MINI-SENTINEL CBER/PRISM SURVEILLANCE

INFRASTRUCTURE FOR EVALUATION OF STATISTICAL ALERTS ARISING FROM VACCINE SAFETY DATA MINING ACTIVITIES IN MINI-SENTINEL

Prepared by: David V. Cole, BM,[1] Martin Kulldorff, PhD,[2] Meghan Baker, MD, ScD,[1] Grace Lee, MD, MPH,[1] Judith C. Maro, PhD, MS,[1] Inna Dashevsky, MS,[1] W. Katherine Yih, PhD, MPH,[1] Carolyn Balsbaugh, MPH,[1] Estelle Russek-Cohen, PhD,[3] David Martin, MD, MPH,[3] Michael Nguyen, MD[3]

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

July 22, 2016

v0.6

sMini-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.

Mini-Sentinel CBER/PRISM Surveillance

Infrastructure for Evaluation of Statistical Alerts Arising from Vaccine Safety Data Mining Activities in Mini-Sentinel

Table of Contents

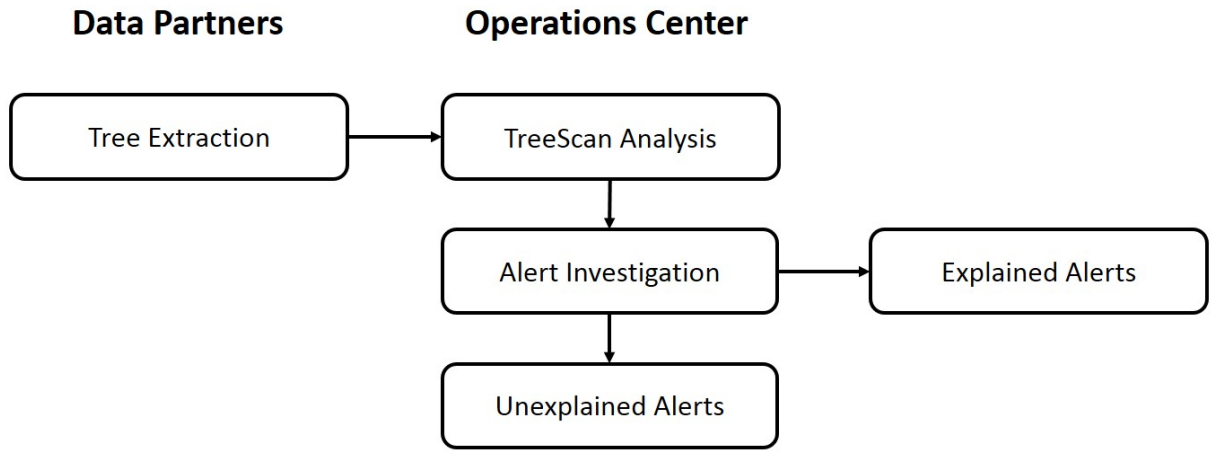
I. INTRODUCTION

Vaccine safety work in Mini-Sentinel to date has been conducted through protocol-based assessments (PBAs) with chart review[1-5] and sequential surveillance in near-real time,[6] all of which look for an association between an exposure and one or more pre-specified outcomes. At the same time, development work has been conducted on data mining methods using a tree-based scan statistic to look for unanticipated adverse events following a specified vaccine exposure.[7-9] Since the purpose of TreeScan is to generate hypotheses that might represent potential vaccine safety concerns, the implementation of TreeScan requires approaches to efficiently investigate alerts. Since chart review is resource- and time-intensive, a more efficient, scalable, and timely approach was needed to obtain some of the important clinical context to help evaluate TreeScan alerts to differentiate between those which are expected or known and alerts that require more study. This project fulfills those goals by creating the infrastructure to freeze, evaluate, and visualize the procedure and diagnostic codes associated with the TreeScan alerts.

Description of TreeScan Operations at Project Start

During the developmental phases of TreeScan, the operational process started with the Tree Extraction workplan, prepared at the Operations Center and run at the participating Data Partner sites. After all sites uploaded their results, a programmer/analyst at the Operations Center combined and summarized the data into the input format required by the TreeScan software. The analysis was run and the results reviewed by the TreeScan workgroup. However, the original data pull was large and not frozen nor formatted specifically to enable follow-up investigations. There were no automated ways for the larger dataset to be reduced in size to focus on alerts, nor were the data captured in a way that enabled further evaluation either in aggregate or at an individual patient level in de-identified manner.

Figure 1: TreeScan Process Flow During Methods Development Stage



This report illustrates how the original TreeScan operational process was modified in 3 key ways to achieve the goal of creating a new alert follow-up approach. It also details how we tested the software enhancements using vaccine safety data from Sentinel.

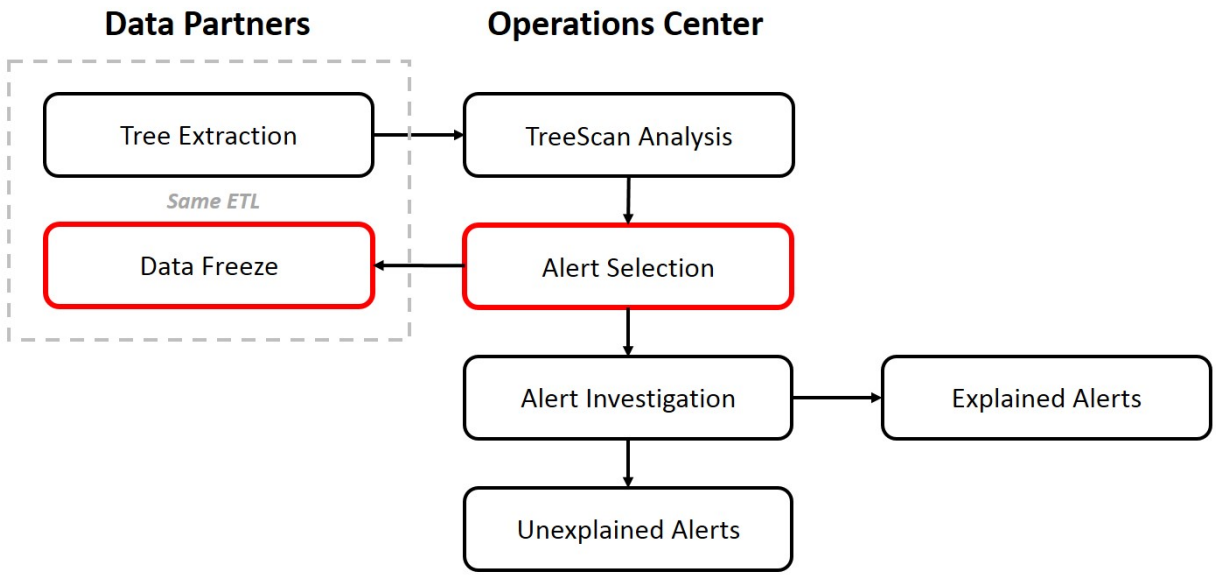
The three aims of this project were:

1. To integrate a data freeze into the Tree Extraction and Analysis process for one TreeScan method
2. To develop a means by which limited, de-identified, patient-level case data associated with selected alerts could be retrieved from the Data Partner sites
3. To create a reporting tool for FDA and Mini-Sentinel clinicians and investigators to use in reviewing the case data

II. DEVELOPMENT

A. AIM 1: DATA FREEZE

Figure 2: TreeScan Process Flow Adding Data Freeze



The first aim addresses the dynamic nature of the Mini-Sentinel Distributed Database (MSDD) (http://mini-sentinel.org/data_activities/distributed_db_and_data/default.aspx), where Data Partners periodically extract, transform, and load (ETL) their data into the Mini-Sentinel Common Data Model (CDM) format on a staggered schedule to add data for the newly-available time period as well as to add, delete, and update data for the previously reported period. Since there is no existing requirement that the patient identifier (PatID) value be consistent from one ETL to the next, a PatID value in one ETL may or may not be associated with the same patient in subsequent ETLs. Even if we kept a list of PatIDs which contributed to an alert, we would have no guarantee these were the exact same *patients* in a more recent ETL. By freezing case data from the same ETL used during Tree Extraction at each site, we can avoid such issues.

Time pressure is also somewhat mitigated, since frozen data may be retrieved and reviewed at any time without regard to refresh schedules. The main time pressure is thus limited to the requirement to complete Tree Extraction, TreeScan analysis, and data freeze on the same ETL for all Data Partners. Further, by freezing data for patients associated only with selected alerts rather than for the entire

cohort, we minimize the amount of data that must be frozen and stored at the Data Partner sites, and we also gain efficiency in subsequent program runs on these much smaller datasets.

1. Tree Extraction Crosswalk

The first step in the Data Freeze process is to identify patients who contributed exposure-event episodes to each alert. Since the Tree Extraction program already identifies those episodes, we have added code to the Tree Extraction program to create and save at the local Data Partner sites a crosswalk dataset containing one record per incident event. For self-controlled analyses, the incident event is recorded as a combination of vaccine exposure and a subsequent event occurring within a pre-specified time window following exposure.9

At minimum, the following variables are required in the crosswalk dataset: patient ID, admission date of exposure, admission date of event, diagnosis code, diagnosis code type, and the analysis group to which the patient episode applies. (See Appendix A for the data dictionary.) The analysis group is used to distinguish between analysis looks at more than one vaccine (for example, MMR and MMRV) and/or at more than one age group (for example, 1 to 2 years and 3 to 5 years). Using the preceding examples, we would have four analysis groups, one for each unique combination of vaccine and age group.

2. Alert Selection

Once Tree Extraction has been completed by all Data Partners, the output data are aggregated at the Operations Center and run through the TreeScan analysis software (http://www.treescan.org), which in turn outputs a results file. To automate the steps between analysis and running the data freeze, five SAS macro programs were developed:

a. Convert the Tree Temporal Scan Analysis Results Flat File Into a SAS Dataset

The first macro converts the Tree Temporal Scan analysis results from a delimited flat file into a SAS dataset. Two versions of the macro were developed: one to be used if the relative risk and excess case count are already included in the file, and another to be used if those variables are not included. This alternative macro calculates the relative risk and excess case count using the other variables in the file, using the following formulas:[12]

$$\text{relative risk} = \frac{\frac{\text{observed}}{\text{expected}}}{\frac{\text{node cases} - \text{observed}}{\text{node cases} - \text{expected}}}$$

$$\text{excess cases} = \text{observed} - \left(\text{expected} \times \frac{\text{node cases} - \text{observed}}{\text{node cases} - \text{expected}}\right)$$

b. Convert the Horizontal Diagnosis Tree Dataset Into a Vertical Child-Parent[1] Structure

The second macro converts the diagnosis tree used by Tree Extraction in horizontal form, i.e. one record per diagnosis code and additional variables for each higher-level node of the tree, to a vertical form with one record per child-parent relationship which will be useful when identifying the next-higher node for any node which produces an alert. To illustrate, take a simple example from the current form of the tree based on the Multi-level Clinical Classifications Software (MLCCS), a product of the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project (http://www.hcupus.ahrq.gov/toolssoftware/ccs/ccs.jsp). The ICD-9 diagnosis code for febrile seizures is 780.31, and the record in the diagnosis tree lookup dataset looks like this:

dx_codetype	Dx	mlccs5	mlccs4	mlccs3	mlccs2	mlccs1
09	780.31	780.31	06.04.02.00	06.04.02	06.04	06

When this horizontal record is converted into the vertical child-parent form, we have four records:

child	parent
780.31	06.04.02.00
06.04.02.00	06.04.02
06.04.02	06.04
06.04	06

Since future versions of the diagnosis tree could conceivably have a different number of levels than the current five, the conversion macro includes code to automatically determine the number of node variables in the horizontal tree and extract the nodes one by one into the vertical child-parent tree.

c. Convert the Child-Parent Tree Dataset Into dx-node Structure

In order to construct the alerts SAS lookup file to be used with the Data Freeze program, the third macro converts the child-parent tree into a different vertical form with one record per unique combination of diagnosis code and node. Taking the above example, we have the following five records:

dx	node
780.31	780.31
780.31	06.04.02.00
780.31	06.04.02
780.31	06.04
780.31	06

d. Automatically Identify Statistical Alerts Using Criteria Agreed Upon by the TreeScan Workgroup Prior to Analysis

The fourth macro selects the alerts for data freeze based on criteria that should be established by the FDA and TreeScan workgroup prior to analysis. First, primary alerts are selected from nodes meeting both of the following criteria: 1) p-value less than or equal to a maximum value; and 2) relative risk greater than or equal to a minimum value. These primary alerts are then compared to the child-parent

[1] The term "child-parent" here refers to tree structure terminology and not to a familial relationship between people.

tree, and if the parent for any primary alert has a p-value less than or equal to a maximum value, the node is added to the set of alerts to be frozen. All of the p-value and relative risk inputs are represented by macro parameters to allow investigators the flexibility to adjust the criteria.

Note that this data freeze is meant as a defensive measure to preserve data related to alerts that *may* require further investigation, and to do so in a timely fashion, we select a broader set of alerts than will actually require review. In reality, most if not all of the alerts will cover outcomes that already have an established association with vaccines in general or with the specific vaccine of interest. Further investigation will be limited to those in the "unexplained" category, as illustrated in **Figure 2**, which represents a small subset of the alerts that are selected for data freeze.

e. Create a Lookup Dataset for Use with the Data Freeze Program

The fifth and final macro compares all selected nodes to the dx-node tree to create a lookup table with the diagnosis codes associated with each node along with an arbitrarily assigned alert ID to be used to distinguish between frozen datasets if further investigation is required. For this step in particular, it is essential that the exact tree from Tree Extraction has been used. The tree may be pruned differently depending on the vaccine and age group under evaluation, and new versions of the tree will be developed to account for the transition to ICD-10, meaning the nodes of different versions of the tree may have different sets of underlying diagnosis codes. Finally, an additional table is created to display the selected alerts for the FDA and PRISM investigators to review and approve as the final step before distributing the Data Freeze program. At this point, additional nodes that did not meet the primary or secondary alert selection criteria may be added manually to the set of alerts for data freeze.

The Alert Selection program package was developed to automate processes, minimize opportunities for human error, and, most importantly, to shorten the time needed between analysis and data freeze. The timing of the data freeze is critical, as it is most desirable to freeze from the same ETL used to generate the alerts. Data Partners refresh as often as every three months on differing schedules, so the data freeze must be done as quickly as possible to avoid issues with CDM data at any site being overwritten with a new ETL. Before running Tree Extraction, care should also be taken to choose the optimal time window to complete all steps from Tree Extraction through Data Freeze on the same ETL at all Data Partner sites. To maintain efficiency in the Alert Selection phase, any alterations to the TreeScan analysis output file structure should be communicated in advance so the programmer can ensure compatibility with the programs.

3. Data Freeze Program

The Data Freeze workplan package includes the PRISM TreeScan Data Freeze macro, required utility macros (including the standard ms_freezedata macro), the alert lookup dataset with AlertID and associated Dx values, and a master SAS program.

The Data Freeze macro uses the standard ms_freezedate macro to create and save a snapshot of each available CDM table, populated with data for only the patients associated with one or more alerts. A preliminary step in the macro obtains a list of the PatIDs associated with each alert, and then creates a list of unique values across all alerts. Then a single set of CDM tables is saved to avoid duplication of data that must be stored at the Data Partner sites. The macro also creates and saves two subsets of the Crosswalk dataset for each alert. The first contains the exact exposure-event records which contributed to the alert. The second contains all other records in the Crosswalk dataset for the patients associated with the alert; that is, it contains records for incident events that are unrelated to the alert in question.

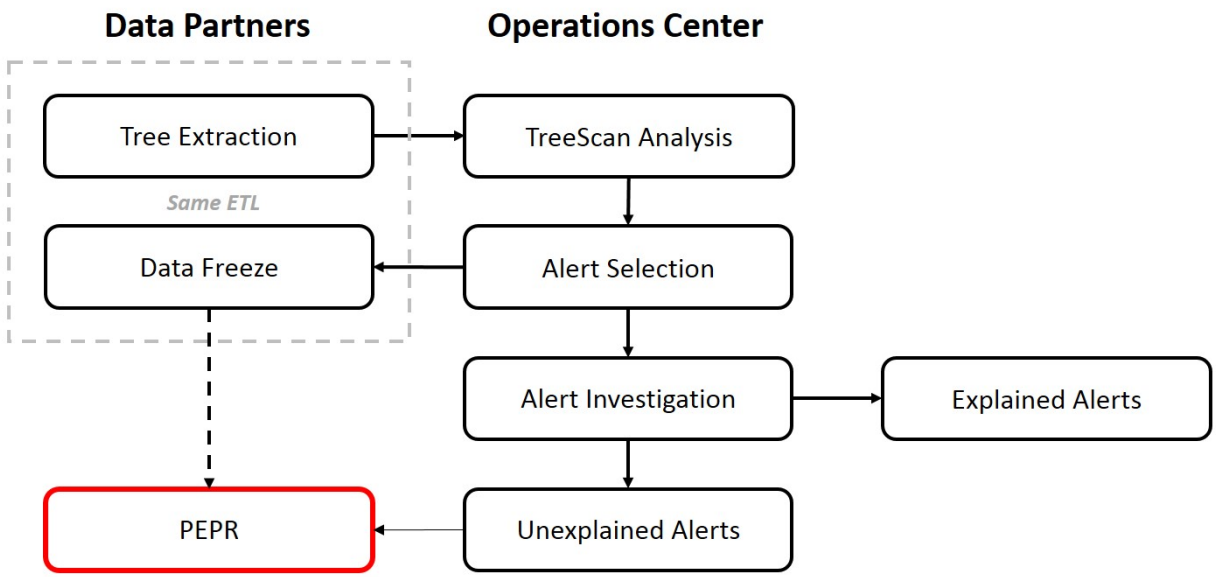
The Data Freeze package does not create any output datasets to be returned for review to the Operations Center. Only the SAS log and signature data set are returned for review by the programmer to confirm that the program ran without error.

Once the data are frozen, time may be taken to consider what, if any, follow-up investigation is necessary. Since Data Freeze is the time-critical step, and then follow-up analysis can be done at any point with the frozen datasets, care should be taken to select the widest range of alerts that may require follow-up while still being thoughtful in minimizing the size of files to be stored at the Data Partner sites. In general, far more alerts will be frozen than are of concern to the FDA, but due diligence requires the data be available in case a review becomes necessary. Additionally, alerts on the larger branches of the tree will likely have smaller branches that also produce alerts, which implies many of the saved alerts will be related and involve the same patients. Take a simple example of febrile seizures, which feeds into a higher node for convulsions. If we save the patients related to the convulsions node, then we would automatically capture those related to the febrile seizures node.

The frozen datasets are subject to standard Mini-Sentinel data retention policies (www.mini-sentinel.org/work_products/About_Us/Mini-Sentinel-Principles-and-Policies.pdf).

B. AIM 2: PATIENT EPISODE PROFILE RETRIEVAL (PEPR)

Figure 3: TreeScan Process Flow Adding PEPR. Note That PEPR Can Be Run at Any Time After Data Freeze, Even if the Original ETL is no Longer Available.



Many alert investigation tools are already available, including (but not limited to): checking programs and data for possible errors; conducting literature review for the exposure/outcome pair and coding practices for the outcome; reviewing descriptive statistics from the Tree Extraction data and the Mini-Sentinel modular programs or summary tables; and when using methods other than Tree Temporal Scan, looking for clusters in time from exposure to event.[10,11] PRISM and FDA also have experience with PBAs using chart review to validate outcomes and investigate associations with vaccine exposures,[1,2,3]

but PBAs are time-consuming and resource-intensive, making them cost-prohibitive as a routine tool for investigation of statistical alerts arising from data mining.

The second aim of this activity fills the gap between broader investigation tools and detailed PBAs by creating re-usable programs to extract and retrieve patient-level case data for review by FDA and PRISM investigators. This type of review is not intended to validate the outcome or determine the validity of an alert but rather to determine whether further investigation is warranted. It is also not intended as an automatic first option but instead should be used only on a small subset of alerts, if at all, and then only after careful consideration of the circumstances by the surveillance team.

Interest in similar capabilities from outside the TreeScan workgroup led us to develop the Patient Episode Profile Retrieval (PEPR) as a self-contained macro requiring a single input dataset stored at the Data Partner site to identify patient episodes of interest. At minimum, the input dataset must contain a PatID variable and at least one date variable. Since the PEPR macro itself does not determine the patient episodes to be included, other methods must first be used to identify patient episodes of interest and complete any required sampling or sub-setting.

1. Security

The PEPR output datasets are based on the CDM tables with certain modifications made to protect patient privacy. In order to strike a balance between the need for robust patient-level case data and minimum-necessary data requirements that are fundamental to the distributed network model, we implemented both mandatory and optional security measures.

a. Mandatory

Pseudo-identifiers are automatically assigned to replace four identifiers – PatID, EncounterID, Provider, and Facility_code – in the output files, and crosswalks are saved at the local sites to allow translation to the original. Each pseudo-identifier is assigned using sequential numbering to assure uniqueness, but the original values are first randomly sorted to further mask the identifier. A random seed parameter is included to assure the randomization process can be reproduced, if necessary.
Additional care is taken in assignment of the PatID pseudo-identifier to account for situations in which a single patient may contribute more than one episode to an analysis. To distinguish between separate episodes for the same patient in the output datasets, particularly when using relative dates (described in the next section), the PatID pseudo-identifier contains two parts: the first identifies the patient, and the second identifies the episode. For example, the current Tree Extraction program defines incidence at the third level of the tree. If an alert occurred at a higher level, a patient could contribute more than one incident event to the alert, and if the events occurred on different days with the 1 – 56 days after exposure, we would not be able to distinguish between those events to assign appropriate relative dates and preserve data integrity unless we assign the pseudo-identifier to the patient-episode combination rather than to the patient alone.

Future TreeScan methods development will add multi-dose analysis, i.e. inclusion of doses beyond the first observed per patient. Once again, depending on the incidence definition and the spacing of doses, a patient could conceivably contribute the same or a similar incident event following each dose. Further, with an eye toward extending use of PEPR beyond TreeScan, we considered PBAs such as the PRISM evaluation of febrile seizures following influenza vaccination in young children.3 Incidence was defined as first observed in 42 days. Since small children are recommended to have two doses of the vaccine, a single patient could have been identified as a case once for each dose if a seizure followed each dose.

That evaluation also looked at PCV and DTaP-containing vaccines, and those additional exposures could have resulted in identification of separate adverse event episodes if the vaccines were administered on different days.

b. Optional

Date variables may be masked by calibrating values to a meaningful relative index variable specified by macro parameter. The relative index date value is subtracted from each CDM date variable so that the index date now has a value of 0 (SAS date value = Jan. 1, 1960), and then any other date now represents the number of days before or after that index date. For Tree Temporal Scan using a self-controlled analysis, the natural choice for relative index is the exposure date, which is then set to 0, and then all other dates represent how many days before or after exposure the encounter, drug dispensing, enrollment start, or enrollment end occurred. The format of the date variables is preserved but the identifiable information removed. Thus, the reviewer will not know the actual calendar date of an event, only the number of days before or after the relative index.

Exposure date	Original ADate value	New ADate value	Unformatted numeric value
10/01/2011	10/01/2011	01/01/1960	0
10/01/2011	10/10/2011	01/10/1960	10
04/15/2006	04/15/2006	01/01/1960	0
04/15/2006	03/15/2006	12/01/1959	-31

If this option is not selected, the original calendar date values are included in the output datasets. An activity involving chart review, for example, would need the original calendar dates in order to match chart data with electronic data.
The only date variable not included in this relative date option is the birth date, since application of the rule would result in representation of the patient's exact age in days on any given relative date. Instead, another option allows for the birth date value to be set to missing for all patients. This rule is optional since the birth date value is necessary for activities involving chart review.
As a compromise, an additional option allows the programmer to specify age groupings to more broadly categorize each patient's age at a selected index date variable.

The final option concerns two variables in the CDM that contain geographic information represented by the ZIP code: Zip in the Demographic table, representing the last known patient ZIP code, and Facility location in the Encounter table, valued with the first three digits of the facility ZIP. An optional macro parameter allows the values to be set to missing, converted to standard postal state abbreviation, or to retain their original value. If geographic clustering is suspected as a confounder, the postal state value may be used while still transmitting a lower level of specificity than the actual five- or three-digit ZIP values. The actual ZIP values would rarely – if ever – be required for anything less than full chart review and should only be used with extreme caution.

Any altered variables follow the same data type and format as the original CDM variables, but the variable name is changed by adding an underscore character "_" to the beginning of the original variable name. For example, PatID becomes _PatID, and ADate becomes _ADate. This convention serves as a reminder to anyone working with the datasets that those variables have been altered for security purposes. In order to run programs that were written to run on the CDM tables, the programmer only

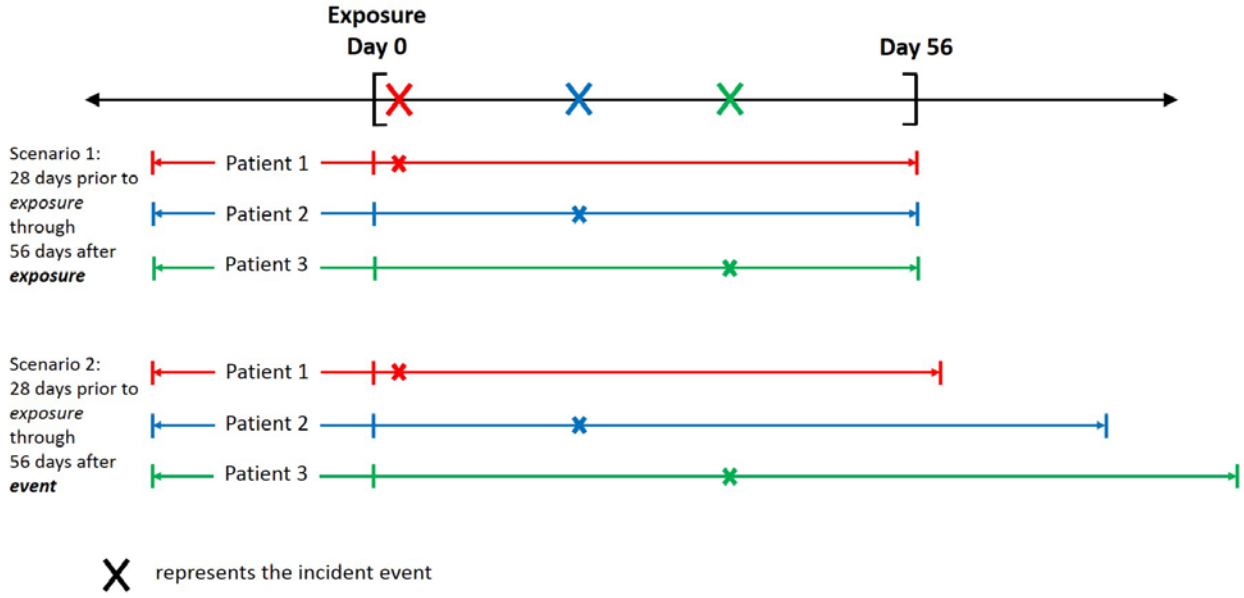
needs to change the variable names back to the original. (See Appendix B for a listing of altered variables.)

2. PEPR Window

In order to limit the amount of patient-level data returned to the Operations Center, the PEPR program selects a limited window of data for each patient episode included in the PEPR datasets as determined by two additional index date variables. Macro parameters are used to define these index dates, which means they can be changed as appropriate for each run of the PEPR macro. The PEPR window, then, is defined as follows: "pre" index date – "pre" index days prior through "post" index date + "post" index days after.

The simplest case is to use the same date variable for both indices, as we might do for a TreeScan analysis by indexing the entire window to the exposure date. That is, we could define the window as 28 days prior to exposure through 56 days after exposure, giving us equal history and follow-up time for all patients, relative to exposure. If instead we wish to have equal follow-up time relative to the event, we would define the window as 28 days prior to exposure through 56 days after event.

Figure 4: Effect of Using Different Index Date Variables to Determine PEPR Window.



Scenario 1 ensures equal follow-up time after exposure and equal calendar time for all episodes but results in varied follow-up time after event. Scenario 2 ensures equal follow-up time after even but results in varying total calendar time per episode

This use of macro parameters to specify the index dates extends the utility of the PEPR macro to a wider variety of evaluations. For example, in a pregnancy-related study, investigators may use an algorithm to estimate the pregnancy start date. Similarly, they may also use an algorithm to estimate the delivery date in an effort to be more precise than simply using the admission date of the encounter containing codes that relate to delivery. The estimated pregnancy start date could serve as the "pre" index date and the estimated delivery date as the "post" index date simply by entering the names of these two variables in the appropriate macro parameters. Another pregnancy-related evaluation may wish to look at mother-baby pairs, where records in the input dataset include information on both mom and baby.

PEPR could be run twice within the same workplan, once to extract the mother's data as described above and a second time to extract the baby's data.[4, 5]

3. Output Datasets

The PEPR output datasets include the modified CDM tables, as described above, and others that are not in CDM format, including the Core dataset and prevalence datasets for diagnosis, procedure, and dispensing. The Core dataset contains information pertaining to the index dates for each patient episode, most importantly the age grouping. All date values in the Core dataset are subject to the same security options as the other PEPR datasets. The prevalence datasets contain one record per unique combination of admission (or dispensing) date and clinical code, along with the number of days since the previous appearance of the same code in the patient's history. For first-observed, the prior days variable will have a missing value. The patient's entire history is used to determine the prior days value, but only those records occurring within the retrieval window will be included in the output datasets, which allows accurate categorization of diagnosis, procedure, and drug codes as first-observed vs. prevalent while adhering to the minimum-necessary data standard. Future development could improve the determination of prevalence by comparing codes across all tables, particularly since NDCs may appear in both the Dispensing and Procedure tables.

By retaining the CDM format of the main PEPR datasets, we allow for maximum flexibility and general use across any number of different types of evaluation. The familiar structure of the CDM tables can also make additional downstream programming easier and more efficient. A workgroup conducting a PBA with chart review could use these datasets in the chart selection process to determine which encounters the Data Partners should pursue and then use them as the basis for the chart review database. A different workgroup that finds a safety signal while conducting sequential analysis may also need to go to chart review, where PEPR could be used in the same manner. The general approach to both input and output allows for a wide range of applications, and the whole system benefits from access to a reliable, familiar tool while also saving resources associated with custom program development.

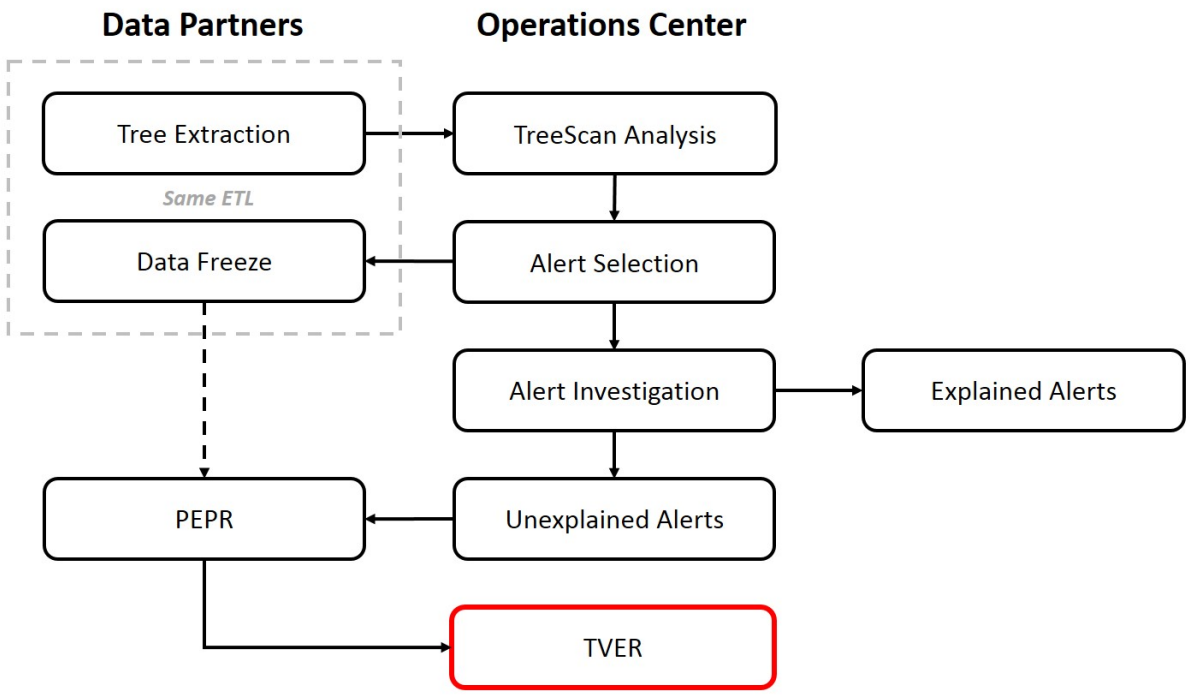
4. PEPR Workplan

The PEPR workplan requires a master program, the PEPR macro, various utility macros, and input macro parameters for each run of PEPR within the workplan. The PEPR macro itself may be run more than once within the same master program if multiple alerts are under investigation, saving the time and expense of multiple program runs by the Data Partners.

Auxiliary macros can be included in the workplan to prepare the input dataset, apply sampling or filters to the output datasets, or to retrieve additional information for output beyond the standard PEPR datasets. For TreeScan, we developed a macro to create and output two additional datasets for each alert. The first contains records from Crosswalk dataset with only the exact episodes which led to inclusion in the node, with the data de-identified in similar manner to the PEPR datasets. An optional variable may be added to this dataset with the calendar month of exposure to provide information on seasonality. The second output dataset is a copy of the Crosswalk dataset containing records of all incident events, regardless of whether they were related to the alert under investigation. This information will used in the report to help the clinician distinguish between events that were identified as incident by the Tree Extraction program, those that were otherwise first-observed, or those that were prevalent, i.e. neither incident nor first-observed.

5. Aim 3: TreeScan Vaccine Episode Report (TVER)

Figure 5: TreeScan Process Flow Adding TVER



The TVER answers the third aim of this activity to transform the PEPR output datasets into a format which facilitates review of vaccine-related TreeScan alerts. The report format is based in part on PRISM experiences with chart review studies, where chart selection reports are provided to clinician reviewers to prioritize encounters for which the Data Partners will pursue medical records. These reports have thus far been highly customized to each specific study population and exposure-outcome pair, but a more generic approach is necessary for TreeScan.

The report contains two sections: the Header with information pertaining to the overall vaccine episode; and the Detail with medical encounter and prescription drug dispensing information. The following sections describe the TVER after the format was modified following initial review during the MMR/MMRV beta-test.

6. TVER Data Preparation

A master dataset for each PEPR table is created by combining datasets from all Data Partners, and a new variable is added to distinguish between records from the different Data Partners. The unique values of this DataPartner variable are assigned a random numeric value and saved to a translation table, and then the programmer has the option to use the actual DataPartner value or the masked value in the report. Finally, a case identifier (CaseID) variable is added to provide a unique value to each patient-episode across all Data Partners.

7. TVER Header

The Header for each case consists of a single record containing the high-level information that relates to the patient-episode as a whole. Unique patient-episodes are identified by CaseID and DataPartner, masked or unmasked. The patient's demographics are represented by sex and age grouping at exposure. The calendar month of exposure and the number of days from exposure to event illustrate the timing of exposure and event, while the node, node description, and risk window of the alert are included to display overall alert information on every report screen.

The PEPR window gives the range of days captured in the PEPR datasets for each patient-episode, relative to exposure, showing the reviewer the range of potential days on which medical encounters and drug dispensings may be observed. If the exposure date has been used as both "pre" and "post" index, the PEPR window values will be the same for every patient-episode. If the event date is used as "post" index in order to ensure equal follow-up time after event for every episode, the PEPR window values will vary based on how many days after exposure the event occurred.

Example scenario 1: PEPR window selected as 28 days before exposure through 56 days after ***exposure***

Days between exposure and event	PEPR window start	PEPR window end
7	-28	56
28	-28	56
55	-28	56

Example scenario 2: PEPR window selected as 28 days before exposure through 56 days after ***event***

Days between exposure and event	PEPR window start	PEPR window end
7	-28	63
28	-28	84
55	-28	111

The coverage window variables serve the similar purpose of informing the reviewer of the potential range of days on which data may be seen in the Detail section, providing some assurance that data are not missing simply because the patient didn't have relevant coverage on certain days.

8. TVER Detail

The Detail section contains records from medical encounters and drug dispensings that occurred on any day within the PEPR window, with one record per unique combination of date (displayed as the number of days before or after exposure), clinical code (i.e. diagnosis, procedure, NDC), and encounter setting (blank for dispensing records). In addition to this basic information, we also include the length of stay (LOS) for inpatient encounters; primary diagnosis indication for inpatient encounters; incidence (incident as defined by the Tree Extraction program, first-observed in the patient's entire history, or prevalent); a node indicator to denote any diagnosis code which applies to the node under review; a main exposure indicator to denote any code which applies to the vaccine under review; an "any vaccine" indicator for any codes relating to vaccine administration; Rx days supply and Rx amount, relevant only to dispensing records; and coverage start and end dates (displayed as number of days relative to exposure) for the medical enrollment segment containing the admission date for medical encounter records or, for dispensing records, the drug coverage segment containing the dispensing date.

Since medical encounter and drug dispensing records are both included in the Detail section, the easiest way to recognize drug dispensing records is to look to the Rx days supply and amount variables, since these will only be populated for dispensing records. The setting (or encounter type) variable will also be missing for dispensing records.

In order to remove clutter in this potentially dense report, a blank is used as default value in several variables. For example, the incidence variable only shows a value for incident or first-observed records. If the code is prevalent, the variable is left blank. The same is true for the node, vaccine, and exposure indicators, where a value of 1 indicates yes and a blank indicates no.

9. Discussion

The current report is meant more as a prototype rather than a permanent solution. A SAS program is used to generate the Header and Detail datasets, but then the data are exported to Excel for presentation to the reviewer(s). This solution is not ideal, since it isn't flexible or scalable and can become quite cumbersome when more than a handful of patients are under review. Excel does have advantages in that it allows reviewers at least some minimal interactive capabilities to sort data, add custom flags, make electronic notes, etc. A more ideal solution would be web-based to enable real-time editing and sharing of custom flags, views, and notes between reviewers.

III. MMR/MMRV BETA TEST

In order to beta-test both local and distributed SAS programs, as well as to test the overall system using actual data to populate the TVER, we chose to re-enact an assessment of MMR and MMRV vaccines used in development of the Tree Temporal Scan statistic methods. We used three of the four original Data Partners (Harvard Pilgrim, HealthCore, and Humana) and the same time period that each contributed to the previous analysis on the assumption that no new alerts would be detected, since the power should be less than with the original four Data Partners. This allowed us to focus on the beta-test by using a previously studied vaccine and established set of statistical alerts. If the results were notably different, we could first suspect an issue with the SAS programs and review for bugs before proceeding.

The study period ranged from January 2004 through November 2011 for one Data Partner, June 2007 through October 2011 for another, and January 2000 through December 2011 for the third. MMR vaccine exposure were identified using ICD-9 procedure code 99.48, ICD-9 diagnosis code v06.4, and CPT4 code 90707. MMRV vaccine exposure was identified using CPT4 code 90170. Two age cohorts were selected, with group 1 representing 330 days to 760 days of age at exposure, roughly 11 to 25 months, and group 2 representing 1430 to 2220 days, roughly 47 to 73 months. The two vaccine exposures and two age groupings provided four analytic groups. The Tree Temporal Scan method was used for analysis, and the Tree Extraction program was set to identify and collect incident adverse events in the 1 to 56 days following the first-observed exposure in each analytic group.

After Tree Extraction datasets were returned from all Data Partner sites, the data were aggregated and run through the Tree Temporal Scan analysis. The Alert Selection package was then run on the results, with a total of 30 primary alerts selected according to pre-specified criteria of p-value less than or equal to .05 and relative risk greater than or equal to 1.2. We also applied a rule to select secondary alerts if the next-higher node on the tree for any primary alert had a p-value less than or equal to .20, but no additional nodes were selected with this criterion.

The selected alerts were mostly related to skin conditions (rash), convulsions (febrile), allergic reactions, and fluid and electrolyte disorders (nausea and vomiting). None of these were surprising, and in particular, the cluster of febrile seizures in 7 to 10 days following vaccination is consistent with findings in the literature.[13] The workgroup decided to manually add immune thrombocytopenic purpura (ITP) to the data freeze as a precaution, since excess cases were detected in the younger age group following MMR. However, the p-value was too high to meet the selection threshold.

From this set of alerts, two were chosen for the beta-test of the PEPR program, with both occurring in age group 1 (330 to 760 days). The first, alert 25 for Nausea and Vomiting, represented ICD-9 diagnosis codes 787.0, 787.01, 787.02, 787.03, and 787.04. There were 14 cases in the risk window of 5 to 9 days after MMRV exposure and 39 cases overall. The TreeScan workgroup decided to retrieve data from four weeks, or 28 days, prior to exposure through eight weeks, or 56 days, after the exposure to ensure capture of the entire 1 to 56 day exposure window and to gather uniform calendar time for all patients. PEPR was run at all three Data Partners for this alert.

Table 1: Nodes related to Alert 25 (Convulsions in 7 – 10 days after MMR vaccine exposure)

Alert ID	Vaccine	Age group	Node	Node description	Total node cases	Risk window	Observed	Expected	Relative risk	Excess Cases	P-value
26	MMRV	1	17.01.06	Nausea and vomiting	39	5 - 9	14	3.48	5.71	11.55	0.01856
25	*MMRV*	*1*	*17.01.06.00*	*Nausea and vomiting*	*39*	*5 - 9*	*14*	*3.48*	*5.71*	*11.55*	*0.01856*
24	MMRV	1	787.03	Vomiting alone	34	5 - 9	13	3.04	6.31	10.94	0.01779

The second, alert 4 for Convulsions, represented ICD-9 diagnosis codes 780.3, 780.31, 780.32, 780.33, and 780.39. There were 80 cases in the risk window of 7 to 10 days after MMR exposure and 290 cases overall. The TreeScan workgroup decided to retrieve data from eight weeks (56 days) prior to exposure through twelve weeks (84 days) after the event to ensure uniform follow-up to event for all patients. Because of the relatively high number of cases, we chose to send run PEPR for this alert at only two Data Partners, which limited the number of cases for review to 13.

Table 2: Nodes related to Alert 4 (Convulsions in 7 – 10 days after MMR vaccine exposure)

Alert ID	Vaccine	Age group	Node	Node description	Total node cases	Risk window	Observed	Expected	Relative risk	Excess Cases	P-value
2	MMR	1	06	Diseases of the nervous system and sense organs	678	7 - 11	156	60.54	3.05	104.8	0.00001
1	MMR	1	06.04	Epilepsy; convulsions	317	7 - 10	88	22.64	5	70.38	0.00001
4	*MMR*	*1*	*06.04.02*	*Convulsions*	*290*	*7 - 10*	*80*	*20.71*	*4.95*	*63.85*	*0.00001*
3	MMR	1	06.04.02.00	Convulsions	290	7 - 10	80	20.71	4.95	63.85	0.00001
10	MMR	1	780.31	Febrile convulsions simple unspec	178	7 - 10	50	12.71	5.08	40.15	0.00001
11	MMR	1	780.39	Other convulsions	95	7 - 9	25	5.09	6.31	21.04	0.00001

For both alerts, the following options were selected:

1) Set the patient birth date to missing.
2) Set the patient and facility ZIP codes to missing.
3) Categorize age at exposure in three-month bands, i.e. 0M-12M 12M-14M 15M-17M 18M-20M 21M-23M 24M+, where the first and last groupings represent the tails of the overall age cohort.
4) Set all calendar dates relative to the exposure date.
5) Include the calendar month of exposure in the PEPR Core output table.
6) Output the Demographic, Enrollment, Encounter, Diagnosis, Procedure, and Dispensing tables.
7) Output the PEPR Core table.
8) Run the Alert Crosswalk Retrieval auxiliary macro to output de-identified copies of the two alert crosswalk datasets described in the PEPR section above.

The TVER was produced for both alerts, and since neither outcome was surprising, the group did not conduct a serious review of the clinical data but instead reviewed the reports to tune format and content as well as to gain familiarity with actual data. This initial review led to modifications of the TVER format to collapse separate tables for medical encounters and drug dispensing detail into a single Detail section, add the PEPR window information to the Header section, and set all default values in the Detail section to blank to remove clutter.

The group also used this review as an opportunity to evaluate whether these data could be useful for the stated purpose of determining the likelihood of an event being caused by vaccine, mostly as a "likely rule-out" tool. As an example, we have included below in Tables 3 and 4 the TVER for a composite case from the nausea and vomiting alert.

In the header we can see from the "Days from expos to event" variable that the nausea and vomiting event occurred 6 days after exposure to MMRV vaccine, placing this case within the 5 to 9 day risk window identified by Tree Temporal Scan analysis. In the detail section, we see that the patient received the MMRV vaccination as part of a routine visit and that PCV7 vaccine was also given on the same day.

Four days after exposure, the patient had another ambulatory visit encounter, where we see a first-observed diagnosis code of 009.0 (infectious colitis, enteritis, and gastroenteritis), a code that was pruned from the tree because the cause is specifically noted as infection.

Three days later, the patient was apparently taken to the emergency department and then admitted to the hospital with the primary diagnosis of 276.51 (dehydration) as well as secondary diagnosis of 787.03 (vomiting alone), the incident code which led to inclusion in this particular node. The patient was administered fluids through IV, and multiple tests were ordered. The patient stayed overnight, as indicated by the value of 1 in the LOS (length of stay) variable. Note that all four diagnosis codes listed for this inpatient encounter, including 786.2 (cough) and 535.50 (unspecified gastritis and gastroduodenitis without mention of hemorrhage), were identified as incident by the Tree Extraction program, but only the code for vomiting belongs to the node for this alert.

Two days after discharge, the patient had one more ambulatory visit encounter where the only diagnosis given was again 009.0, which may have been based on confirmation from test results or simply as an indication of history from the previous ambulatory visit on day 4 after exposure.

Although the patient had medical and drug coverage far beyond the end of the PEPR window (56 days after exposure), no further follow-up visits were observed during that time.

Based on these data, we would probably conclude the vomiting episode was more likely due to gastro-intestinal infection than the exposure to MMRV vaccine.

Table 3: TVER example header and detail for composite patient

Episode Header		~ Coverage refers to the enrollment segment containing the admission date of the encounter													
Node	Node desc	Data Partner	Case ID	Sex	Age band at expos	Month of expos	Days from expos to event	Risk win start	Risk win end	PEPR win start	PEPR win end	Med cov start~	Med cov end~	Drug cov start~	Drug cov end~
17.01.06.00	Nausea and vomiting	1	40	F	12M-14M	JAN	7	5	9	-28	56	-386	1260	-386	1260

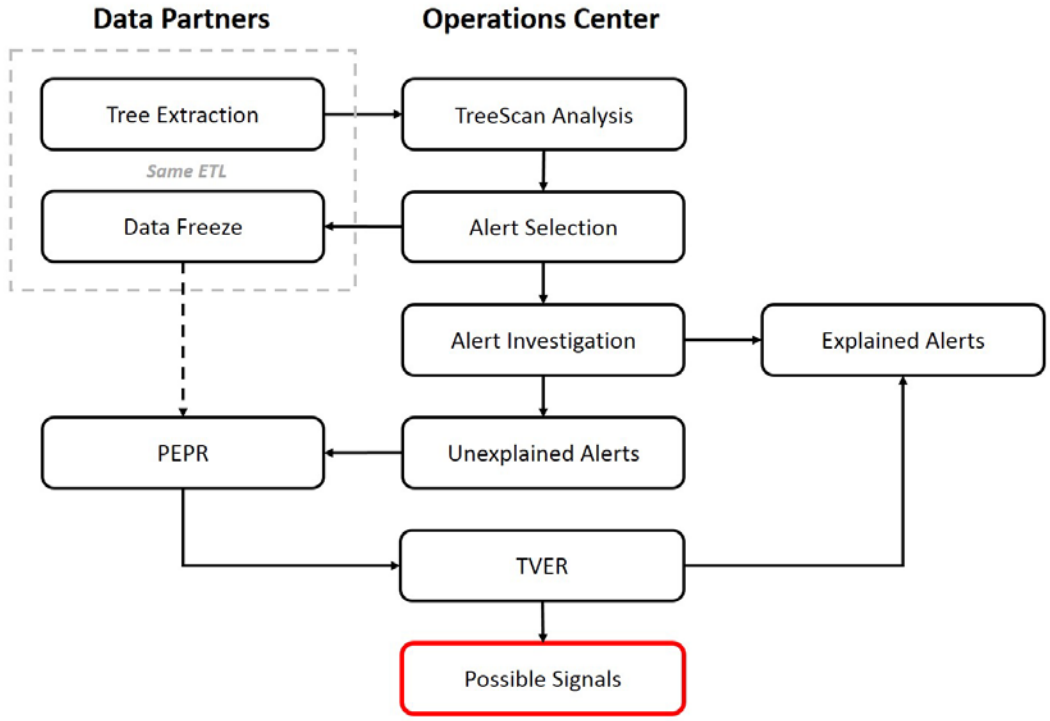
Table 4: TVER example header and detail for composite patient

Episode Detail	^ Incidence: F = first observed; I = incident; blank = prevalent # Primary Dx: P = primary; S = secondary; X = N/A ~ Med enroll segment containing the admission date of the encounter or the drug enroll segment containing the dispensing date

Days from expos	Enc type	LOS	Clinical code			Code description	Incidence^	P Dx#	Node (Y/N)	Main expos (Y/N)	Any vacc (Y/N)	Rx days supp	Rx amt	Cov start~	Cov end~
			Cat	Type	Code										
0	AV		DX	09	V0382	Need Proph Vacc Agnst Strep Pne					1			-386	1260
0	AV		DX	09	V068	Need Proph Vacc Against Oth Comb Dz	F				1			-386	1260
0	AV		DX	09	V202	Routine Infant/Child Health Check								-386	1260
0	AV		PX	C4	90471	Immunization Admin	F				1			-386	1260
0	AV		PX	C4	90472	Immunization Admin Each Add	F				1			-386	1260
0	AV		PX	C4	90669	PCV7 Vaccine Im					1			-386	1260
0	AV		PX	C4	90710	MMRV Vaccine Sc	F			1	1			-386	1260
0	AV		PX	C4	99392	Prev Visit Est Age 1-4	F							-386	1260
4	AV		DX	09	0090	Inf Colitis Enterit & Gastroenterit	F							-386	1260
4	AV		PX	C4	99213	Office/Outpatient Visit Est	F							-386	1260
7	IP	1	DX	09	27651	Dehydration	I	P						-386	1260
7	IP	1	DX	09	53550	Uns Gastrit & Gastroduodit No Hemorr	I	X						-386	1260
7	IP	1	DX	09	7862	Cough	I	X						-386	1260
7	IP	1	DX	09	78703	Vomiting Alone	I	S	1					-386	1260
7	IP	1	PX	C4	71020	Chest X-Ray 2Vw Frontal & Latl	F							-386	1260
7	IP	1	PX	C4	74000	X-Ray Exam Of Abdomen	F							-386	1260

IV. CONCLUSIONS AND RECOMMENDATIONS

Figure 6: TreeScan Process Flow with Full Programming and Data Management Infrastructure



The data captured by PEPR and displayed in TVER can fill the gap in our investigative toolset between broader preliminary investigation and full-scale PBAs with chart review by:

1) Providing clinical context
    a. on key co-morbidities, pre-exposure events, and risk factors that occur before the exposure of interest;
    b. on the temporal sequence of events as coded in administrative data;
    c. on key medical procedures and fuller details of the clinical evaluation on the actual date of the event, to potentially increase or decrease the index of suspicion about the accuracy of the code;
    d. on key follow-up healthcare encounters and their evaluation that might increase the precision of a health outcome of interest (i.e., if the patient subsequently is treated with anti-coagulants, this increases the likelihood that the original VTE event was accurate)
2) Providing stakeholders an additional opportunity to increase the precision when alerts contain ambiguous diagnostic codes or codes that contain numerous actual diagnoses
3) Displaying potential for use as a follow-up tool for statistical signals emerging from use of the Routine Querying System (http://www.mini-sentinel.org/data_activities/modular_programs/details.aspx?ID=166) along with standard analytical tools such as Prospective Routine Observational Monitoring Program Tools (PROMPT) (http://www.mini-sentinel.org/methods/methods_development/details.aspx?ID=1045)

4) Maintaining consistency with Congressional FDAAA mandate to maximize the use of electronic healthcare data for active risk identification and risk analysis (ARIA)
5) Allowing FDA to prioritize finite resources for chart review by using these tools to refine and triage alerts

This last point is particularly important, as triage of statistical alerts by stakeholders will likely **decrease** the need to obtain more granular, individual-level health information that would be required as part of medical record review.

For wider applications, the TVER format could be modified for use with other types of exposures beyond vaccines or other types of evaluations beyond data mining. For drug exposures, dispensing records could be grouped into dispensing episodes with or without allowances for stockpiling, and variables could be added to describe drug therapeutic classes and other higher-level groupings of NDCs that would be of interest to reviewers. Such development could include a report for use with DrugScan (http://www.mini-sentinel.org/methods/methods_development/details.aspx?ID=1061), an outcome-based version of TreeScan that chooses a single adverse event and looks for unanticipated exposures that may be associated with the adverse event. Functionality could be added to the existing Routine Querying System to enable use of PEPR to gather data for initial review or for chart review of statistical alerts or signals resulting from standard analytical tools such as PROMPT.

Alerts involving a large number of patients present problems that should be addressed through further development. The potential need to gather patient-level data, even if de-identified, on hundreds of patients to investigate a data mining alert presents a challenge to the principle of minimum-necessary data disclosure, and TVER in its current state has limited functionality, making high-volume review difficult, particularly for more complex outcomes.

These issues could be addressed by:

1) developing standard aggregate analyses, such as rank order listing of most frequent diagnosis, procedure, and drug codes, to be run on the frozen datasets behind the firewall;
2) adding an auxiliary macro for use with PEPR to take a random sample from each Data Partner;
3) adding an auxiliary macro for use with PEPR to subset the cases according to specific criteria, such as limiting to cases in the risk window (Tree Temporal Scan) or limiting to exposed cases (Poisson);
4) enhancing TVER and adding more sophisticated reporting tools at the Operations Center to facilitate review of larger numbers of cases and of more complex outcomes;
5) working with the Data Partners to adjudicate some or all cases at the sites.

Ultimately, the infrastructure developed through this activity and described in this report can only remain effective if ongoing maintenance and enhancements are included as part of standard operating procedures to ensure continued compatibility with all current and future TreeScan activities. Since this activity only had resources to address one analysis method, the Tree Temporal Scan, all pieces of the infrastructure should be assessed, adapted, and implemented for other existing methods, including Poisson and self-controlled. Any modifications or enhancements to Tree Extraction, the diagnosis tree (including implementation of an ICD-10 tree or a combined ICD-9/ICD-10 tree), or the TreeScan analysis results file should be communicated in advance to the responsible programmer to ensure compatibility with existing infrastructure.

V. APPENDIX A: TREE EXTRACTION OUTCOME CROSSWALK DATA DICTIONARY

Variable	Data Type	Format	Label	Valid Values	Source/Comments
PatID	C(varies by Site)	$##.	Patient ID	Unique alpha-numeric identifier	Pseudoidentifier assigned to a unique patient by the data partner and which can be used to link across tables in the MSCDM
Dx	C(18)	$18.	Diagnosis code	Valid diagnosis code	Code used to identify incident outcome event
Dx_codetype	C(2)	$2.	Diagnosis code type	09 = ICD-9-CM 10 = ICD-10-CM 11 = ICD-11-CM SM = SNOMED CT OT = Other	Includes all code types for current and possible future use.
DX_ADate	N(4)	MMDDYY 10.	Diagnosis admission date	Valid calendar date within parameters of the study period	ADate of the encounter from which the incident diagnosis was identified.
Exp_ADate	N(4)	MMDDYY 10.	Exposure admission date	Valid calendar date within parameters of the study period. Will be blank if the outcome did not occur during the risk or control window.	ADate of the encounter from which the exposure was identified. For AV and ED, this generally corresponds to the date of service. For IP, this is the admission date but not necessarily the date of service.
Group	C(30)	$30.	Group	alpha-numeric	Describes analysis groupings, usually a combination of vaccine exposure and age

VI. APPENDIX B: PEPR ALTERATIONS TO CDM TABLE VARIABLES

CDM Table	Variable	Transformation
All	_PatID	Always masked; crosswalk retained in local folder
Demographic	Birth_date	If prefixed with underscore, set to missing. Otherwise, original value
Demographic	Zip	If prefixed with underscore, set to missing or converted to postal state abbreviation, depending on macro parameter selection. Otherwise, original value
Demographic	Zip_Date	If prefixed with underscore, calculated as Zip_Date – relative index date value, i.e. days +/- the relative index date. Otherwise, original value
Enrollment	Enr_start	If prefixed with underscore, calculated as Enr_start – relative index date value, i.e. days +/- the relative index date. Otherwise, original value

CDM Table	Variable	Transformation
Enrollment	Enr_end	If prefixed with underscore, calculated as Enr_end – relative index date value, i.e. days +/- the relative index date. Otherwise, original value
Encounter, Diagnosis, Procedure	_EncounterID	Always masked; crosswalk retained in local folder
Encounter, Diagnosis, Procedure	ADate	If prefixed with underscore, calculated as ADate – relative index date value, i.e. days +/- the relative index date. Otherwise, original value
Encounter	DDate	If prefixed with underscore, calculated as DDate – relative index date value, i.e. days +/- the relative index date. Otherwise, original value
Encounter, Diagnosis, Procedure, State Vaccine	_Provider	Always masked; crosswalk retained in local folder
Encounter	Facility_location	If prefixed with underscore, set to missing or converted to postal state abbreviation, depending on macro parameter selection. Otherwise, original value
Encounter, Lab Results	_Facility_code	Always masked; crosswalk retained in local folder
Dispensing	RxDate	If prefixed with underscore, calculated as RxDate – relative index date value, i.e. days +/- the relative index date. Otherwise, original value
State Vaccine	SIIS	If prefixed with underscore, set to missing Otherwise, original value
State Vaccine	VaxDate	If prefixed with underscore, calculated as VaxDate – relative index date value, i.e. days +/- the relative index date. Otherwise, original value
Lab Results	Order_dt	If prefixed with underscore, calculated as Order_dt – relative index date value, i.e. days +/- the relative index date. Otherwise, original value
Lab Results	Lab_dt	If prefixed with underscore, calculated as Lab_dt – relative index date value, i.e. days +/- the relative index date. Otherwise, original value
Lab Results	Result_dt	If prefixed with underscore, calculated as Result_dt – relative index date value, i.e. days +/- the relative index date. Otherwise, original value
Vital Signs	Measure_date	If prefixed with underscore, calculated as Measure_date – relative index date value, i.e. days +/- the relative index date. Otherwise, original value
Death	DeathDate	If prefixed with underscore, calculated as DeathDate – relative index date value, i.e. days +/- the relative index date. Otherwise, original value

VII. ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of the following organizations and individuals:

Harvard Pilgrim Health Care Institute: Jeffrey Brown, Claudia Colonel-Moreno, Alison Kawai, Matt Lakoma, Jen Popovic, Lisa Ortendahl, Megan Reidy, Schyler Versteeg, Tiffany Woodworth, Diana Santiago, Robert Rosofsky

HealthCore: Bolarinwa Ekezue, Chunfu Liu, Fang Tian

Humana: Mano Selvan, Yunping Zhou

VIII. REFERENCES

1. Yih WK, Lieu TA, Kulldorff M, Martin D, McMahill-Walraven CN, Platt R, Selvam N, Selvan M, Lee GM, Nguyen M. Intussusception Risk after Rotavirus Vaccination in U.S. Infants. N Engl J Med, 2014, 370:503-12.
2. Yih WK, Zichittella LJ, Greene SK, et al. Evaluation of the Risk of Venous Thromboembolism After Gardisil Vaccination. Report to FDA; April 23, 2015. (Accessed at http://mini-sentinel.org/work_products/Assessments/Mini-Sentinel_PRISM_Gardasil-and-Venous-Thromboembolism-Report.pdf)
3. Kawai AT, Martin D, Kulldorff M, Li L, Cole DV, McMahill-Walraven CN, Selvam N, Selvan MS, Lee GM. Febrile Seizures After 2010–2011 Trivalent Inactivated Influenza Vaccine. Pediatrics, 2015-0635.
4. Kawai AT, Li L, Kulldorff M, et al. Mini-Sentinel CBER/PRISM Surveillance: Influenza Vaccines and Pregnancy Outcomes. Protocol for FDA; March 28, 2014. (Accessed at http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Influenza-Vaccines-and-Pregnancy-Outcomes-Protocol.pdf)
5. Kawai AT, Li L, Andrade SE, et al. Mini-Sentinel/CBER Protocol: Influenza Vaccines and Birth Outcomes. Protocol for FDA; December 5, 2014. (Accessed at http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Influenza-Vaccines-and-Birth-Outcomes-Protocol.pdf)
6. Yih WK, Zichittella LJ, Sandhu SK, et al. Accessing the Freshest Feasible Data for Conducting Active Influenza Vaccine Safety Surveillance. Report to FDA; April 8, 2015. (Accessed at http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Active-Influenza-Vaccine-Safety-Surveillance-Report.pdf)
7. TreeScan Extension (PRISM). Methods project for FDA; in progress. (Accessed at http://www.mini-sentinel.org/methods/methods_development/details.aspx?ID=1056)
8. TreeScan Power. Methods project for FDA; in progress. (Accessed at http://mini-sentinel.org/methods/methods_development/details.aspx?ID=1055)
9. Yih WK, Nguyen M, Maro JM, et al. Mini-Sentinel CBER/PRISM Surveillance: Pilot of Self-Controlled Tree-Temporal Scan Analysis for Gardasil Vaccine. Protocol for FDA; March 30, 2015. (Accessed at http://mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Pilot-Self-Controlled-Tree-Temporal-Scan-Analysis-Gardasil-Vaccine-Protocol.pdf)
10. McClure DL, Raebel MA, Yih WK, Shoaibi A, Mullersman JE, Anderson-Smits C, Glanz JM Mini-Sentinel methods: framework for assessment of positive results from signal refinement. Pharmacoepidemiology and Drug Safety. 2014;23:3-8.
11. Yih WK, Kulldorff M, Fireman BH, et al. Active surveillance for adverse events: the experience of the Vaccine Safety Datalink Project. Pediatrics. 2011;127:S54–64.
12. Kulldorff M. TreeScan User Guide, v1.1. November, 2014; 37-38. (Accessed at http://www.treescan.org/cgi-bin/treescan/register.pl/treescanv1.1-userguide.pdf?todo=process_userguide_download)
13. Klein NP, Fireman B, Yih WK, Lewis E, Kulldorff M, Ray P, Baxter R, Hambidge S, Nordin J, Naleway A, Belongia EA, Lieu T, Baggs J, Weintraub E, for the Vaccine Safety Datalink. Measles-mumps-rubella-varicella combination vaccine and the risk of febrile seizures. Pediatrics, 2010, 126, e1-8.