gardasil vaccination will be identified using CPT code 90649.

"Contributed time" is defined as Days 1-56 after vaccination. "Contributed time," "follow-up period," and "observation window" are used interchangeably in this protocol.

Only the first apparent dose will be included in analysis. We will consider an HPV dose to be a first apparent dose if there is no prior record of an HPV dose for that patient, going back the maximum amount of available time for the patient but no earlier than his/her 9th birthday. All subsequent doses will be ignored, regardless of the timing of their occurrence. Descriptive statistics from the PRISM Gardasil-Venous thromboembolism study indicate that only 5.7% of Dose 2s are given within 56 days of Dose 1, hence we do not expect the 56 days of contributed time to be impacted by a subsequent dose in any major way.

The ability to evaluate subsequent doses is a future planned enhancement.

VIII. HIERARCHICAL DIAGNOSIS TREE

Outcomes will be identified and defined using ICD-9 codes and a hierarchical classification of these codes. 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:

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 fourth level for all outcomes. For example, ICD-9 code 729.5, "pain in soft tissues of the limb," corresponds to a second level outcome on the tree, 13.08, in which the first level 13 is "diseases of the musculoskeletal system and connective tissue" and the second level 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.

Some ICD-9 codes will be excluded from the tree and therefore from the analysis, for example, those representing:

- Outcomes that are common and of an unspecific or less serious nature, such as fever, croup, and acute pharyngitis
- 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)
- Other outcomes very unlikely to be caused by vaccination, such as well-care visits, delivery of a baby, vitamin deficiencies, or fracture of a lower limb

IX. INCIDENT DIAGNOSES OF INTEREST

The study will focus on incident diagnoses observed during contributed time. A diagnosis is an incident diagnosis if it was observed in the inpatient or emergency department setting during the contributed time and if there was no other diagnosis in the same third-level branch of the MLCCS diagnosis tree in any setting during the prior 183 days. This means that, even if it 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.

The third-level branch was chosen based on results of testing and seems to avoid the identification of alerts for outcomes that are either too broad or general to be meaningful (second-level) or so specific

that multiple alerts for essentially the same category of outcome, produced by the same mechanism, are generated (fourth-level). Nonetheless, as a simple exploration of the effect of using different levels, we will compare the numbers of ascertained outcomes in the tree structure when using second, third, and fourth levels to define incident diagnoses.

Each member can contribute multiple incident diagnoses during his/her contributed days, as long as they are not part of the same third level branch of the MLCCS tree. If a member has multiple incident diagnoses on the same 3rd level branch, the program will select the rarest incident outcome. The rarest incident outcome is defined as the one for which the MLCCS level 4 code is the least common based on an outcome frequency list. While very unlikely, in the event that there are two or more or codes that occur with equal frequency, the one with the lowest numerical MLCCS code will be selected. If they are also part of the same fourth level of the MLCCS tree, the rarest outcome is defined as the one for which the ICD9 code is the least common based on a frequency list. While very unlikely, in the event that there are two or more codes that occur with equal frequency, the order of priority will be the one with the lowest numerical ICD9 code, followed by V codes, and E codes. The outcome frequency lists will be created based on ED and inpatient data for 9-26.99 year olds.

Appendix 2 illustrates these rules by means of two fictional example patients.

X. RISK AND COMPARISON WINDOWS

We will examine a set of possible risk windows, namely those intervals that start 1-28 days after vaccination and that end 2-42 days after vaccination. We will only consider risk windows that are at least 2 days long and at most 28 days long. That is, we will ignore all one-day long risk windows, since it seems unlikely that the vaccine will increase the risk only on a single specific day after vaccination for all patients, such as Day 6, but not on any other days for any patients, such as Days 5 or 7.

The comparison period consists of those days within the Days 1-56 follow-up period that are not in the risk window being evaluated.

XI. DATA AND RESULTS FORMATS

A. DATA FROM DATA PARTNER SITES

The data for analysis will be provided by the sites and will include:

- Patient_index –artificial patient ID that looks like 1,2,3, etc., to provide the ability to list all outcomes for the same person
- DX – ICD9
- Distance – time to event in terms of number of days after vaccination

The Distance variable is the number of days after vaccination that the incident diagnosis occurred. For example, it is 12 if the diagnosis occurred 12 days after vaccination. All these numbers will be in the range [1,56].

Note that a patient can have more than one row, if they had more than one incident diagnosis during the 56 days after vaccination.

B. DATA FORMAT FOR TREESCAN ANALYSIS

For each tree-temporal scan analysis there will be a separate data set, with columns for the following three variables:

- ICD9Code
- #Patients
- #DaysAfterVaccination

For example, if there were three patients who had a febrile convulsion on the seventh day after vaccination, the data set will have the following row for those three patients:

780.31 3 7

C. FORMATTING FOR REPORTING RESULTS

For each tree-temporal scan analysis, results will be reported in a table format. We will report all statistically significant branches (cuts) on the tree, together with selected other branches that are either above or below a statistically significant cut. For the primary analysis (conditional temporal-tree without day-of-week adjustment), statistical significance is defined as $p \leq 0.05$, where the p-values are adjusted for the multiple testing inherent in the many branches and risk intervals evaluated. The columns in the table will be:

- MLCCS code or ICD-9 code for the cut/branch
- Total number of events on the branch, summed over the 1-56 days follow-up period
- The detected risk interval
- Number of events on the branch that are in the detected risk interval
- Expected number of events on the branch and in the risk interval
- Relative risk
- Attributable risk
- Log likelihood ratio
- P-value

XII. STATISTICAL ALERT FOLLOW-UP

The TreeScan method evaluates thousands of unspecified clinical outcomes to identify statistical relationships ("alerts") between exposures and outcomes. Data related to alerts will be frozen for possible later follow-up. These alerts do not necessarily imply a causal relationship between the exposure and outcome. 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 Limitations (Section XIII) below. The following are unlikely sources of confounding:

- *Age:* The self-controlled tree-temporal scan statistic automatically adjusts for all non-time varying confounders. Since all vaccinees are at least 9 years old and the follow-up period is only 8 weeks, we do not expect to see confounding by age for any of the outcomes.
- *Indication or contraindication:* Adolescents are unlikely to be temporarily indicated or contraindicated for vaccination. However, to minimize the risk for confounding by indication or contraindication, we use a post-vaccination rather than a pre-vaccination comparison window.
- *Differential loss to follow-up:* By design, a minimum enrollment period is required to prevent differential loss to follow-up.

Every statistical alert must be evaluated to assess whether the alert represents a known relationship, and if it does not, the likelihood that the alert is the result of confounding or bias. All statistical alerts 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 working group, involving other expertise as needed.

The initial assessment will review publicly available information (e.g., literature), data available to FDA (e.g., clinical trial reports), and data available "in-house" (i.e., at the Operations Center and available to the TreeScan working group). All statistical alerts will be categorized (as above) for review to determine next steps. If it is determined that additional analyses are appropriate, FDA will work with the TreeScan working group, other experts, and the Mini-Sentinel Operations Center to plan and carry out the additional analyses. These additional analyses could include examining the data that were frozen or conducting new analyses using the entire Mini-Sentinel Distributed Database.

XIII. 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.

First of all, 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 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.

Also, in this pilot project, we will be able to evaluate only the first dose of the HPV vaccine series. This is a consequence of the way the current data extraction program was constructed. Building the ability to evaluate subsequent doses is a future planned enhancement.

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, although we have tried only one other tree. While we expect that most trees developed with clinical expertise will generate similar results, some trees could potentially miss alerts generated by other trees.

Multiple outcomes from the same patient will be used, but only if they are not on the same third level branch of the MLCCS tree. This could potentially create problems when evaluating higher levels of a

tree. For example, since myocardial infarction (07.02.03) is often preceded by chest pain (07.02.05) a few days before, a patient may be counted twice at the 07 and 07.02 level analyses, and the tree based scan statistic will erroneously ignore the dependence of these events. While we have not seen any evidence of problems caused by this dependence in previous work, the possibility of such a problem should be kept in mind when evaluating statistical alerts at the two highest levels of the tree.

To capture 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 or a disadvantage. A potential disadvantage is reduced power to detect adverse reactions that are primarily treated in an outpatient setting.

While the self-control tree-based scan statistic automatically adjusts for all non-time varying confounders, 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. This bias could go in either direction.
- *Concomitant exposures:* By design, 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 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 potential bias resulting from this can be adjusted for, and such adjustment is planned in secondary analyses.

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 determine *potential* problems 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. 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.

Beyond the limitations inherent to our data source, there are also constraints to the overall objective to develop and evaluate a new statistical method to detect unknown and unsuspected adverse events following vaccination: the lack of a gold-standard reference database. To overcome this constraint, prior projects have used real healthcare data to verify that TreeScan can identify known associations (febrile seizures after MMRV, ITP after MMR) and not produce numerous alerts that are the result of confounding. In this project, we also plan to inject a range of simulated risks into background rate data to test whether the method can detect outcomes that occur at various frequencies, outcomes that occur with varying risk window onset and duration, and different magnitudes of risk. We also plan to conduct follow-up investigations of any identified alerts using electronic data, as described in Section XII, to categorize the alerts. Only through repeated use and follow-up of statistical alerts can we fully characterize the performance of TreeScan and determine whether it can successfully identify serious, unknown, and unsuspected adverse reactions. We believe our careful, iterative, and staged approach is the most suitable and feasible way to evaluate the appropriate use and implementation of this method.

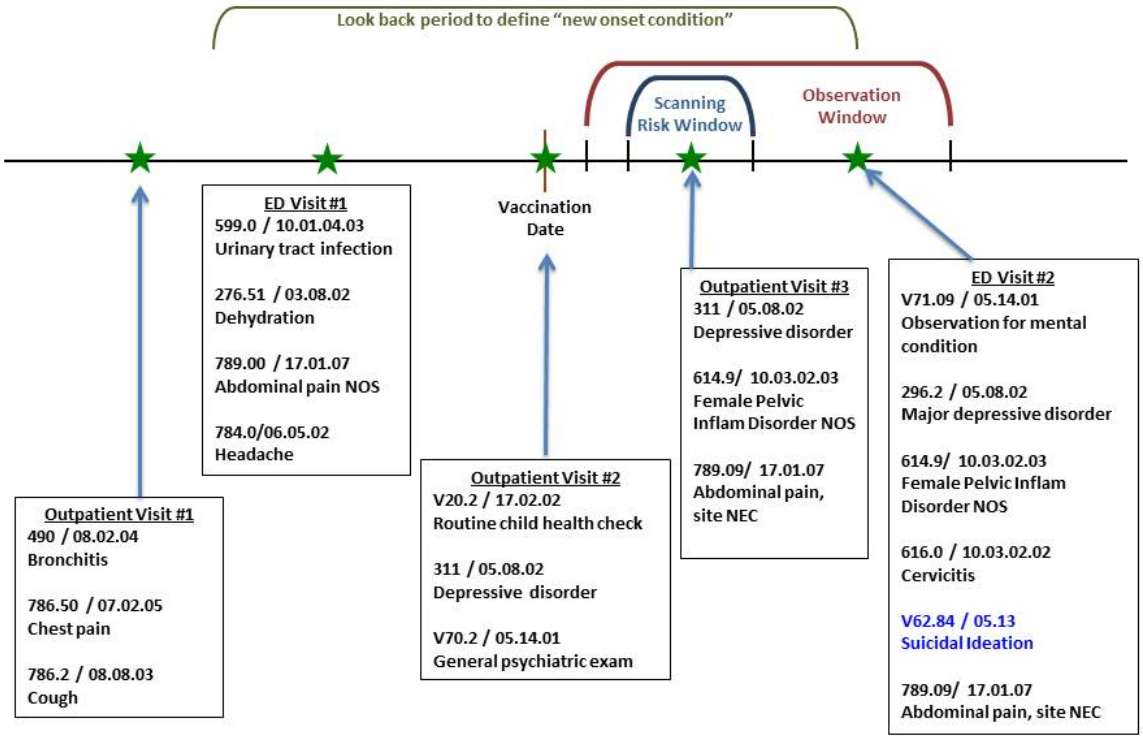
XIV. APPENDIX 1: COMPARISON OF TREESCAN METHODS

	Tree Scan Statistic				Tree-Temporal Scan	
	Cohort / Poisson Model		Self-Control / Bernoulli Model		Self-Control	
	Unconditional	Conditional	Unconditional	Conditional	Unconditional	Conditional
DATA NEEDS						
Exposure definition	One vaccine or multiple vaccines with any AND, OR, NOT logical operators. May differentiate between doses of the same vaccine.					
Outcome definition	Incident diagnosis, i.e., a diagnosis for which there was not the same or similar diagnosis in prior X days. Similar is defined as not being on the same 2nd, 3rd or 4th level of the tree.					
Adverse events (AEs) in risk interval	AE count in risk interval of, e.g., 1-2 or 1-28 days post vaccination				AE count in follow-up period of, e.g., 1-56 days post-vaccination, with information about the exact number of days post-vaccination	
AEs in comparison group	Many options, e.g., AEs in unexposed pre-vaccination time, after risk window, and among non-vaccinated. Used to generate age-adjusted expected counts.		AE count in control interval of, e.g., 29-56 days post vaccination			
TREESCAN INPUT						
Tree structure	A set of ICD9 codes and MLCCS codes, with information about the parent of each one					
AEs in risk interval	For each ICD9 code, number of AEs in risk interval				For each ICD9 code, number of AEs by days post-vaccination	
AEs in control interval	N/a		For each ICD9 code, number of AEs in control interval			
Expected count	For each ICD9 code, expected AEs under the null		N/a		N/a	
NULL HYPOTHESIS	AEs are generated from the expected counts.	The relative risk of different AEs is determined by the relative magnitude of the expected counts, but the total number of AEs is fixed and non-random.	The probability of an event occuring in the risk versus control interval is proportional to the lengths of those intervals. For each ICD9 code, total number of AEs is fixed and non-random.	The probability of an event occuring in the risk versus control interval is proportional to the total number of AEs in those intervals summed over all ICD9 codes. For each ICD9 code, total number of AEs is fixed and non-random.	An AE occurs uniformly over the follow-up period, with equal probability on each day.	Irrespective of the ICD9 code, all AEs have the same probability of occurring on a specific day. The probability is equal to the total number of AEs on that day divided by the total number of AEs in the follow-up period.
ALTERNATIVE HYPOTHESIS	There is at least one leaf or one branch on the tree where there are more expected AEs than what is defined under the null hypothesis.					
TREESCAN ANALYSIS OPTIONS						
Common parameters	Input file names; Type of Scan (tree only or tree-time); Probability Model; Conditional Analysis; Number of Monte Carlo replications; Output options					
Model-specific parameters	None		Risk Interval Probability	None	Temporal window start and end times; Maximum and minimum temporal window length; Data time range (length of follow-up)	
Power evaluation	Yes, available			No, not yet available		
TREESCAN EXECUTION						
Scan the tree	Scan the tree to find most likely and secondary cuts/clusters.				Perform the temporal scan on each potential cut on the tree, to find most likely and secondary clusters.	
Random Monte Carlo data sets	Generated from expected counts, using Poisson distribution. Random data sets may have more or less total AEs than the real data.	Conditioned on total number of AEs , so that each random data set has exactly the same number of total AEs as the real data set. The random number of AEs for a particular ICD9 code is binomially distributed as Bin(n,p), where n is the total number of AEs and p is the expected count in the leaf divided by the total expected count summed over the whole tree.	Generated using the specified risk interval probability. For each ICD9 code, the sum of the AEs in the risk and control intervals will be the same in each random data set and the real data set. The total number of AEs in the risk interval, summed over all ICD9 codes, may be different in the random and real data sets.	For each ICD9 code, the sum of the AEs in the risk and control intervals will be the same in each random data set and the real data set. The total number of AEs in the risk interval summed over all ICD9 codes, C, is the same in each random data set and the real data. Randomization is conducted by randomly picking C of the AEs to be assigned to the risk interval.	For each ICD9 code, the sum of the AEs taken over all days will be the same in each random data set and the real data set. Independently of every other AE, an AE is assigned to a day using a uniform distribution where each day in the follow-up period is equally likely to be chosen.	For each ICD9 code, the sum of the AEs taken over all days will be the same in each random data set and the real data set. For each day, the total number of AEs on day X, summed over all ICD9 codes, is the same in each random data set and the real data. Randomization is conducted by randomly permuting the days and the ICD9 code pairings, keeping the marginals fixed.
INTERPRETATION						
Self-controlled	No		Yes			
Pre-defined risk interval	Yes				No	
Adjusts for multiple testing	Yes					
Adjusts for temporal variation common to all ICD9 codes	No	Yes	No	Yes	No	Yes

XV. APPENDIX 2: EXAMPLES OF FICTIONAL PATIENTS

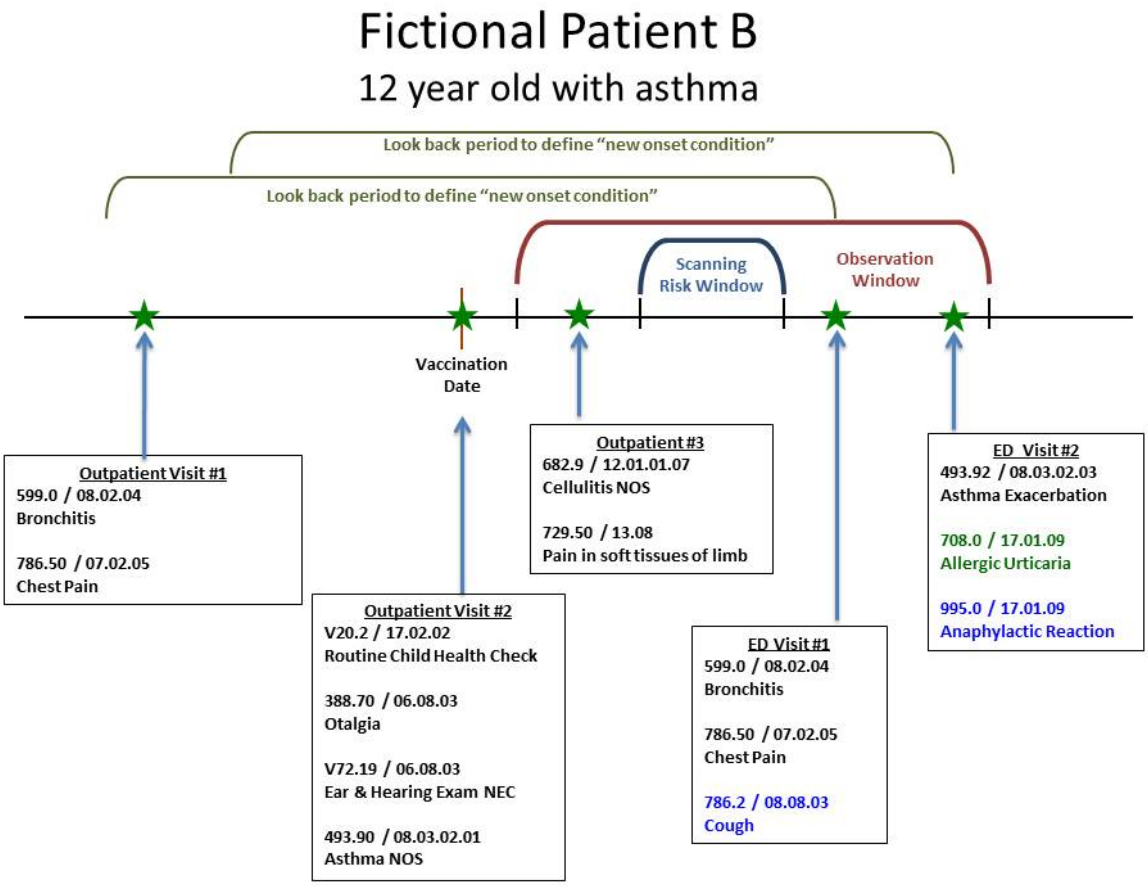
Fictional Patient A

18 year old with depression and pelvic inflammatory disease



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.

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



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 Tree-Scan 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.

Excluded Condition	Shares 3rd Level MLCCS with prior diagnosis in	Shares 3rd Level MLCCS with diagnosis on same day	Prior Diagnosis in Look back Period in
Bronchitis			OV1
Chest pain			OV1
Asthma exacerbation	OV2		
Allergic urticaria		ED2	

XVI. APPENDIX 3: FREQUENTLY ASKED QUESTIONS

TREESCAN DOCUMENTATION

1. How can I access and use the TreeScan software?

   Answer: TreeScan is free software, available for download at www.treescan.org. The TreeScan User Guide, also available at www.treescan.org, provides detailed instructions on how to use the software.

2. Where can I find more information about the TreeScan statistical method?

   Answer: The basic TreeScan methodology, using a Poisson probability model, is described by Kulldorff, Fang, and Walsh (2003).[6] For its use for pharmacovigilance, see the paper by Kulldorff et al (2013).[5] For information about the various probability models, please consult the TreeScan User Guide. A version of this FAQ document will exist as a living document on the TreeScan website, www.treescan.org, to be updated and added to over time.

STATISTICAL METHOD VALIDATION

3. What is meant by "validation" in the context of TreeScan?

   Answer: There are two principal ways of thinking about validation of this method:

   a. Validating the TreeScan statistical method itself. Here, we are asking, "Given a true adverse event, what approaches can be used to help confirm that TreeScan is capable of detecting a true risk?"
   b. Validating TreeScan findings. Here, the question is, "Given a TreeScan alert, what methods can be used to help confirm (or refute) that the alert represents a true vaccine-related risk?"

4. Since there is no gold standard against which to compare the TreeScan results, could some or all of the following kinds of analyses be done for validation purposes?

   a. Conduct a power evaluation by creating simulated data spiked with an excess frequency of a particular outcome of interest during the risk window using different relative risks, different sample sizes, and different baseline frequencies of the outcome of interest.

      Answer: Yes, this is one important way to evaluate the ability of the method to detect statistically significant differences in the frequency of the outcome of interest in the control vs. risk window. Such a power evaluation study is currently underway.

   b. Divide the dataset into two random subsets, perform a TreeScan analysis on each one, and compare the results.

      Answer: Not recommended for either validating the TreeScan method or validating TreeScan findings. Since the two datasets are randomly obtained from the same larger dataset, any

difference between the results of the two analyses will, by definition, be due to chance. Moreover, once we have shown that the results from the comparison are the same or very similar, we will still not have proven that they are both accurate results instead of both erroneous results. Hence, it is not a useful way to evaluate the method. Using only part of the available data would reduce statistical power, so it is also not a recommended approach to use only half the data for the actual analysis and save the other half for validating the findings.

c. Divide the dataset in two (e.g., 2006-2011 vs. 2012-2014, or Data Partners A and B vs. C and D), perform TreeScan analysis on each half, and compare the results.

   Answer: Not recommended for validating either TreeScan findings or the method. Using only part of the available data would reduce statistical power, so it is not a recommended approach for validating actual findings. However, the investigator might choose to subset the data by time period based on their awareness of secular trends (e.g., significant changes in coding practices over time) or of other forms of time-varying confounding that could introduce a bias. That being said, unlike Q4b above, subsets of data originating from different periods of time and/or different Data Partners could differ. Hence, any difference in TreeScan results could be due either to that reason or to random chance. Also, as in Q4b, even if we were to find that the results from the comparison were the same or very similar, we would not have proven that they were both accurate results instead of both erroneous results.

d. Divide the available data into two random subsets and perform a TreeScan analysis on one and a conventional regression analysis on the other.

   Answer: Not recommended for validating TreeScan findings, but could be used for methods validation. We could run TreeScan on one randomly selected half of the dataset, note the statistically significant alerts, and then use a non-data-mining method, such as standard logistic regression, on the other half of the dataset to demonstrate that the TreeScan alerts are statistically significant using standard methods. When using the standard method, we would not have to adjust for multiple testing, since we would be running the analysis on a completely independent, non-overlapping dataset, for a few previously detected outcomes. This exercise could serve an educational purpose in confirming the TreeScan alerts using a widely known and established method. It must be noted, however, that this approach should not be used for validating actual TreeScan findings, as power would be greatly reduced.

e. Use another established signal detection method on the same dataset as TreeScan and compare results.

   Answer: Recommended for methods validation, and acceptable, with a caveat, for data analysis and validation of actual findings. One such comparison has already been done, for evaluation purposes, in which TreeScan was compared with the Gamma Poisson Shrinker (GPS), a method that is commonly applied in spontaneous reporting systems.[1] Doing other similar comparisons would be very valuable, using other datasets and/or other signal generation methods. For actual analysis, it is sometimes good to use multiple methods, but one must also be careful not to use so many different methods that at least one of them will generate a statistically significant result just by chance.

5. How is the adjustment for multiple testing implemented?

   Answer: Considering the thousands of overlapping disease outcome definitions 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, and if the null hypothesis is true in the real dataset, the test statistics come from exactly the same probability distribution. This means that, if the null hypothesis is true, the rank of the 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.

6. Why isn't a Bonferroni type of adjustment used to adjust for the multiple testing?

   Answer: Because the overlapping data are obtained from different cuts on the tree, where the data for one cut are a subset of the data in another cut, there is a lot of dependence between the likelihoods calculated for closely related cuts. With such dependence, a Bonferroni type of adjustment becomes too conservative.

EPIDEMIOLOGIC ISSUES

7. What does a "TreeScan alert" mean?

   Answer: A TreeScan alert occurs when TreeScan detects a statistically significant difference in the frequency of a coded outcome when comparing counts of the coded outcome in the risk and control windows while making various adjustments, depending on the specific tree-based scan statistic being utilized. Adjustments are made for multiple testing, so the alerts are unlikely to be due to chance. An alert could be due to confounding, so a TreeScan alert should be investigated using traditional pharmacoepidemiologic methods.

8. Is it likely or unlikely that there will be statistical alerts just due to chance?

   Answer: It is unlikely, since the method adjusts for the multiple testing inherent in the thousands of outcomes and groups of related outcomes that are evaluated. Suppose you use an alpha level of 0.05, declaring an alert when $p \leq 0.05$. If you perform 100 TreeScan analyses, the expected number of TreeScan analyses without any alerts due to chance is 95. If the null hypothesis is true, the expected number of TreeScan analyses without any alerts at all is 95. In other words, in the long run, 5 percent of your TreeScan analyses will contain a false positive alert that is due to chance. If this is considered too high, you can instead use an alpha level of 0.01. Then, in the long run, only 1

percent of your TreeScan analyses will contain a false positive alert that is due to chance.

9. A higher frequency of healthcare visits and diagnosis codes has been observed in the first two weeks after vaccination than subsequently, usually due to workup of conditions that were present at the time of vaccination and diagnosed shortly thereafter. Wouldn't you expect false TreeScan alerts to arise due to this phenomenon?

   Answer: It depends on the type of TreeScan analysis. It is easy to test whether this phenomenon is present in a particular dataset by comparing the total outcomes in different periods after vaccination. Whether it is present depends on the age group and the vaccine exposure being studied. When present, it can bias the unconditional self-control and the unconditional tree-temporal scan statistics. On the other hand, the *conditional* versions are explicitly designed to adjust for this type of bias, although they will only adjust for the general phenomenon across all outcomes, not for an exceptionally strong time bias for a specific outcome. The latter must be dealt with either using an adjusted Poisson TreeScan analysis or in the alert investigation stage.

10. Could bias enter a TreeScan analysis as a result of differences in recommendations for use of the product (including the timing of rollout) among different demographic subgroups, such as males vs. females?

    Answer: In using a self-controlled version of TreeScan, there would be no bias introduced by differential use of the product by different subgroups, nor by differences in the timing of rollout to different subgroups.

11. What if only one subgroup of the population has an excess product-associated risk of an outcome, while there is no excess risk for the rest of the population? For example, maybe a vaccine increases the risk of seizures in just females or in just a particular age group. Will TreeScan be able to detect the risk and identify the subgroup?

    Answer: This would be a case of effect modification (interaction). If an excess risk were limited to a particular demographic subgroup of the population, the statistical power to see it would be reduced due to the random noise caused by the random timing of the outcome events in the unaffected subgroups, reducing the signal-to-noise ratio. If the sample size were big enough, TreeScan could detect it, but without identifying the subgroup. The group at risk could then be identified during an investigation of the TreeScan alert. That being said, if an investigator were concerned about subgroup-specific effects *a priori*, then the investigator could subset the population under investigation at the outset. For example, the population of interest could be defined as pregnant women as opposed to all persons receiving a vaccination, provided such a population could be identified in administrative data.

12. What if the outcome is such that it can occur only in one subgroup of the population, such as a sex-specific outcome? Can TreeScan detect such problems?

    Answer: Yes. Since there are no instances of the outcome for the other subgroup(s), there is no random noise added by that group, and the power will be approximately the same as for a subgroup-specific analysis.

13. What are likely sources of confounding in TreeScan analysis?

    Answer: The self-controlled versions of TreeScan automatically adjust for all time-invariant confounders, but they do not adjust for time-varying confounding (except for an overall greater number of visits and diagnoses in the first couple of weeks after vaccination, which the conditional version can adjust for). The potential source of confounding of primary concern is exposure to concomitant vaccines. Other examples of possible sources of time-varying confounding are seasonality, concomitant drug exposures, concomitant routine screenings, concomitant conditions evaluated at the exposure visit, and latent or unobservable illness at the time of exposure. Additionally, there is time-varying confounding by indication when a person is temporarily indicated or contraindicated to receive a particular vaccine, e.g., prophylactic pneumococcal polysaccharide vaccine administered preceding splenectomy.

TREE STRUCTURE

14. How is the tree generated from the data?

    Answer: It is not. The tree is pre-specified by the user, before collecting the data and doing the analysis.

15. Is it possible to use a different tree, rather than the MLCCS classification? Why is the MLCCS tree used, and what are its general characteristics?

    Answer: Any tree that aggregates data into logical, hierarchical groupings would be sufficient for our purpose. Conveniently, the Agency for Healthcare Research and Quality has created the Multi-Level Clinical Classification Software (MLCCS), which groups over 14,000 diagnosis codes into a smaller number of clinically meaningful categories that are useful for our purpose here. Please see http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp for more information.

    The MLCCS is a hierarchical system that is defined using both single-level CCS groupings and ICD-9-CM codes. Four levels exist in the multi-level *diagnosis CCS*. That is, there are four levels of groupings in addition to one level of the base codes. The first of the four levels in the multi-level diagnosis CCS is broken into 18 categories. These categories broadly group the single-level CCS and ICD-9-CM codes into 18 major groupings such as “Diseases of the Circulatory System,” “Diseases of the Digestive System,” “Injury and Poisoning,” and “Mental Illness.” The specificity of categories increases with the higher levels, such that second-level categories are more specific than first-level categories, third-level categories are more specific than second-level categories, and fourth-level categories are more specific than third-level categories. An example table is shown below.

Table 1: Example Outcome Hierarchical Tree Structure

Node	Level	ICD9	Description
06	1st		Diseases Of The Nervous System And Sense Organs
06.04	2nd		..Epilepsy; convulsions
06.04.02	3rd	780.3	……Convulsions
	3rd	780.31	……Febrile convulsions
	3rd	780.32	……Complex febrile convulsions
	3rd	780.33	……Post traumatic seizures
	3rd	780.39	……Other convulsions

16. How will the transition from ICD-9 to ICD-10 coding of diagnoses affect the results?

    Answer: The current pilot uses ICD-9 codes only, so the transition to ICD-10 codes will not affect this study. TreeScan can be used with ICD-10 codes (or any others) as long as an appropriate tree exists that utilizes those codes. The AHRQ has already developed an ICD-10 based MLCCS tree, and therefore, it will not be difficult to transition to an ICD-10 based coding system. ICD-10 codes are more detailed than ICD-9 codes, therefore it is possible that we would see alerts for more specific outcomes when using ICD-10 codes.

17. Is it possible to execute a TreeScan analysis with both ICD-9 and ICD-10 codes in the same study, i.e., perhaps because the study spans periods when both ICD-9 and ICD-10 codes were in use?

    Answer: Yes. One common tree must be used. This can be created by mapping ICD-9 codes onto ICD-10 codes or vice versa or by basing the common tree on a combination of ICD-9 and ICD-10 codes.

18. How does the TreeScan method account for potential differences in how outcomes are coded in administrative data?

    Answer: Relying on electronic healthcare databases has key advantages, including the efficient capture of the healthcare experiences of a large patient population for analysis. 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. For this reason, all TreeScan alerts must be evaluated with coding variability in mind.

    One of TreeScan's strengths is that an investigator is not required to pre-specify how a clinical concept will be coded in administrative data. For example, different physicians could code an outcome slightly differently, for example, coding a febrile seizure as

    - convulsions (780.3), or
    - febrile convulsions NOS (780.31), or
    - complex febrile convulsions (780.32).

    The tree structure can detect an alert by combining related diagnoses at more coarsely-aggregated levels of the tree even when a singular diagnosis does not have a large enough sample size to generate an alert independently.

19. What is the purpose of defining and using incident diagnoses, rather than using all recorded diagnoses?

    Answer: A diagnosis is an incident diagnosis if there is no prior diagnosis of the same or similar kind in the prior "X" days. The primary goal of the incident diagnosis definition is to identify only new-onset events and distinguish them from ongoing conditions that existed before vaccination. By removing such ongoing conditions, the signal-to-noise ratio increases, which in turn increases statistical power. A second reason for using only incident diagnoses is to ensure that a person can only contribute one event during the post-exposure follow-up period in the same user-defined level of the tree. This is further accomplished by ensuring that the number of days used in the incident diagnosis definition is longer than the post-exposure follow-up time. Together, these operational characteristics help to ensure the independence assumptions that underlie the TreeScan probability models.

20. Why is an incident diagnosis defined based on the second, third, or fourth level of the diagnosis tree rather than on ICD-9 codes?

    Answer: Many ICD-9 codes are similar, and the same disease condition may be recorded using slightly different ICD-9 codes during follow-up visits.

21. In this pilot study, a diagnosis is only counted as incident if there were no other diagnoses in the same third level of the tree during the preceding pre-specified X number of days. What are the pros and cons of using the second vs. third vs. fourth level of the hierarchical tree to identify incident diagnoses? How does the definition of incident outcome impact the TreeScan results and the ability to detect alerts?

    Answer: A trade-off exists in choosing any tier of the diagnosis tree (e.g., second, third, fourth levels) for defining incident diagnoses. Defining incident events at the second level results in fewer diagnostic events compared to defining them at the third or fourth level, leading to a smaller sample size. Using the fourth level leads to a larger sample size. The additional diagnostic events can either reduce statistical power if they are random noise due to pre-existing conditions recorded during follow-up visits, or they may increase statistical power if they are informative diagnostic events unrelated to pre-existing conditions. Increased power will increase the sensitivity of the method, that is, the ability to detect true adverse reactions. Using the third tier in this pilot was a reasonable approach to start with in view of the trade-offs.

22. Wouldn't the choice of the third level to define incident outcomes cause an increased frequency of second- or fourth-level outcomes in the risk window to be missed?

    Answer: The third level is only used to determine if an outcome is incident or not. However the incident diagnosis is defined, all five levels of the tree are evaluated, and the TreeScan method can generate an alert at the first, second, third, fourth, or fifth level of the tree.

23. What happens if there are outcomes that are not more finely differentiated than at the second level of the MLCCS tree? How is such an outcome taken into consideration if incident diagnoses are determined using a higher level?

    Answer: We refer to the ICD-9-CM codes that form the basis of the tree that we use as the fifth-level (i.e., in tree parlance, the leaf level), and hence, that level exists for all outcomes. Aggregation of multiple leaves (i.e., into a branch or trunk of the tree) occurs at the first through fourth levels. Sometimes, the MLCCS tree ends at the second or third levels. In such cases, we have artificially filled in the tree up to the finest level of aggregation, i.e. the fourth level. For example, ICD-9 code 729.5, "pain in soft tissues of the limb," corresponds to a second level outcome on the tree, 13.08, in which the first level 13 is "diseases of the musculoskeletal system and connective tissue" and the second level 08 is "other connective tissue disease." There is no finer MLCCS differentiation of this outcome, but additional levels were created for it by adding zeroes — the third level is 13.08.00 and the fourth level is 13.08.00.00.

IMPLEMENTATION AND TREESCAN PARAMETER SELECTION

24. Why was an observation period of Days 1-56 chosen, and why will only risk windows that are at most 28 days long (and that start 1-28 days after vaccination and end 2-42 days after vaccination) be considered?

    Answer: In theory, clusters of any duration can be evaluated. In practice, a self-control design is most suited to detect outcomes that happen during a well-defined risk period after vaccination, so our early efforts are focused on early-onset outcomes. In this pilot of the TreeScan method, we chose not to use an observation period of longer than Days 1-56 after vaccination, because Gardasil Dose 2 is commonly given approximately 2 months after Dose 1, and we wished to avoid having control time for Dose 1 overlap with the risk period for Dose 2. With an observation period of 56 days, if we were to look for and detect risk intervals that were longer than 28 days, it would more accurately reflect a *decreased* risk in the few days *outside* the interval rather than an *increased* risk *inside* the interval. Moreover, it is desirable, for reasons of statistical power, to use an observation period of twice the length of the maximum cluster length one is looking for. So clusters lasting more than 28 days could be assessed, but in that case we would want to use a longer follow-up period.

25. How does TreeScan handle multiple doses of a vaccine or other exposure?

    Answer: Currently, due to the nature of the data extraction program (as opposed to the TreeScan™ software), only first doses can be identified. The ability to identify and evaluate multiple doses will be developed in future enhancements to the data extraction program.

26. What would the consequences be, for the TreeScan analysis with a 56-day follow-up period, of Gardasil Dose 2 being administered within 56 days of Dose 1?

    Answer: The recommended spacing of Gardasil doses has been 0, 1-2, and 6 months. Dose 2 is sometimes given within 56 days of Dose 1. If this were not an unusual occurrence, then a true association between (the unascertained) Dose 2 and one or more outcomes (concentrated, for example, during the first 2 weeks after (the unascertained) Dose 2 and often falling during Days 43-

56 after Dose 1) would lead to an attenuation of the apparent risk of these outcomes from Dose 1. However, most Dose 2's are given more than 56 days after Dose 1, so we would not expect such attenuation to be a major problem.

27. Instead of using the same rules to define incident outcomes, set the observation period, etc., for all outcomes, wouldn't it make more sense to tailor the rules to the various outcomes, some of which would have more acute and/or immediate onsets than others?

    Answer: Customization of the rules for specific outcomes is not feasible with a method of this kind, where there are thousands of different outcomes being assessed. This is an early-warning, alert-detection system. Definitions of incident outcomes, risk intervals, and other parameters can and should be customized in subsequent alert investigations.

LOOKING AHEAD

28. Can TreeScan be used to identify statistically protective effects as well as excess risks? If so, how?

    Answer: Currently, this cannot be done when using the Poisson or tree-temporal versions of the TreeScan software, because a one-sided statistical test is implemented. A future version of TreeScan will be able to do two-sided tests. Theoretically, when using the TreeScan Bernoulli option for a self-control analysis, it is possible to do a one-sided test for a protective effect by simply redefining the risk interval as the control interval and the control interval as the "protective interval." However, the evaluation of effectiveness using TreeScan is more complex and is not a current goal for Sentinel.

29. What kind of investigations will be undertaken if and when unexpected safety alerts arise from TreeScan analysis, and how will decisions about those investigations be made?

    Answer: A manuscript is in development that will present the planned approach to investigating alerts from TreeScan. In general, the approach is as follows: We will first identify whether alerts correspond to a known or expected association, such as those noted in the product label. Then, of the remaining alerts, we will consider the potential biases that might have been operating and the strength of the bias needed to generate a false alert. We may perform additional alert follow-up under the guidance of the Food and Drug Administration.

30. If TreeScan alerts arise, how can we be sure that the alert is not related to one particular Data Partner?

    Answer: Any alert investigation will consider the possibility that a TreeScan alert is Data Partner-specific and will include stratification of data by Data Partner.

31. Can TreeScan be used to monitor the safety of drugs as well as vaccines?

    Answer: TreeScan has been applied to pharmaceutical drugs. When the method was applied to two antifungal drugs and two diabetes drugs, known adverse reactions were found.[1,5] However,

TreeScan is still being piloted and is not in widespread use for either drug or vaccine safety monitoring.

XVII. REFERENCES

1) 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.
2) Jacobsen SJ, Ackerson BK, Sy LS, et al. Observational safety study of febrile convulsion following first dose MMRV vaccination in a managed care setting. Vaccine, 2009;27:4656-4661.
3) Klein NP, Fireman B, Yih WK, et al. Vaccine Safety Datalink. Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. Pediatrics, 2010;126:e1-8. doi: 10.1542/peds.2010-0665.
4) Kulldorff M, et al. A Tree-Temporal Scan Statistic for Drug and Vaccine Safety Data Mining. Manuscript, 2014.
5) Kulldorff M, Dashevsky I, Avery TR, et al. Drug safety data mining with a tree-based scan statistic. Pharmacoepidemiology and Drug Safety, 2013;22:517-523.
6) Kulldorff M, Fang Z, Walsh SJ. A tree-based scan statistic for database disease surveillance. Biometrics, 2003;59:323-331.
7) McClure DL, Raebel MA, Yih WK, et al. Mini-Sentinel methods: Framework for assessment of positive results from signal refinement. Pharmacoepidemiology and Drug Safety, 2014;23:3–8. doi: 10.1002/pds.3547.
8) Yih WK, Kulldorff M, Fireman BH, et al. Active surveillance for adverse events: The experience of the Vaccine Safety Datalink project. Pediatrics, 2011;127:S54–S64.