SENTINEL METHODS REPORT

MONITORING ALL DRUGS FOR A SPECIFIC OUTCOME IN THE SENTINEL SYSTEM

Prepared by: Judith C. Maro, PhD, MS,[1] Martin Kulldorff, PhD,[2] Inna Dashevsky, MS,[1] David V. Cole, BM,[1] Jeffrey S. Brown, PhD,[1] Gerald J. Dal Pan, MD, MHS,[3]

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

February 15, 2017

The Sentinel System is sponsored by the U.S. Food and Drug Administration (FDA) to proactively monitor the safety of FDA-regulated medical products and complements other existing FDA safety surveillance capabilities. The Sentinel System is one piece of FDA's Sentinel Initiative, a long-term, multi-faceted effort to develop a national electronic system. Sentinel Collaborators include Data and Academic Partners that provide access to healthcare data and ongoing scientific, technical, methodological, and organizational expertise. The Sentinel Coordinating Center is funded by the FDA through the Department of Health and Human Services (HHS) Contract number HHSF223201400030I. This project was funded by the FDA through HHS Mini-Sentinel contract number HHSF22301007T.

Sentinel

Sentinel Methods Report

Monitoring All Drugs For A Specific Outcome In The Sentinel System

Table of Contents

I. INTRODUCTION

The tree-based scan statistic method is a signal detection approach for postmarket medical product surveillance.[1,2] This method allows an investigator to cast a wide net to search for unexpected potential associations between exposures and outcomes of interest. The tree-based scan statistic detects statistically significant associations in electronic health data that have been grouped into hierarchical tree structures.[3] Specifically, using log-likelihood ratios and Monte Carlo hypothesis testing, the tree-based scan statistic tests whether any position in the tree structure is associated with an elevated frequency of observed events (i.e., in excess of the expected value), adjusting for the multiple testing inherent in the many overlapping hypotheses considered. More information on the method and its tree-based scan statistics can be found in Appendix A.

Outcomes with statistically significant elevated frequencies generate "alerts" that must be carefully evaluated using other clinical and pharmacoepidemiologic methods that are more tailored to evaluate risks for a specific exposure-outcome pair of concern. In effect, these data-mining analyses are hypothesis-generating in that they produce an initial warning with respect to potential associations, and are a form of signal detection.

Historically, the FDA and others have focused on individual vaccines or drugs, or drug classes, and specific medical outcomes experienced by users of those medical products. In the same vein, Nelson et al. have published a review paper on data-mining methods that are grounded in a vaccine (i.e., exposure-based) context.[4] Most studies that have used the tree-based scan statistic have been exposure-focused in that they set the exposure, such as a drug or vaccine, and scan a hierarchical tree of outcomes.[1,2,5]

However, it is also possible to set the outcome of interest and scan a hierarchical tree of drug exposures. Certain outcomes are of immediate interest to FDA, and are proposed for expedited reporting.[6] In an outcome-oriented setting, the main advantage of the tree-based scan statistic is the ability to detect alerts arising from drug-specific effects as well as class-wide effects all while formally controlling for multiple hypothesis testing. Similar outcome-centered data-mining projects using different methodologies and alerting algorithms have been performed in data used by the EU-ADR consortium,[7,8] the Observational Medical Outcomes Partnership,[9] and others.[10] By assessing individual medical outcomes across thousands of drug exposures, there is potential to generate new knowledge about these outcomes in a broader context, which may influence the data collected for future postmarket surveillance or premarket testing.

The primary objective of this study was to use tree-based scan statistics with a self-controlled case-crossover design using the TreeScan™ software (http://www.treescan.org) in the Sentinel Distributed Database (SDD) to determine which drugs, or classes of drugs, are associated with a particular subsequent medical outcome. This was the first evaluation in this setting and was intended to answer preliminary questions regarding the feasibility of TreeScan™ to produce information on drug-associated serious adverse events that are "rare" to "infrequent" and likely to appear in the FDA Adverse Event Reporting System (FAERS).[11] Because this evaluation was primarily exploratory, we performed multiple related analyses to understand their strengths and limitations without naming a primary or secondary analysis.

II. METHODS

A. STUDY PERIODS, POPULATIONS, DATA SOURCES

We performed data-mining analyses for the period 2000-2014 using claims data from three Sentinel Data Partners. Each site contributed data from their earliest available date through their latest date of complete data availability, meaning that not all partners contributed for all years. Eligible members of the study population were females and males greater than or equal to 18 years old at the time of the incident diagnosis of interest who had both medical and drug coverage. Enrollment gaps of 45 days or less were bridged and treated as continuously enrolled time.

Appendix B provides a step-by-step walkthrough of how the study cohort was created.

B. SELF-CONTROLLED CASE-CROSSOVER DESIGN AND OUTCOMES OF INTEREST

We chose a self-controlled case-crossover design[12,13] because it is outcome- or case-oriented (as opposed to exposure-oriented), and minimizes bias by automatically adjusting for time-invariant confounding. In this design, incident outcomes were identified. Then, incident exposures in a period preceding the outcome were identified. These incident exposures then serve as the index date for an observation window that followed the incident exposure. The observation window always contained the incident outcome by design. In a self-controlled case-crossover design, the important analytic information is the time from the incident exposure to the incident outcome. More details of this design construction are in Appendix B.

We chose angioedema and Achilles tendon rupture as outcomes of interest because:

1. They have validated algorithms in claims databases.
2. They each have a positive control (i.e., these outcomes are known to be associated with one or more specific drug classes).
3. Outcome onset is temporally close to drug exposure (i.e., little or no latency/induction period).
4. They are unlikely to be subject to confounding by indication.
5. There are expected to have a sufficient number of events in the data to perform the analyses.

1. Angioedema

Angioedema has been well-studied within the Sentinel Distributed Database, and has a positive predictive value of 90-95% according to validation studies done outside the Sentinel system.[14] This outcome was expected to occur on the order of 1-2 cases /1000 person-years contributed, with the majority of cases occurring in the first 30 days following various drug exposures.[14] These data were comparable with other studies.[15]

Angioedema was identified by ICD-9-CM diagnosis code 995.1 in the outpatient, inpatient, or emergency department setting, for which there were no other occurrences of this diagnosis code in the outpatient, inpatient, or emergency department settings in the preceding 64 days. For patients who had more than one qualifying incident diagnosis (i.e., diagnoses separated by more than 64 days), we limited their contributed data to the first incident diagnosis.

2. Achilles Tendon Rupture

The algorithm for Achilles tendon rupture had a positive predictive value of 86%.[16] Achilles tendon rupture was expected to occur on the order of 0.6-1.8 cases /10,000 person-years contributed, and had been shown to occur within weeks of certain drug exposures, but also up to 90 days.[17]

Achilles tendon rupture was identified by ICD-9-CM code 727.67 in the outpatient, inpatient, or emergency department setting. Of this universe of patients, the cohort was further restricted to patients that also had a CPT code for Achilles tendon rupture repair (27605, 27606, 27650, 27652, 27654, 01472) in any setting in the 30 days following or prior to the appearance of 727.67. The index date was set to the earlier of the ICD-9-CM code or CPT code date. Once the index date was set, neither the original ICD-9-CM code nor any of the related CPT codes could have occurred in the outpatient, inpatient, or emergency department setting in the preceding 127 days, i.e. four months. For patients that had more than one qualifying incident diagnosis (i.e., diagnoses separated by more than 127 days), we limited their contributed data to the first incident diagnosis.

C. INCIDENT EXPOSURE DEFINITIONS AND DRUG TREE

The study focused on identifying incident exposures of interest observed prior to the outcome of interest. Exposures were identified and defined using National Drug Codes (NDCs) and classified into a hierarchical tree defined using Medi-Span's Therapeutic Classification System (Wolters Kluwer Health, Inc., Conshohocken, PA). An example tree is in Table 1 below.

Table 1. Example Multi-level Therapeutic Classification System

Level	Generic product identifier	Coding	Example
1	36-	Drug Group	Antihypertensives
2	36-10	Drug Class	Angiotensin converting enzymes (ACE) inhibitors
3	36-10-00	Drug Sub-class	ACE inhibitors
4	36-10-00-30	Drug name	Lisinopril
5	36-10-00-30-00	Drug name extension	Lisinopril
6	36-10-00-30-00-03	Dosage form	Lisinopril Tablet
7	36-10-00-30-00-03-03	Strength	Lisinopril Tablet 2.5 MG

An exposure was an incident exposure if it was first dispensed within the 64 days preceding an angioedema event or the 127 days preceding an Achilles tendon rupture event, respectively, and there was no previous dispensing of that drug or any other drug located in the same level in Medi-Span's Therapeutic Classification System in the preceding 127 days relative to the identified dispensing. We performed analyses where we assigned incidence at the 3rd level (i.e., Drug sub-class level) and at the 4th level (i.e., Drug name level). We use an example to illustrate. If a person who experienced incident angioedema were dispensed lisinopril 60 days prior to the event and dispensed enalapril (i.e., another drug in the same sub-class as lisinopril) 30 days prior to the event and met all the other required criteria, then both drug products would be counted as exposure-outcome pairs in the analytic dataset when incidence was assigned to the 4th level. However, because both drugs are in the same drug sub-class (i.e., ACE inhibitors), when incidence criteria were assigned to the 3rd level, only the lisinopril exposure-outcome pair would be counted in the analytic dataset because it was the first ACE inhibitor exposure in the period preceding the incident outcome.

One person may have contributed multiple exposure-outcome pairs in these data as long as they met the appropriate incidence criteria. If a member were dispensed multiple drugs at the *same level of the tree on the same day*, then only one exposure was allowed in the analytic dataset. In this situation, the exposure was selected by:

1. Choosing the exposure with the longest days of supply dispensed.
2. If there were multiple exposures that could be incident on the same day with the same days of supply dispensed, the exposure with the smallest count of unique users over the study period was used.

D. DATA-MINING ANALYSES AND STATISTICAL METHODS

1. Tree-based Scan Statistics for Self-controlled Data

Recall, that the important analytic information in the self-controlled case-crossover design is the time-to-event from the incident exposure to the incident outcome, which occurs during an observation window following the incident exposure. The observation window for analyses (i.e., days 1-63 post-exposure in the angioedema analyses and days 1-126 post-exposure in the Achilles tendon rupture analyses) is divided into a risk window and a comparison window. In a fixed risk window analysis, these values are defined *a priori* as listed in **Table 2** below (e.g., days 1-28 are the risk window and days 29-63 are the comparison window). In a varying risk window analysis, these windows are allowed to move within the overall observation window. See Appendix B for illustrations of these windows.

The self-controlled case-crossover design is compatible with the following four tree-based scan statistics: conditional and unconditional Bernoulli tree-based scan statistics for fixed risk window analyses, and conditional and unconditional tree-temporal analyses for varying risk window analyses.

The tree-based scan statistic uses a log-likelihood ratio test to detect elevated frequencies of outcomes in electronic health data that have been grouped into hierarchical tree structures.[1,3] The test statistic cannot be determined analytically and therefore is computed via Monte Carlo hypothesis testing. The tree-based scan statistic automatically adjusts for multiple overlapping testing inherent to data-mining. Performing a conditional analysis is a mechanism to control for situations when there is an across-the-board increase in healthcare utilization during a particular period that is unrelated to the exposure of interest. This situation might occur commonly when the cohort has follow-up tests or visits in the days immediately following their initial dispensing of a medication. The conditional variants of the tree-based scan statistic attenuate the effect of this increased healthcare utilization potentially unrelated to the exposure by standardizing all diagnoses by the frequency with which they appear in the dataset. Mathematical expressions for all versions of the tree-based scan statistic used in this report can be found in Appendix A.

In the unconditional forms of the tree-based scan statistic, the null hypothesis assumes outcomes to be uniformly distributed across the observation window following the incident dispensing. Under the alternative hypothesis, there is at least one drug, or class of drugs, for which there is a temporal cluster of study outcomes during some time interval in the observation window. In the conditional forms of the tree-based scan statistic, the outcomes are standardized by the frequency with which they appear in the overall dataset on any given day within the observation window.

2. Summary of Analyses

We performed multiple variations of these self-controlled analyses to understand their strengths and limitations without naming a primary or secondary analysis. All analyses were treated as point exposures, meaning that any incident drug dispensing (i.e., to a new user of a medication) was considered an exposure regardless of the number of days supplied of the drug. All analyses were performed with a threshold for alerting set to 0.05 (1-sided). A summary is shown in **Table 2** below.

Table 2. Summary of Analyses

Analysis #	Outcome	Fixed/varying risk	Tree-based scan statistic	Drug tree incidence
1	Angioedema	Fixed (1-28)	Unconditional Bernoulli	3
2	Angioedema	Fixed (1-28)	Conditional Bernoulli	3
3	Angioedema	Variable	Unconditional Tree-Temporal	3
4	Angioedema	Variable	Conditional Tree-Temporal	3
5	ATR	Fixed (1-28)	Unconditional Bernoulli	3
6	ATR	Fixed (1-28)	Conditional Bernoulli	3
7	ATR	Fixed (1-63)	Unconditional Bernoulli	3
8	ATR	Fixed (1-63)	Conditional Bernoulli	3
9	ATR	Variable	Unconditional Tree-Temporal	3
10	ATR	Variable	Conditional Tree-Temporal	3
11	Angioedema	Fixed (1-28)	Unconditional Bernoulli	4
12	Angioedema	Fixed (1-28)	Conditional Bernoulli	4
13	Angioedema	Variable	Unconditional Tree-Temporal	4
14	Angioedema	Variable	Conditional Tree-Temporal	4
15	ATR	Fixed (1-28)	Unconditional Bernoulli	4
16	ATR	Fixed (1-28)	Conditional Bernoulli	4
17	ATR	Fixed (1-63)	Unconditional Bernoulli	4
18	ATR	Fixed (1-63)	Conditional Bernoulli	4
19	ATR	Variable	Unconditional Tree-Temporal	4
20	ATR	Variable	Conditional Tree-Temporal	4

ATR: Achilles tendon rupture

We did not consider scenarios in which the incident dispensing and health outcome of interest occurred on the same day (i.e., Day 0 analyses) because of an inability to distinguish the sequence of events within the same day in the data.[18]

For angioedema, we evaluated an overall post-exposure observation window of 1-63 days. We ran analyses on the following fixed risk windows: 1-28 days post-exposure. We considered the following variable risk windows: those that were at least two days long, at most 30 days long, and that were contained in the 1-63 day post-exposure observation window. All three conditions were required.

For Achilles tendon rupture, we evaluated an overall post-exposure observation window of 1-126 days. We evaluated the following fixed risk windows: 1-28 days post-exposure and 1-63 days post-exposure. We evaluated the following variable risk windows: those that were at least two days long, at most 60 days long, and that were contained in the 1-126 day post-exposure observation window. All three conditions were required.

For both outcomes, the comparison window consisted of the days within the observation window (i.e., follow-up period) that were not in the risk window being evaluated. Consequently, only post-exposure time was considered for the comparison window.

3. Multiple Hypothesis Testing and Pruning the Tree

One person may have contributed multiple exposure-outcome pairs in these data. However, at more coarsely-aggregated nodes on the tree (i.e., closer to the root), it is possible for one individual to have been dispensed two incident drugs (e.g., two blood pressure medications) that would have been represented twice in nodes with more aggregation, resulting in dependencies in the data. To prevent this occurrence, we did not conduct hypothesis tests at numerically lower levels than the pre-specified

incidence level. That is, for analyses where the incidence level was set to the 3rd level (i.e., the drug sub-class level), we did not conduct hypothesis testing at the 1st and 2nd levels.

Further, we did not conduct hypothesis testing at the 6th, 7th, and national drug code (NDC) levels (i.e., the "leaf" level or finest degree of granularity of the tree) because we did not believe that elevated frequencies were likely to be important at those levels. Further, eliminating hypothesis testing at those levels allowed us to maintain higher statistical power. **Table 3** indicates the frequency of "nodes" for potential hypothesis testing in the aggregated Medi-Span Therapeutic Drug Classification system. Additionally, there are 326,497 NDC level nodes.

Table 3. Frequency Distribution of Nodes in the Medi-Span Therapeutic Drug Classification

Level and name	Frequency	Percent	Cumulative frequency
1 - Drug group	94	0.26	94
2 - Drug class	663	1.86	757
3 - Drug subclass	1387	3.90	2144
4 - Drug base name	4132	11.61	6276
5 - Drug name/drug name extension	6144	17.27	12420
6 - Drug name and dosage form	8932	25.10	21352
7 - Drug name and strength	14231	39.99	35583

4. Alert Investigations

Alerts were first explored by reviewing the product label and the scientific literature to determine whether the statistical alerts were due to known adverse reactions. Further alert evaluation procedures using the Patient Episode Profile and Retrieval (PEPR) tool were conducted.[19]

III. ANGIOEDEMA RESULTS

A. ANGIOEDEMA SUMMARY DATA

Figure 1 displays the cohort attrition table for the angioedema analyses. 45,580 incident angioedema outcomes were ascertained when incidence criteria was set at the 3rd level and 46,360 incident angioedema outcomes were ascertained at the 4th level. More outcomes are expected to be ascertained when incidence criteria was set at the 4th level because the criteria are less stringent. The number of exposure-outcome pairs in the dataset with incidence set at the 3rd level was 110,785 and the 4th level was 117,498, making an average of 2.4-2.5 incident drug dispensings per incident angioedema event.

The number of pairs in the 28-day fixed risk window angioedema analyses was 61,066 and 64,730 for the 3rd and 4th incidence levels, respectively. Therefore, 55% of the total pairs occurred in the fixed risk window. The assumption of the null hypothesis in the unconditional analyses was that pairs are distributed uniformly across the observation window, meaning that 44% (28/63 days) were expected to occur in the 28-day fixed risk window. Therefore, this assumption of the unconditional analyses was not met.

Figure 1. Cohort Attrition Table for Angioedema Analytic Datasets

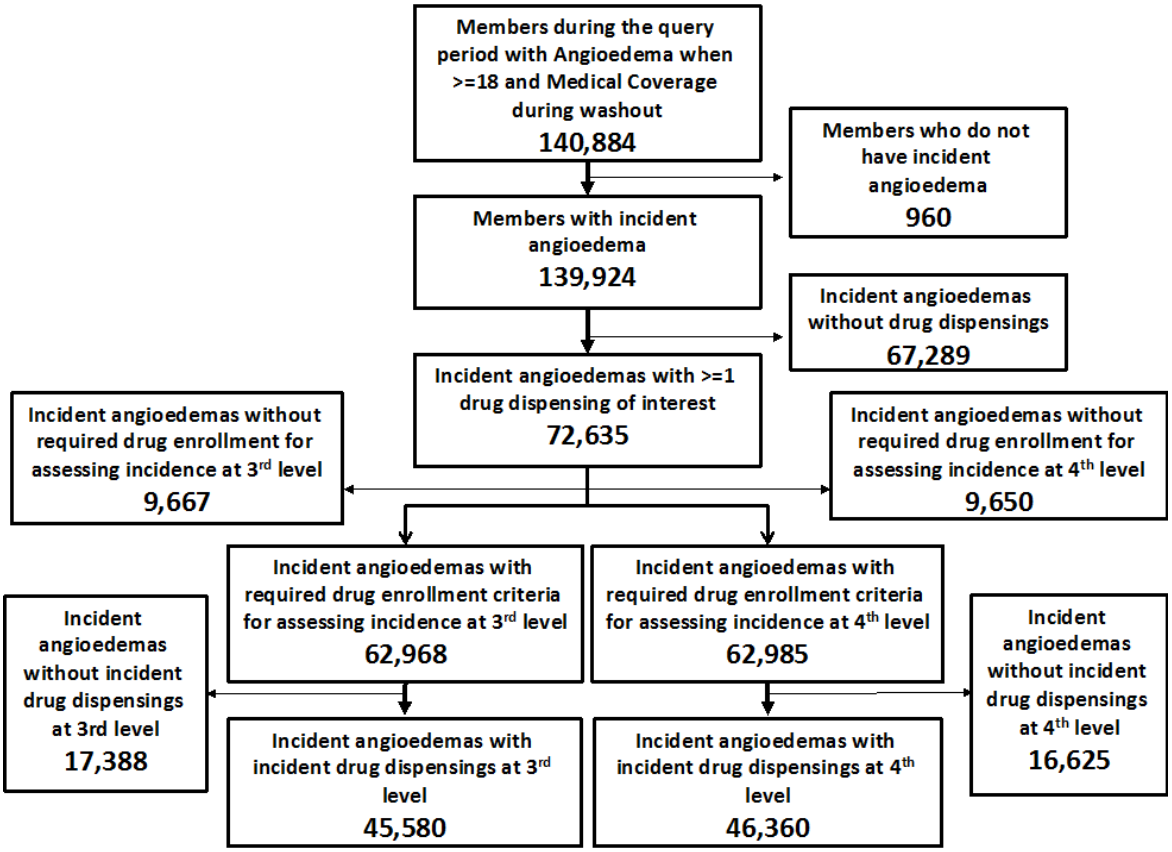
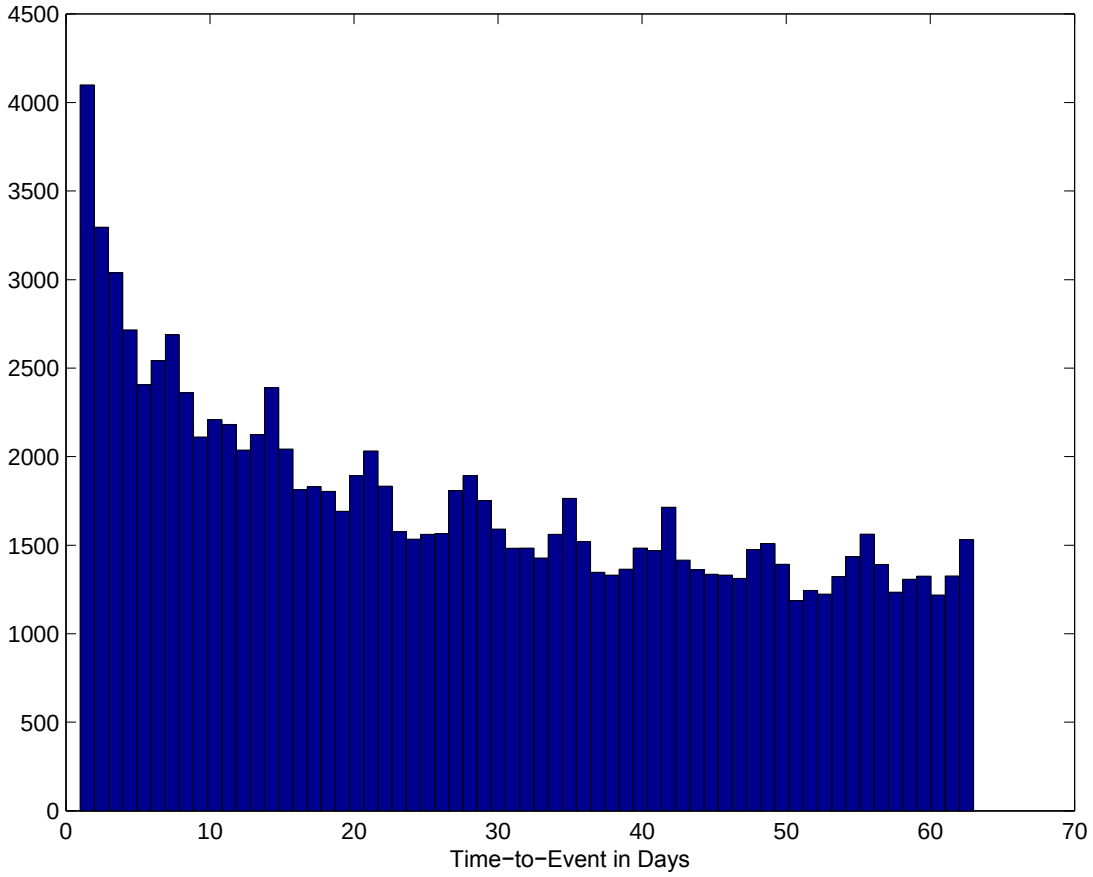


Figure 2 shows the time-to-event of all 110,785 exposure-outcome pairs in the analytic dataset when incidence was set at the 3rd level. The graph for the 4th level is very similar and not shown here. Visual inspection of the time-to-event data indicated that there was not equal probability of an outcome on any day in the observation window, an assumption of the unconditional forms of the tree-based scan statistic. Further, one can see a day-of-the-week pattern for angioedema events.

Figure 2. Time-to-Event in Days for 110,785 exposure-outcome pairs in the Angioedema Dataset with Incidence set to the 3rd Level.



B. ANGIOEDEMA PLANNED ANALYSES

As indicated in **Table 2**, we performed eight different angioedema analyses. The assumption of the unconditional tree-based scan statistic was not met and we do not show those results here for that reason.

Table 4 reports the results of the conditional tree-temporal scan statistic with a varying risk window applied to the angioedema dataset when incidence was set to the 3rd level (i.e., drug sub-class level). In **Table 4**, there were 41 nodes with exposures that met our pre-specified criteria for an "alert," (i.e., p-value at or below 0.05). However, many of the 4th and 5th level nodes have data that are identical because there was only a singular formulation of the active drug ingredient (e.g., prednisone). There were 28 unique alerting nodes and 15 nodes that were meaningfully different (i.e., that represent different alerts in different parts of the tree). Alerts in the same parts of the tree are grouped together with a "…" indication in all tables.

Table 5 reports the results of the conditional Bernoulli tree-based scan statistic with a fixed risk window applied to the angioedema dataset when incidence was set to the 3rd level. There were 22 unique alerting nodes and 8 that were in different parts of the tree.

Table 4 and **Table 5** are very similar as expected. Appendix C has additional result tables for the angioedema analyses where incidence was set to the 4th level. They are very similar.

Table 4. Data-Mining Results of the Conditional Tree-Temporal Scan Statistic with a Varying Risk Window Analysis in the Angioedema Dataset with Incidence set to the 3rd Level.

Node name	Tree Level	Node cases	Risk window	Cases in window	Expected Cases	Test statistic	P-value
Glucocorticosteroids	3	11127	1-8	3526	2324.3	274.4	<0.001
..Prednisone	4	7009	1-7	2145	1314.9	222.8	<0.001
..Methylprednisolone	4	3893	1-8	1141	813.2	59.1	<0.001
...Methylprednisolone	5	3891	1-8	1140	812.8	58.9	<0.001
Epinephrine	5	3640	1-15	1925	1256.3	154.9	<0.001
H-2 Antagonists	3	1917	1-6	606	313.1	107.7	<0.001
..Famotidine	4	781	1-6	296	127.6	80.8	<0.001
..Ranitidine	4	939	1-6	251	153.4	26.1	<0.001
Antianxiety Agents - Misc.	3	2268	1-4	472	269.2	62.5	<0.001
..Hydroxyzine	4	2174	1-4	467	258.0	68.3	<0.001
...Hydroxyzine HCl	5	1931	1-4	407	229.2	56.1	<0.001
...Hydroxyzine Pamoate	5	243	1-10	104	60.2	13.0	0.005
Sulfamethoxazole-Trimethoprim	5	1812	3-13	641	432.0	44.2	<0.001
Antihistamines - Ethanolamines	3	169	1-6	73	27.6	25.6	<0.001
..Diphenhydramine	4	160	1-6	70	26.1	25.1	<0.001
...Diphenhydramine HCl	5	159	1-6	70	26.0	25.4	<0.001
ACE Inhibitors	3	2649	2-8	594	455.3	19.3	<0.001
..Lisinopril	4	2169	2-8	502	372.8	20.2	<0.001
Minocycline HCl	5	199	12-19	63	28.3	15.8	<0.001
Bupropion HCl	5	390	19-30	122	73.0	13.7	0.003
Central Muscle Relaxants	3	1243	54-63	222	153.1	13.6	0.003
Levofloxacin	4	1116	1-3	161	105.1	12.8	0.007
...Levofloxacin	5	1115	1-3	161	105.0	12.8	0.007
Benzodiazepines	3	1491	41-63	531	425.5	12.2	0.015
HMG CoA Reductase Inhibitors	3	2008	42-54	414	323.0	11.8	0.022
..Simvastatin	4	802	42-54	188	129.0	11.8	0.022
Selective Serotonin Reuptake Inhibitors (SSRIs)	3	1044	32-59	469	372.2	11.7	0.025
Triazolam	4	35	51-52	8	0.8	11.5	0.030

Table 5. Data-Mining Results of the Conditional Bernoulli Scan Statistic with a Fixed Risk Window Analysis in the Angioedema Dataset with Incidence set to the 3rd Level.

Node name	Tree Level	Node cases	Risk window	Cases in window	Expected cases	Test statistic	P-value
Glucocorticosteroids	3	11127	1-28	7547	6133.3	414.3	<0.001
..Prednisone	4	7009	1-28	4861	3863.4	316.5	<0.001
..Methylprednisolone	4	3893	1-28	2555	2145.9	92.1	<0.001
...Methylprednisolone	5	3891	1-28	2553	2144.8	91.7	<0.001
Epinephrine	4	3640	1-28	2656	2006.4	254.1	<0.001
H-2 Antagonists	3	1917	1-28	1378	1056.7	115.8	<0.001
..Famotidine	4	781	1-28	601	430.5	81.1	<0.001
..Ranitidine	4	939	1-28	634	517.6	30.3	<0.001
..Cimetidine	4	192	1-28	140	105.8	12.9	<0.001
...Cimetidine	5	191	1-28	139	105.3	12.6	<0.001
Antianxiety Agents - Misc.	3	2268	1-28	1538	1250.2	77.6	<0.001
..Hydroxyzine	4	2174	1-28	1491	1198.3	83.8	<0.001
...Hydroxyzine HCl	5	1931	1-28	1316	1064.4	69.5	<0.001
...Hydroxyzine Pamoate	5	243	1-28	175	133.9	14.7	<0.001
Sulfamethoxazole-Trimethoprim	5	1812	1-28	1169	998.8	33.5	<0.001
Antihistamines – Ethanolamines	3	169	1-28	134	93.2	21.7	<0.001
..Diphenhydramine	4	160	1-28	129	88.2	23.0	<0.001
...Diphenhydramine HCl	5	159	1-28	128	87.6	22.7	<0.001
ACE Inhibitors	3	2649	1-28	1621	1460.2	20.4	<0.001
..Lisinopril	4	2169	1-28	1345	1195.6	21.5	<0.001
Antihistamines - Non-Sedating	3	1503	1-28	906	828.5	8.3	0.03
..Levocetirizine	4	385	1-28	251	212.2	8.1	0.03

The alerting nodes were dominated by exposures that are treatments for angioedema and related diagnoses, suggesting that patients were being treated for some sort of allergic reaction or hypersensitivity that was later recognized to be, and coded as, angioedema. A delayed recording of angioedema in the data would result in misclassification of the onset of the outcome. Thus, we hypothesized we were capturing individuals at a prodromal stage and the angioedema was not likely to be caused by the exposures in **Table 4** that are treatments for allergic conditions, but rather the outcome of some other exposure or pre-existing allergic condition. We explored this hypothesis further as described in Section 0. In a conditional analysis (i.e., one that uses a conditional variant of the tree-based scan statistic), the presence of a temporal pattern created by misclassification of the onset can have adverse effects by attenuating other potential alerts and/or inappropriately amplifying other noise.

C. ANGIOEDEMA ALERT INVESTIGATION

To confirm our hypothesis of misclassification of the onset of angioedema, we randomly extracted fourteen patient claims profiles for patients alerting in the identified risk window of the glucocorticosteroids node (days 1-8), and fifteen patient claims profiles for patients alerting in the identified hydroxyzine node (days 1-4). We extracted all claims in the 30 days preceding their exposure and the 30 days following their exposure.

Among the identified patient profiles in the glucocorticosteroids node, 10 of the 14 patient profiles retrieved had evidence of symptoms suggestive of an allergy or hypersensitivity reaction on the index date or in the days preceding it (e.g., ICD-9-CM code 995.3, allergy unspecified not elsewhere classified). Six of the 14 had codes for ICD-9-CM 782.1, rash and other non-specific skin eruption. Of the 4 patients without evidence of prior allergy, one had systemic lupus erythematosus and another was dispensed glucocorticosteroids following the angioedema episode, implying that the glucocorticosteroids were not a suspected agent. The remaining two patient profiles did not contain enough information to allow further interpretation.

Among the hydroxyzine nodes, 13/15 had pre-existing allergy codes in the week prior to the exposure, and 15/15 were dispensed or administered steroids on the same day or in the week preceding the hydroxyzine exposure. The most commonly occurring predecessor codes were ICD-9-CM 708.9, 708.0, and 995.3, which are unspecified urticaria, allergic urticaria, and allergy unspecified not elsewhere classified, respectively.

We concluded that misclassification of the onset of angioedema was probable as evidenced by the numerous precursor allergy diagnoses codes.

D. ANGIOEDEMA SENSITIVITY ANALYSES

We pruned all anti-histamines and other known angioedema treatments (e.g., hydroxyzine, glucocorticosteroids) from the Medi-Span Therapeutic Classification System and re-performed the analysis with the new dataset in order to remove unwanted effects of conditioning. As a result of pruning, we lost 21,249 exposure-outcome pairs from the angioedema treatments, which was 19.2% of the original dataset.

Table 6 reports the results of the conditional tree-temporal scan statistic with a varying risk window applied to the pruned angioedema dataset when incidence was set to the 3rd level (i.e., drug sub-class level). There were 13 unique alerting nodes located in 8 parts of the tree. **Table 7** reports the results of the conditional Bernoulli tree-based scan statistic with a fixed risk window analysis applied to the pruned angioedema dataset when incidence was set to the 3rd level. There were 11 unique alerting nodes located in 7 parts of the tree.

Table 6. Data-Mining Results of the Conditional Tree-Temporal Scan Statistic with a Varying Risk Window Analysis in the Pruned Angioedema Dataset with Incidence set to the 3rd Level.

Node name	Tree Level	Node cases	Risk window	Cases in window	Expected cases	Test statistic	P-value
Sulfamethoxazole-Trimethoprim	5	1812	3-12	598	373.1	57.5	<0.001
ACE Inhibitors	3	2649	2-8	594	407.1	37.7	<0.001
..Lisinopril	4	2169	1-8	590	402.7	38.2	<0.001
Fluoroquinolones	3	3159	1-6	588	448.0	20.0	<0.001
..Levofloxacin	4	1116	1-6	249	158.4	22.1	<0.001
...Levofloxacin	5	1115	1-6	249	158.2	22.2	<0.001
Minocycline	4	199	15-19	44	16.0	16.6	<0.001
Clindamycin	4	770	1-7	189	127.3	13.0	0.001
...Clindamycin HCl	5	769	1-7	188	127.1	12.7	0.002
Bupropion	4	390	18-31	136	85.6	12.6	0.002
ACE Inhibitors & Thiazide/Thiazide-Like	4	814	2-9	201	140.0	11.7	0.006
..Lisinopril & Hydrochlorothiazide	5	748	2-9	188	128.8	11.9	0.005
Triazolam	4	35	51-52	8	0.9	10.0	0.046

Table 7. Data-Mining Results of the Conditional Bernoulli Scan Statistic with a Fixed Risk Window Analysis in the Pruned Angioedema Dataset with Incidence set to the 3rd Level.

Node name	Tree level	Node cases	Risk window	Cases in window	Expected cases	Test statistic	P-value
Sulfamethoxazole-Trimethoprim	5	1812	1-28	1169	942.8	58.8	<0.001
ACE Inhibitors	3	2649	1-28	1621	1378.3	46.4	<0.001
..Lisinopril	4	2169	1-28	1345	1128.5	44.9	<0.001
Thienopyridine Derivatives	3	432	1-28	277	224.8	12.9	<0.001
..Clopidogrel	4	414	1-28	267	215.4	13.2	<0.001
ACE Inhibitors & Thiazide/Thiazide-Like	4	814	1-28	488	423.5	10.4	0.003
..Lisinopril & Hydrochlorothiazide	5	748	1-28	449	389.2	9.7	0.005
Doxepin	4	253	1-28	167	131.6	10.1	0.003
Corticosteroids - Topical	3	2897	1-28	1620	1507.3	9.1	0.011
Clindamycin	4	770	1-28	459	400.6	9.0	0.011
..Clindamycin HCl	5	769	1-28	458	400.1	8.9	0.013

Concerning the alerts in **Table 6** and **Table 7**, there was evidence in the literature that some of these nodes were true positive associations (e.g., lisinopril[20] and its combinations, bupropion hydrochloride[21], flouroquinolones[22]) and these drugs contained angioedema in either the *Contraindications* or *Warnings and Precautions* section of the label. The labels for triazolam and clopidogrel also included angioedema in the *Warnings and Precautions* section. We note that the triazolam alert had few cases and an unusually short identified risk window that occurred multiple weeks after incident dispensing. There was also evidence that the other antibiotics were likely true positive associations[23,24]; angioedema was listed in the *Adverse Reactions* portions of these labels.

In **Table 7**, there were additional allergy treatments (e.g., doxepin hydrochloride, topical corticosteroids) that were not pruned out, but that we believed to be further evidence of misclassification of the onset.

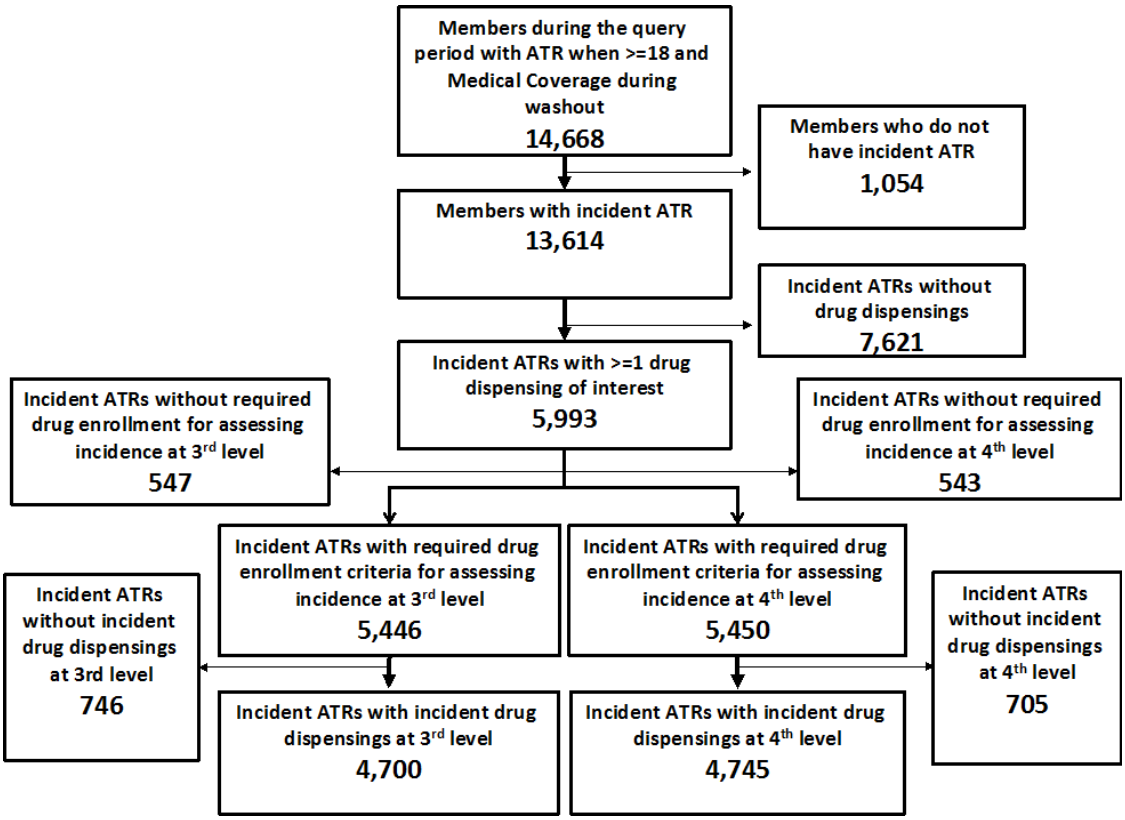
In summary, the angioedema alerts seen in the pruned datasets were either true positive associations or likely positive associations, and no false positive alerts were detected.

IV. ACHILLES TENDON RUPTURE RESULTS

A. ACHILLES TENDON RUPTURE SUMMARY DATA

Figure 3 displays the cohort attrition table for the Achilles tendon rupture analyses. More Achilles tendon rupture outcomes were ascertained when incidence criteria were set at the 4th level, which was less stringent. Similarly, the number of exposure-outcome pairs in the dataset with incidence set at the 3rd level was 12,576 and the 4th level was 13,186, making an average of 2.7-2.8 incident drug dispensings per incident Achilles tendon rupture event.

Figure 3. Cohort Attrition Table for Achilles Tendon Rupture Analytic Datasets



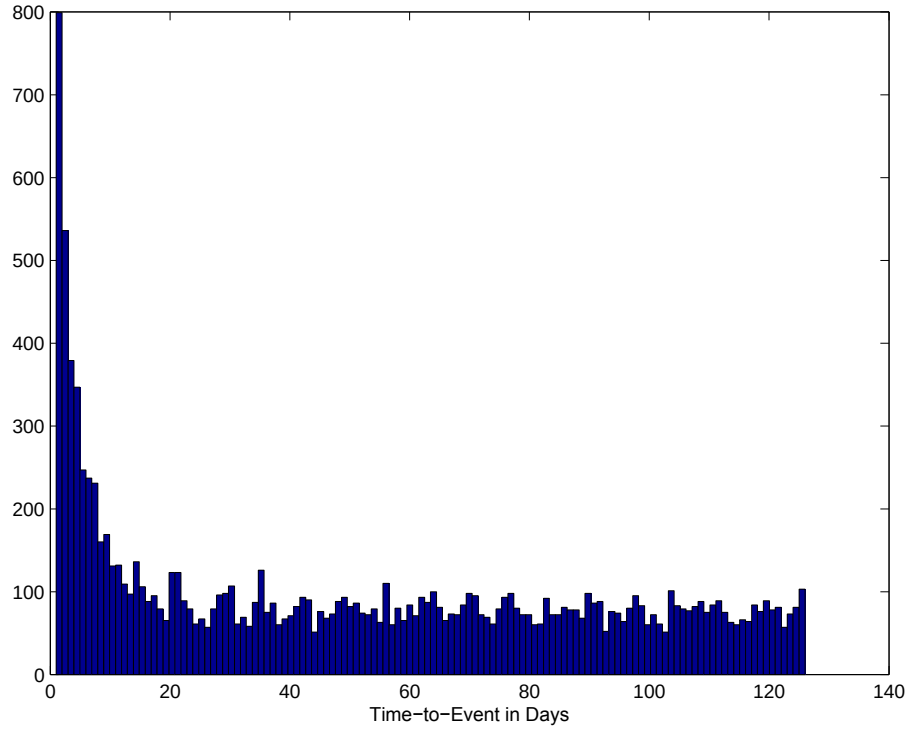
The numbers of exposure-outcome pairs in the 28-day risk window for the Achilles tendon rupture fixed risk window analyses were 4,917 and 5,080 for the 3rd and 4th incidence levels, respectively. Therefore, 39% of the total pairs observed during the 126-day observation window occurred in the fixed risk window. The assumption of the null hypothesis in the unconditional analyses was that pairs are distributed uniformly across the observation window, meaning that 22% (28/126 days) were expected to occur in the risk window.

The numbers of exposure-outcome pairs in the 63-day risk window for the Achilles tendon rupture fixed risk window analyses were 7,702 and 8,045 for the 3rd and 4th incidence levels, respectively. Therefore, 61% of the total pairs observed during the 126-day observation window occurred in the fixed risk window. The assumption of the null hypothesis in the unconditional analyses was that pairs are distributed uniformly across the observation window, meaning that 50% (63/126 days) were expected to occur in the risk window.

Therefore, this assumption of the unconditional analyses was not met in either the 28-day or 63-day fixed risk window analysis.

Figure 4 shows the time-to-event of all 12,576 exposure-outcome pairs in the analytic dataset when incidence was set at the 3rd level. The graph for the 4th level is very similar and not shown here. Visual inspection of the time-to-event data indicated that there was not equal probability of an outcome on any day in the observation window, an assumption of the unconditional tree-based scan statistic. There was much more extreme imbalance with respect to counts immediately following incident dispensings (i.e., more skewed to the origin).

Figure 4. Time-to-Event Data for 12,576 exposure-outcome pairs in the Achilles tendon rupture dataset with Incidence set to the 3rd level.



B. ACHILLES TENDON RUPTURE PLANNED ANALYSES

As indicated in **Table 2**, we performed twelve different Achilles tendon rupture analyses. The assumption of the unconditional tree-based scan statistic was not met and we do not show those results here for that reason.

Table 8 reports the results of the conditional tree-temporal scan statistic with a varying risk window applied to the Achilles tendon rupture dataset when incidence was set to the 3rd level (i.e., drug sub-class level). In **Table 8**, there were 24 nodes with exposures that met our pre-specified criteria for an "alert," (i.e., p-value at or below 0.05). However, many of the 4th and 5th level nodes have data that were identical because there was only a singular formulation of the active drug ingredient (e.g., promethazine hydrochloride). There were 15 unique nodes with "alerts" located in 8 different parts of the tree and those are the ones we show in **Table 8**.

Table 8. Data-Mining Results of the Conditional Tree-Temporal Scan Statistic with a Varying Risk Window Analysis in the Achilles Tendon Rupture Dataset with Incidence set to the 3rd Level.

Node name	Tree level	Node cases	Risk window	Cases in window	Expected cases	Test statistic	P-value
Hydrocodone Combination - Two Ingredient	4	1389	1-9	851	342.9	276.1	<0.001
..Hydrocodone-Acetaminophen	5	1366	1-9	844	337.3	278.1	<0.001
Opioid Combination - Two Ingredient	4	746	1-7	508	164.7	233.7	<0.001
..Oxycodone w/ Acetaminophen	5	745	1-7	507	164.5	233.1	<0.001
Nonsteroidal Anti-inflammatory Agents (NSAIDs)	3	1111	1-20	513	376.9	22.8	<0.001
..Ibuprofen	4	401	1-14	251	118.3	56.8	<0.001
Promethazine	4	185	1-8	113	43.2	39.1	<0.001
Cephalosporins - 1st Generation	3	344	1-7	159	75.9	34.7	<0.001
..Cephalexin	4	309	1-7	135	68.2	25.6	<0.001
..Cefadroxil	4	35	1-8	25	8.2	11.1	0.047
Opioid Agonists	3	324	1-4	116	53.1	27.9	<0.001
..Oxycodone	4	92	1-7	63	20.3	28.7	<0.001
Low Molecular Weight Heparins	3	67	1-15	55	20.3	20.1	<0.001
..Enoxaparin	4	66	1-10	49	17.0	19.9	<0.001
Glucocorticosteroids	3	472	50-104	231	160.4	13.9	0.002

Table 9 reports the results of the conditional Bernoulli tree-based scan statistic with a fixed risk window analysis applied to the Achilles tendon rupture dataset when incidence was set to the 3rd level. There were 21 unique alerting nodes located in 9 parts of the tree.

Table 9. Data-Mining Results of the Conditional Bernoulli Scan Statistic with a Fixed Risk Window Analysis in the Achilles Tendon Rupture Dataset with Incidence set to the 3rd Level.

Node name	Tree level	Node cases	Risk window	Cases in window	Expected cases	Test statistic	P-value
Hydrocodone Combination - Two Ingredient	4	1389	1-28	1024	543.1	386.7	<0.001
..Hydrocodone-Acetaminophen	5	1366	1-28	1010	534.1	384.4	<0.001
Opioid Combination - Two Ingredient	4	746	1-28	624	291.7	333.9	<0.001
..Oxycodone w/ Acetaminophen	5	745	1-28	623	291.3	333.1	<0.001
Nonsteroidal Anti-inflammatory Agents (NSAIDs)	3	1111	1-28	569	434.4	36.7	<0.001
..Ibuprofen	4	401	1-28	287	156.8	89.6	<0.001
..Naproxen	4	288	1-28	149	112.6	9.6	0.003
...Naproxen	5	255	1-28	132	99.7	8.5	0.007
..Ketorolac	4	33	1-28	25	12.9	9.2	0.005
Opioid Agonists	3	324	1-28	208	126.7	42.7	<0.001
..Oxycodone	4	92	1-28	78	36.0	41.3	<0.001
..Meperidine	4	19	1-28	16	7.4	8.2	0.013
Promethazine	4	185	1-28	133	72.3	41.4	<0.001
Cephalosporins - 1st Generation	3	344	1-28	213	134.5	37.5	<0.001
..Cephalexin	4	309	1-28	183	120.8	26.1	<0.001
..Cefadroxil	4	35	1-28	30	13.7	16.3	<0.001
Low Molecular Weight Heparins	3	67	1-28	56	26.2	28.3	<0.001
..Enoxaparin	4	66	1-28	55	25.8	27.5	<0.001
Ondansetron	4	75	1-28	50	29.3	11.7	<0.001
..Ondansetron HCl	5	37	1-28	30	14.5	13.7	<0.001
Hydroxyzine Pamoate	5	26	1-28	21	10.2	9.5	0.003

Table 10 reports the results of the conditional Bernoulli tree-based scan statistic with a fixed risk window analysis applied to the Achilles tendon rupture dataset when incidence was set to the 3rd level. There were 16 unique alerting nodes located in 8 distinct parts of the tree.

Table 10. Data-Mining Results of the Conditional Bernoulli Scan Statistic with a Fixed Risk Window Analysis in the Achilles Tendon Rupture Dataset with Incidence set to the 3rd Level.

Node name	Tree level	Node cases	Risk window	Cases in window	Expected cases	Test statistic	P-value
Hydrocodone Combination - Two Ingredient	4	1389	1-63	1172	850.7	197.7	<0.001
..Hydrocodone-Acetaminophen	5	1366	1-63	1153	836.6	194.6	<0.001
Opioid Combination - Two Ingredient	4	746	1-63	670	456.9	163.0	<0.001
..Oxycodone w/ Acetaminophen	5	745	1-63	669	456.3	162.6	<0.001
Nonsteroidal Anti-inflammatory Agents (NSAIDs)	3	1111	1-63	774	680.4	18.8	<0.001
..Ibuprofen	4	401	1-63	333	245.6	46.5	<0.001
Opioid Agonists	3	324	1-63	259	198.4	26.8	<0.001
..Oxycodone	4	92	1-63	87	56.3	28.1	<0.001
..Meperidine	4	19	1-63	19	11.6	9.3	0.003
Cephalosporins - 1st Generation	3	344	1-63	265	210.7	20.0	<0.001
..Cephalexin	4	309	1-63	233	189.2	14.2	<0.001
..Cefadroxil	4	35	1-63	32	21.4	8.3	0.011
Promethazine	4	185	1-63	150	113.3	17.2	<0.001
Low Molecular Weight Heparins	3	67	1-63	61	41.0	15.5	<0.001
..Enoxaparin	4	66	1-63	60	40.4	15.1	<0.001
Ondansetron HCl	5	37	1-63	33	22.7	7.3	0.031

The alerting nodes for the Achilles tendon rupture datasets were mostly therapeutics used to treat pain from sports injuries that often precede the development or diagnosis of an Achilles tendon rupture, or were therapeutics that are administered peri-operatively (e.g., cephalexin, heparin, promethazine hydrochloride) to patients undergoing surgical repair of an Achilles tendon rupture. Therefore, we hypothesized misclassification of the onset of Achilles tendon rupture and the likely presence of earlier injury (e.g., sports injury).

We had also expected to see fluoroquinolones "alert" and they did not.

Using the power evaluation feature of TreeScan™ that is available for an unconditional Bernoulli tree-based scan statistic, we evaluated what power we had available to see a fluoroquinolone signal. There were 138 incident levofloxacin dispensings, 130 incident ciprofloxacin dispensings, and 286 incident fluoroquinolones dispensings. **Figure 5** shows the time-to-event data for all incident fluoroquinolone dispensings.

Table 11 indicates what the probability of alerting was given those data and given a true relative risk as indicated. Simulations were generated with 100,000 replications under the null hypothesis and 10,000 replications under the alternative hypothesis. The null hypothesis was that there was no increased occurrence of Achilles tendon rupture in the 28 days following fluoroquinolone dispensing. The

alternative hypothesis was that there was an increased occurrence of Achilles tendon rupture in the 28 days following fluoroquinolone dispensing and the strength of the hypothesized effect size was as indicated in **Table 11**. A recent literature review of fluoroquinolones and Achilles tendon rupture reported odd ratios that were highest in the first 30 days following fluoroquinolone exposure but the range was 1.5 to 4.[25]

Figure 5. Time-to-Event in Days for Fluoroquinolones and Achilles tendon rupture with Incidence set to the 3rd level.

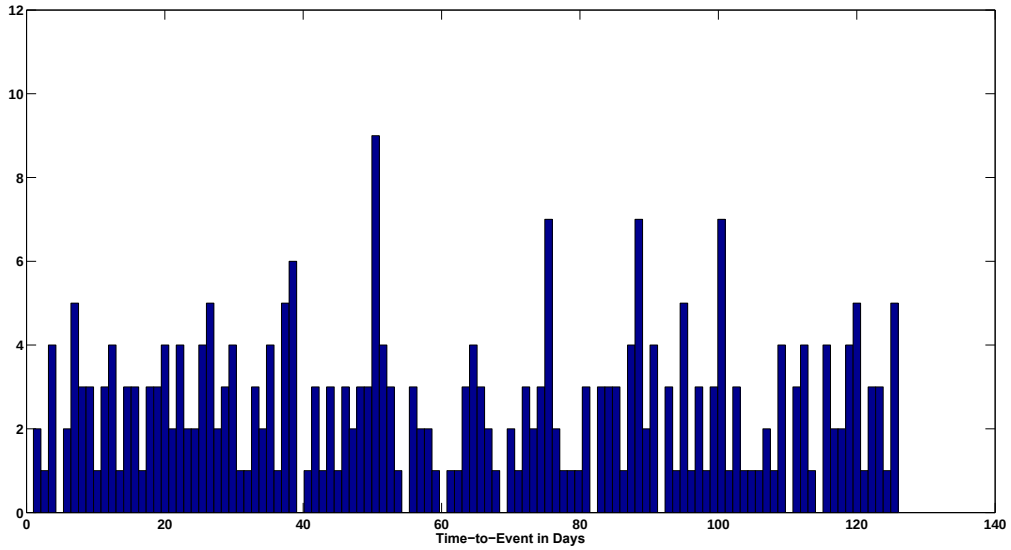


Table 11. Probability of Signaling using the Unconditional Bernoulli Scan Statistic with a Fixed Risk Window Analysis in the Achilles Tendon Rupture Dataset with Incidence set to the 3rd Level.

Alternative hypothesis (relative risk)	All fluoroquinolone		Levofloxacin		Ciprofloxacin	
	28 day	63 day	28 day	63 day	28 day	63 day
1	0.046	0.036	0.046	0.036	0.046	0.036
1.5	0.283	0.401	0.095	0.127	0.101	0.103
2	0.946	0.981	0.505	0.639	0.507	0.540
3	1.000	1.000	0.989	0.994	0.987	0.985
5	1.000	1.000	1.000	1.000	1.000	1.000
7	1.000	1.000	1.000	1.000	1.000	1.000
10	1.000	1.000	1.000	1.000	1.000	1.000

Overall, there was a low probability of signaling for all fluoroquinolones combined using the unconditional Bernoulli tree-based scan statistic with a fixed risk window analysis if the relative risk were less than twofold, which was consistent with some of the Achilles tendon rupture studies cited in the review paper.[25]

C. ACHILLES TENDON RUPTURE ALERT INVESTIGATION

To confirm our hypothesis of misclassification of the onset of Achilles tendon rupture, we randomly extracted sixteen patient claims profiles for patients alerting in the identified risk window of the hydrocodone combinations node (days 1-9). We extracted all claims in the 30 days preceding their exposure and the 45 days following their exposure. The 45 days following exposure ensured that we would see the ascertainment scheme for the Achilles tendon rupture, which required two codes within thirty days.

Among the identified patient profiles in the hydrocodone combinations node, 6 of the 16 patient profiles had evidence of trauma or injury on the index date or in the days preceding it. 7 of the 16 had antecedent Achilles tendon problems, most notably ICD-9-CM 726.71, Achilles bursitis or tendonitis. The remaining three patient profiles did not contain enough information to allow further interpretation.

Additionally, we randomly extracted 48 patient profiles from the non-alerting fluoroquinolone node to discover whether their clinical features were consistent with fluroroquinolone associated-Achilles tendon rupture. We extracted all claims in the 30 days preceding their exposure and the 126 days following their exposure, which comprised the entire observation window.

12/48 had antecedent injury or tendonitis prior to the index date of exposure. 18/48 had concomitant steroid use. 28/48 were 55+. 12 had diagnoses consistent with a urinary tract infection, 18 had diagnoses consistent with upper respiratory infections including bronchitis and sinusitis. In general, the data were consistent with risk factors associated with fluoroquinolone associated-Achilles tendon rupture (e.g., concomitant steroid use, age), but we did not use a case definition to sort cases into definite, probable, or possible fluoroquinolone associated Achilles tendon rupture.

D. ACHILLES TENDON RUPTURE SENSITIVITY ANALYSES

Similar to the strategy we employed for angioedema, we pruned all pain medications and surgery prophylaxis medications (e.g., antibiotics) from the tree, leaving us with 7,805 exposure-outcome pairs (38% loss of data). When we analyzed this pruned dataset, there were no alerting nodes. Because we were concerned about possible misclassification of the onset date based on the claims profiles we retrieved, we removed time-to-event data for days 1-14 and re-performed the conditional tree-temporal scan with a varying risk window analysis on only events that happened from days 15-126. When we analyzed this modified dataset, there were no alerting nodes.

V. DISCUSSION

We used a self-controlled case-crossover design to develop analytic datasets to detect elevated frequencies of incident drug dispensings preceding either incident angioedema or incident Achilles tendon rupture. We analyzed these datasets using various forms of the tree-based scan statistic. This method controlled for multiple hypothesis testing while evaluating data organized into hierarchical trees.

For angioedema events, we detected elevated frequencies of many therapeutics used to treat allergic reactions and suspected that we were capturing an interim point in an allergic process. Upon removing these therapeutics from the hierarchical drug tree and then re-performing the analysis, we detected known true positive associations (e.g., lisinopril) or suspected true positive associations (e.g., antibiotics). There were few total alerts considering the large number of incident drug dispensings evaluated.

For Achilles tendon rupture events, we detected elevated frequencies of many therapeutics used to treat injuries or other surgery prophylaxis agents, leading us to suspect misclassification of the onset date for the rupture event or time-varying confounding due to a pre-existing injury.

We did not detect an elevation of fluoroquinolones and were underpowered to do so if the effect size was less than a twofold increased relative risk.

Outcome-oriented TreeScan™ analyses had the unintended benefit of pinpointing potential areas of concern in health outcome of interest algorithm validation. For example, our analyses, which used validated algorithms, suggested residual misclassification of the onset of disease. It might be more appropriate when identifying drug-induced angioedema to include additional exclusions for prodromal allergy symptoms. Likewise, when identifying drug-induced Achilles tendon rupture, it might be appropriate to exclude those with antecedent injuries.

There were limitations of this evaluation, which are either inherent to secondary-use observational data, the nature of data-mining, or deliberate choices to simplify a primarily exploratory analysis.

First, 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 safety surveillance.[26]

Second, we chose to consider observation windows in the first several months following exposure, and therefore we could not detect drug-associated outcomes that occurred several years after exposure in this analysis. Future studies could be structured to look for those types of events, but would likely be of a different epidemiologic design.

Third, we considered an incident dispensing to be any qualifying exposure regardless of the length of the dispensing as a simplifying measure in this exploratory study because we were unable to tailor the analysis to track the many different exposure patterns that might be present among the thousands of exposures being assessed. Thus, a 1-day incident exposure was dealt with in exactly the same way as an incident exposure that might continue on a daily basis for a long time, e.g. statins.

For these outcomes, there were a few well-known exposure-outcome pairs that could be described as known positive controls, e.g. angiotensin-converting enzyme inhibitor-induced angioedema, fluoroquinolone antibiotic-induced Achilles tendon rupture. Because of these established relationships, prescribers of these medical products may be more likely to follow their patients in anticipation of these adverse effects, and thus provide a differential opportunity to detect the outcome of interest (e.g., providers may look for signs of ankle pain or tendinopathy). Additionally, patients at high risk of Achilles tendon rupture may be channeled away from these medications because of the associated adverse events.

The purpose of our outcome-based TreeScan™ analysis was exploratory signal detection. That is, we sought to investigate a method that could point to *potential* problems that would require further attention, focusing on outcomes that were particularly troubling because they have a drug-induced component. Once an alert is generated, an understanding of the clinical context between the alerting drug and the outcome of interest may be sufficient to dismiss the alert as unlikely to represent a temporally related cause of the outcome of interest. However, it is also possible that TreeScan™ may suggest a previously undetected exposure-outcome pair of interest. The analytic results of TreeScan™ should not by themselves be viewed as evidence of a causal relationship between an exposure and an outcome.

VI. CONCLUSIONS

Signal identification has traditionally been strongly driven by spontaneous reports, which lack population data to provide context. While these reports are the backbone of postmarket surveillance, there is more that can be done to characterize their public health significance. Outcome-based data-mining can generate a complementary stream of new information on these events without being overwhelmed by false positive information. Particular outcomes of interest are events that are serious and often drug-related (e.g., liver failure, bone marrow failure, Stevens-Johnson syndrome). Finally, an unexpected finding was that outcome-based data mining could be used to "pressure-test" health outcome algorithms, suggesting ways to increase positive predictive value and reduce false positive associations.

VII. REFERENCES

1. Brown, J. S. *et al.* Drug Adverse Event Detection in Health Plan Data Using the Gamma Poisson Shrinker and Comparison to the Tree-based Scan Statistic. *Pharmaceutics* **5,** 179–200 (2013).
2. Kulldorff, M. *et al.* Drug safety data mining with a tree-based scan statistic. *Pharmacoepidemiol. Drug Saf.* **22,** 517–523 (2013).
3. Kulldorff, M., Fang, Z. & Walsh, S. J. A tree-based scan statistic for database disease surveillance. *Biometrics* **59,** 323–331 (2003).
4. Nelson, J. C. *et al.* Integrating database knowledge and epidemiological design to improve the implementation of data mining methods that evaluate vaccine safety in large healthcare databases. *Stat. Anal. Data Min.* **7,** 337–351 (2014).
5. Yih, W. K. et al. Pilot of Self-Controlled Tree-Temporal Scan Analysis for Gardasil Vaccine. (U.S. Food and Drug Administration, 2016).
6. Food and Drug Administration & Department of Health and Human Services. Safety Reporting Requirements for Human and Drug Biological Prodcuts (Notice of Proposed Rulemaking). *Fed. Regist.* **68,** 12406–12497 (2003).
7. Coloma, P. M. *et al.* Drug-induced acute myocardial infarction: identifying 'prime suspects' from electronic healthcare records-based surveillance system. *PloS One* **8,** e72148 (2013).
8. Ferrajolo, C. *et al.* Signal detection of potentially drug-induced acute liver injury in children using a multi-country healthcare database network. *Drug Saf. Int. J. Med. Toxicol. Drug Exp.* **37,** 99–108 (2014).
9. Ryan, P. B., Madigan, D., Stang, P. E., Schuemie, M. J. & Hripcsak, G. Medication-wide association studies. *CPT Pharmacomet. Syst. Pharmacol.* **2,** e76 (2013).
10. Patel, C. J., Ji, J., Sundquist, J., Ioannidis, J. P. A. & Sundquist, K. Systematic assessment of pharmaceutical prescriptions in association with cancer risk: a method to conduct a population-wide medication-wide longitudinal study. *Sci. Rep.* **6,** 31308 (2016).
11. The Council for International Organizations of Medical Sciences (CIOMS) Working Group III. *Guidelines for preparing core clinical safety information on drugs.* (World Health Organization (WHO), 1995).
12. Maclure, M. *et al.* When should case-only designs be used for safety monitoring of medical products? *Pharmacoepidemiol. Drug Saf.* **21 Suppl 1,** 50–61 (2012).
13. Maclure, M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am. J. Epidemiol.* **133,** 144–153 (1991).
14. Toh, S. *et al.* Comparative risk for angioedema associated with the use of drugs that target the renin-angiotensin-aldosterone system. *Arch. Intern. Med.* **172,** 1582–1589 (2012).
15. Kostis, J. B. *et al.* Incidence and characteristics of angioedema associated with enalapril. *Arch. Intern. Med.* **165,** 1637–1642 (2005).
16. Seeger, J. D. *et al.* Achilles tendon rupture and its association with fluoroquinolone antibiotics and other potential risk factors in a managed care population. *Pharmacoepidemiol. Drug Saf.* **15,** 784–792 (2006).
17. Khaliq, Y. & Zhanel, G. G. Fluoroquinolone-Associated Tendinopathy: A Critical Review of the Literature. *Clin. Infect. Dis.* **36,** 1404–1410 (2003).
18. Brown, J. S. *et al.* Active influenza vaccine safety surveillance: potential within a healthcare claims environment. *Med. Care* **47,** 1251–1257 (2009).
19. Cole, D. V. *et al.* Infrastructure for Evaluation of Statistical Alerts Arising from Vaccine Safety Data Mining Activities in Mini-Sentinel. *Sentinel Initiative* (2016). Available at:

https://www.sentinelinitiative.org/sites/default/files/Methods/Mini-Sentinel_PRISM_Data-Mining-Infrastructure_Report_0.pdf. (Accessed: 5th January 2017)
20. Baram, M., Kommuri, A., Sellers, S. A. & Cohn, J. R. ACE inhibitor-induced angioedema. *J. Allergy Clin. Immunol. Pract.* **1,** 442–445 (2013).
21. Beyens, M.-N., Guy, C., Mounier, G., Laporte, S. & Ollagnier, M. Serious adverse reactions of bupropion for smoking cessation: analysis of the French Pharmacovigilance Database from 2001 to 2004. *Drug Saf.* **31,** 1017–1026 (2008).
22. Jones, S. C., Budnitz, D. S., Sorbello, A. & Mehta, H. US-based emergency department visits for fluoroquinolone-associated hypersensitivity reactions. *Pharmacoepidemiol. Drug Saf.* **22,** 1099–1106 (2013).
23. Bertazzoni, G. *et al.* Drug-induced angioedema: experience of Italian emergency departments. *Intern. Emerg. Med.* **9,** 455–462 (2014).
24. Lerch, M. Drug-induced angioedema. *Chem. Immunol. Allergy* **97,** 98–105 (2012).
25. Stephenson, A. L., Wu, W., Cortes, D. & Rochon, P. A. Tendon Injury and Fluoroquinolone Use: A Systematic Review. *Drug Saf.* **36,** 709–721 (2013).
26. Schneeweiss, S. & Avorn, J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J. Clin. Epidemiol.* **58,** 323–337 (2005).
27. Dwass, M. Modified Randomization Tests for Nonparametric Hypotheses. *Ann. Math. Stat.* **28,** 181–187 (1957).

VIII. ACKNOWLEDGEMENTS

We acknowledge the contributions of our other workgroup members including Carolyn Balsbaugh and Andrew Petrone.

IX. APPENDIX A - BRIEF DESCRIPTION OF THE METHODS

A. UNCONDITIONAL BERNOULLI SCAN STATISTIC WITH FIXED RISK WINDOW

All exposures are first classified into a hierarchical tree structure described in Section II.C. For each leaf *i* of the tree (i.e., finest granularity) which represents a unique exposure of interest, we note the observed number c_i of outcomes in the risk window and the observed number n_i of outcomes in the comparison window.

The next step is to define nodes on the tree. Each node *G* defines either an exposure (if at the leaf level) or a group of related exposures, i.e., a branch on the tree. In the case of an exposure-based tree, this is a class or sub-class of medical products. The sums of the observed number of outcomes in this node in the risk and comparison window are denoted as c_G and n_G respectively. Note that a single leaf is one potential node, but a node could also be an entire branch of the tree.

The log likelihood ratio is derived from a Binomial-based maximum likelihood estimator and is:

$$LLR = \ln\left(\frac{\left(\frac{c_G}{c_G+n_G}\right)^{c_G}\left(\frac{n_G}{c_G+n_G}\right)^{n_G}}{p^{c_G}(1-p)^{n_G}}\right) I\left(\frac{c_G}{c_G+n_G} > p\right)$$

where:

> p is the length of the risk window divided by the sum of the lengths of the risk and comparison windows. This represents the Bernoulli probability under the null hypothesis that the outcome occurs in proportion to the length of the window.
>
> $I()$ is the indication function, which is 1 when there are more outcomes in the risk window than would be expected by chance. It is included to ensure that we are looking for an excess risk of the having the adverse event rather than a protective decreased risk.

Log likelihood ratios are computed for computational convenience and results from them are equivalent to results based on likelihood ratios. The order in which the nodes are evaluated does not impact the results. The node *G* with the maximum LLR is the most likely cluster of unexplained outcomes in the risk window and its log likelihood ratio is the test statistic:

$$T = \max_G LLR(G)$$

The distribution of T is not known analytically, and so inference is conducted using Monte Carlo hypothesis testing.[27] First, a user-defined number of random data sets (e.g., 99,999) are generated under the null hypothesis that the observed number of outcomes in the risk window should be proportional to the length of the risk window relative to the observation window. *T* is calculated for the 99,999 random data sets and the 1 real data set.

If the *T* in the real data is among the 5% highest of all the maxima from the real and 99,999 random data sets generated under the null hypothesis, then that node constitutes a signal at the *alpha=0.05* statistical significance level. The Monte Carlo based p-value is calculated as *p=R/(99999 + 1)*, where *R* is the rank of the *T* in the real data set in relation to the *T* in the random data sets. That way the method formally adjusts the p-values for the multiple testing generated by the many overlapping groupings of exposures. This means that, when the null hypothesis is true, there is a 95% probability that all p-values

are greater than 0.05, or in other words, that there is not a single exposure-outcome pair or grouping with p≤0.05.

B. CONDITIONAL BERNOULLI SCAN STATISTIC WITH FIXED RISK WINDOW

When using the unconditional Bernoulli tree-based scan statistic described above, the null hypothesis is that any outcome is likely to occur in proportion to the length of the risk and comparison windows. In the conditional version, the lengths of the two windows are ignored, and instead the null hypothesis is based on the proportion of the sum of outcomes in the risk window of a particular node as compared to the total number of outcomes in the risk window observed in the whole tree.

Thus, we calculate the total number of outcomes in the risk window $C = \sum_i c_i$ observed in the whole tree and the total number of outcomes in the comparison window $N = \sum_i n_i$ observed in the whole tree.

So, when comparing the unconditional to the conditional, the probability *p* used above is now replaced by $\left(\frac{C}{C+N}\right)$.

The LLR for the conditional Bernoulli tree-based scan statistic is

$$LLR = \ln\left(\frac{\left(\frac{c_G}{c_G+n_G}\right)^{c_G}\left(\frac{n_G}{c_G+n_G}\right)^{n_G}}{\left(\frac{C}{C+N}\right)^{c_G}\left(\frac{N}{C+N}\right)^{n_G}}\right) I\left(\frac{c_G}{c_G+n_G} > \frac{C}{C+N}\right)$$

I() is the indication function, which is 1 when there are more outcomes in the risk window than would be expected by chance. It is included to ensure that we are looking for an excess risk of the having the adverse event rather than a protective decreased risk.

Again, log likelihood ratios are used for computational convenience as opposed to likelihood ratios. The order in which the nodes are evaluated does not impact the results. The node *G* with the maximum LLR is the most likely cluster of unexplained outcomes in the risk window and its log likelihood ratio is the test statistic:

$$T = \max_G LLR(G)$$

The other difference occurs in the Monte Carlo simulation step. Now, every random data set has to have the same *C* and *N* as the real data, so that the total number of outcomes in the risk window and control windows are the same in both the real and all the random data sets. The rest of the procedure is the same as described above.

C. UNCONDITIONAL TREE-TEMPORAL SCAN STATISTIC WITH VARYING RISK WINDOW

The unconditional tree-temporal scan statistic – also called the tree-temporal scan – adds a temporal dimension to the data. Now, in addition to the multiple hypotheses tested based on the tree structure as in the fixed risk window studies, each node itself contributes multiple temporal hypotheses related to the length of the risk and comparison windows.

$$LLR = \ln\left(\frac{\left(\frac{c_G}{c_G + n_G}\right)^{c_G}\left(\frac{n_G}{c_G + n_G}\right)^{n_G}}{\left(\frac{w}{O}\right)^{c_G}\left(\frac{O-w}{O}\right)^{n_G}}\right) I\left(\frac{c_G}{c_G + n_G} > \frac{w}{O}\right)$$

where:

c_G is the number of outcomes in the node *G* of interest that are also in the variable risk window

n_G is the number of outcomes in the node that are NOT in the variable risk window

w is the length of the variable risk window

O is the length of the total observation window.

$I()$ is the indication function, which is 1 when there are more outcomes in the risk window than expected under the null, and it is included to ensure that we are looking for an excess risk of the having the adverse event rather than a protective decreased risk. Note that *O* is a constant that is the same for every node and every potential risk window (i.e., time interval of interest).

Similar to the unconditional fixed risk window analysis described above, the null hypothesis is again that the outcome occurs in proportion to the length of the risk window relative to the total observation window.

As before, log likelihood ratios are used for computational convenience as opposed to likelihood ratios. The order in which the nodes are evaluated does not impact the results. The node *G* with the maximum LLR is the most likely cluster of unexplained outcomes in the risk window and its log likelihood ratio is the test statistic:

$$T = \max_G LLR(G)$$

The Monte Carlo simulation step occurs similarly as described before.

D. CONDITIONAL TREE-TEMPORAL SCAN STATISTIC WITH VARYING RISK WINDOW

Similar to the conditional Bernoulli tree-based scan statistic used with a fixed risk window analysis, the probability under the null hypothesis is based on the proportion of the sum of outcomes in the risk window as compared to the total outcomes observed. Additionally, the temporal definition further defines the number of hypotheses being tested.

$$LLR = \left(c_G \ln\frac{c_G}{u_G} + ((C+N) - c_G)\ln\left(\frac{(C+N) - c_G}{(C+N) - u_G}\right)\right) I(c_G > u_G)$$

where:

c_G is the number of outcomes in the node *G* of interest that are also in the variable risk window

n_G is the number of outcomes in the node that are NOT in the variable risk window

C is the total number of outcomes in the variable risk time window for all nodes.

N is the total number of outcomes NOT in the variable risk window for all nodes.

u_G is the expected number of outcomes in the variable risk window in the node under the null hypothesis.

$$u_G = (C)\left(\frac{c_G + n_G}{C+N}\right)$$

$I()$ is the indication function, which is 1 when there are more outcomes in the variable risk window than expected under the null, and it is included to ensure that we are looking for an excess risk of the having the adverse event rather than a protective decreased risk.

Again, log likelihood ratios are used for computational convenience as opposed to likelihood ratios. The order in which the nodes are evaluated does not impact the results. The node *G* with the maximum LLR is the most likely cluster of unexplained outcomes in the risk window and its log likelihood ratio is the test statistic:

$$T = \max_{G} LLR(G)$$

As in the conditional Bernoulli scan statistic, every random data set has to have the same total number of outcomes for the same day in both the real and all the random data sets. The rest of the procedure is the same as described above.

X. APPENDIX B – PARAMETER SUMMARY AND PROGRAM STEPS

A. PARAMETER SETTING SUMMARY

Table B 1. Parameter Setting Summary

Variable Symbol	Variable name	Angioedema	Achilles tendon rupture
Y	Enrollment gap	45 days	45 days
	Age strata	18-44, 45-54, 55-64, 65+	
P	Incident encounter settings	AV, IP, ED	AV, IP, ED
F	Incident outcome washout	64 days	127 days
W	Care Setting of Lookback Outcomes	AV, IP, ED	AV, IP, ED
B	Blackout period	0-0 days	0-0 days
R	Observation window	1-63 days	1-126 days
D	Exposure incidence washout	127 days	127 days
Z	Drug tree Incidence level	3rd and 4th	3rd and 4th

Abbreviations: AV: ambulatory, IP: inpatient, ED: emergency department

B. DEFINITIONS

R=Observation Window. Length of Observation Window = *R*end –*R*start +1

B=Blackout Period. Length of Blackout Period = *B*end-*B*start+1

D=Exposure Washout. Length of Exposure Washout=*D*end-*D*start+1

F=Outcome Washout. Length of Outcome Washout=*F*end-*F*start+1

C. ASSUMPTIONS/NOTES

1. Everyone must have drug and medical coverage to meet enrollment requirements
2. First date that incident outcome may be observed = (Query Start Date) + Max (*R*+*B*+1, *D*+1,*F*).
3. Each patient can only have one incident episode of exposure per *Z* grouping. Thus, *D*>=*R*+*B*.

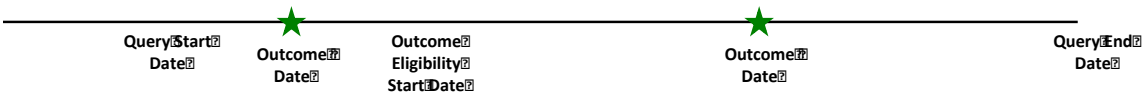
4. Each patient can only have one incident outcome of interest in the same observation window. Thus *F*>=*R*+*B*.
5. Risk Window slides within Observation Window in Tree-Temporal Scan. Whatever is not in the Risk Window is in the Comparison Window.
6. Risk Window Length >2 days AND <=0.5**R* days
7. Patients may contribute multiple drug starts so long as they meet incidence criteria.
8. This is a point exposure analysis.

D. PROGRAM STEPS WITH FIGURES

1. Within query start date and query end date, find the window in which outcomes can occur so adequate washout periods can be assessed: the outcome eligibility period. First outcome can occur is Query Start Date + MAX(*R+B+1, D+1, F*) to allow for washouts.
2. Within the Outcome Eligibility Period, identify all outcomes based on outcome code algorithm in *P* settings.
   a. For Angioedema, this is ICD-9-CM code 995.1 in three settings (IP, AV, ED)
   b. For Achilles tendon rupture, this is ICD-9-CM code 727.67 in the outpatient, inpatient, or emergency department setting. Of this universe of patients, the cohort is further reduced to patients that have one of the following CPT codes that represents a repair procedure: 27605, 27606, 27650, 27652, 27654, 01472 in any setting in the 30 days following or prior to the appearance of 727.67.

In **Figure B 1**, only the 2nd outcome date would count. The first outcome would be eliminated because it would occur before the outcome eligibility start date, which prevents outcomes from entering the dataset for which complete washout periods are not available.

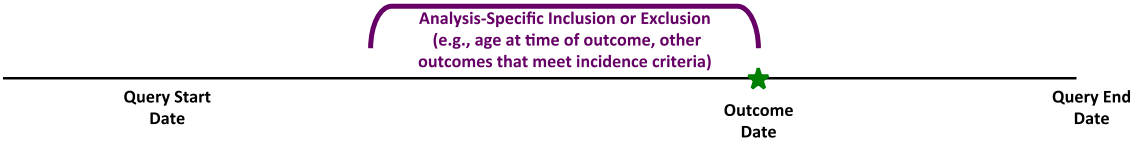
Figure B 1. Example illustrating outcome eligibility



3. Check that outcomes meet age-criteria, i.e., outcomes occurred when patient is in one of the valid age strata requested for the program.
   a. For Angioedema and Achilles tendon rupture, the strata are 18-44, 45-54, 55-64, 65+.

In **Figure B 2**, valid outcomes have already been ascertained based on the age at the time of the outcome.

Figure B 2. Example illustrating age-specific inclusions



4. Check that the outcome has enough prior enrollment in order to determine incidence. Outcome washout period is *F* days prior to the outcome (i.e., if the outcome is Day 0, then outcome washout period is [-*F*, -1].) Therefore member requires continuous enrollment during [-*F*, -1].

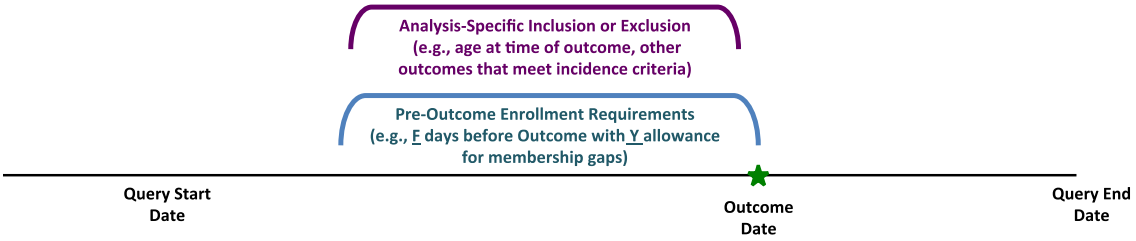
Check that member has medical coverage (note, here drug coverage is not required) with *Y* days allowance of membership gap.

a. For Angioedema, *F*=64 days, *Y*=45 days.

b. For Achilles tendon rupture, *F*=127 days, *Y*=45 days.

In **Figure B 3**, outcomes must have required pre-outcome enrollment.

Figure B 3. Example illustrating enrollment requirements.



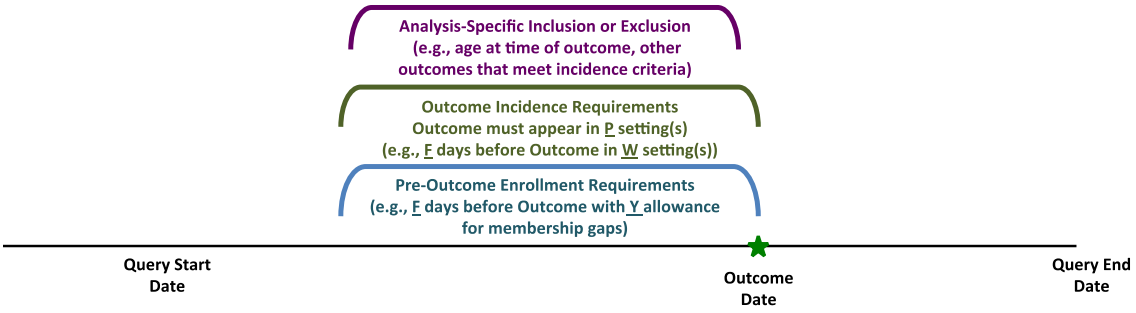
5. If member has required coverage, then check that outcome is incident, i.e., some set of electronic codes did not occur in *W* settings in the [-*F*,-1] days prior to the identified outcome.

   a. For Angioedema, ICD-9-CM code is 995.1 and *W* = IP, ED, AV.

   b. For Achilles tendon rupture, ICD-9-CM codes are 727.67 and CPT codes 27605, 27606, 27650, 27652, 27654, 01472 and *W*=IP, ED, AV.

In **Figure B 4**, outcomes have required enrollment and now have to be assessed for incidence.

Figure B 4. Example illustrating outcome incidence requirements.



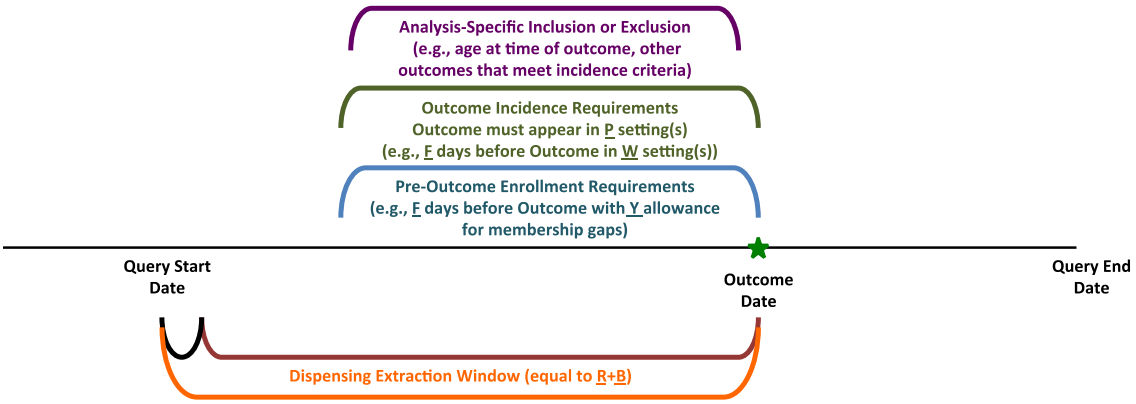
6. If multiple outcomes meet incidence criteria, then only keep the first outcome.
7. Extract potential drug dispensings based on date of outcome. If outcome occurrence is given t=0, then all drug dispensings in [(-*R*+ -*B*) ,-1] will be extracted.

   a. For Angioedema, *B*start=0, *B*end=0, *R*start=1, *R*end=63.

   b. For Achilles tendon rupture, *B*start=0, *B*end=0, *R*start=1, *R*end=126.

In **Figure B 5**, the outcome has been established as incident. Now, dispensings must be extracted in the period preceding the outcome.

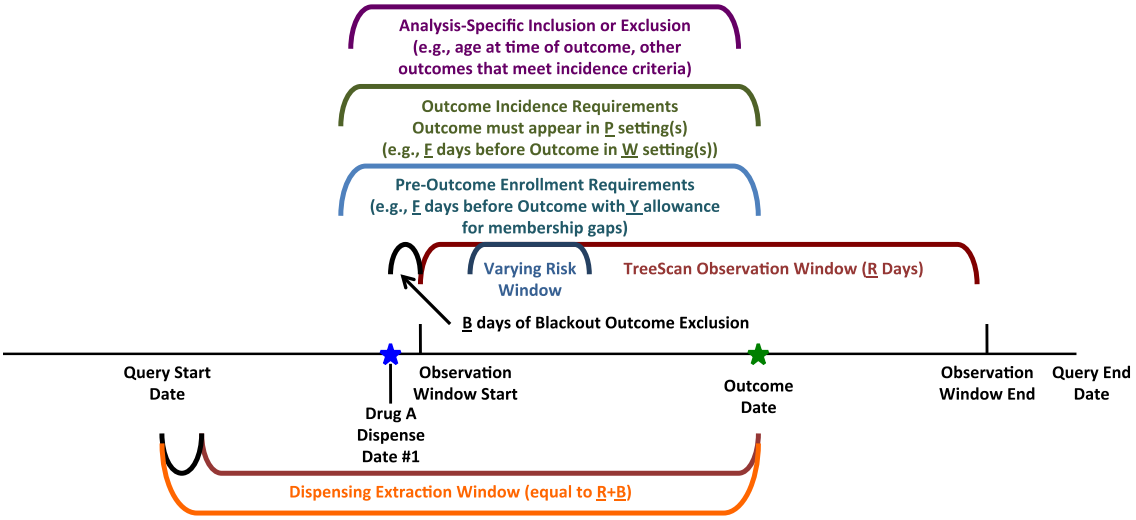
Figure B 5. Example illustrating dispensing extraction window



8. Check that outcome doesn't fall in the blackout window. That is, if time-to-event from exposure index date to outcome index date <=length(*B*) (blackout period), discard the dispensing.

In **Figure B 6**, a potential drug dispensing / exposure of interest is identified and a check is performed to ensure that the outcome of interest does not fall in the blackout period.

Figure B 6. Example step illustrating blackout window check.

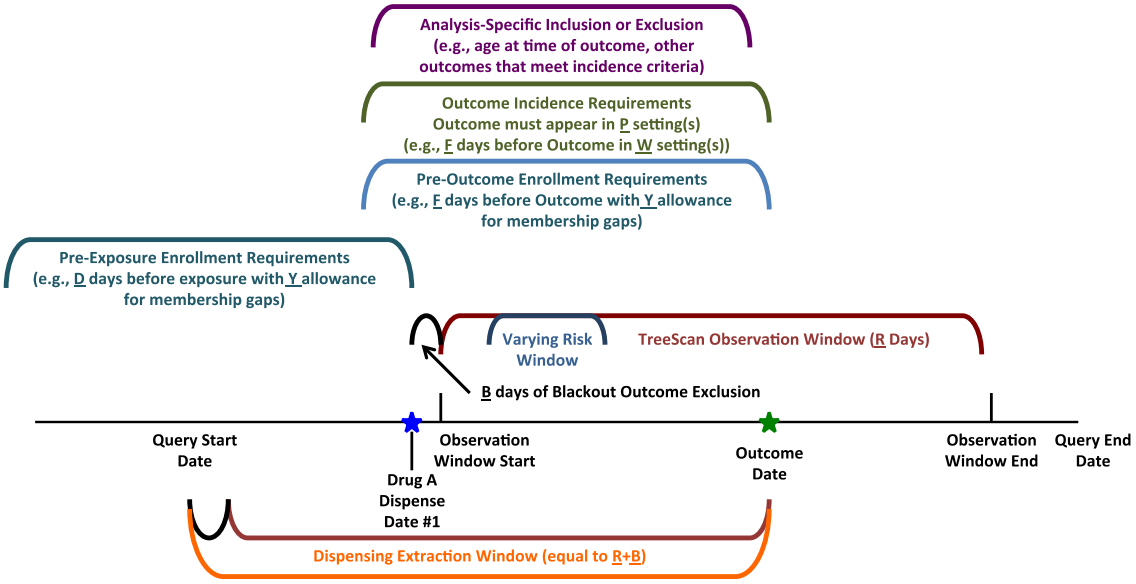


9. Check that exposure has enough prior enrollment in order to determine incidence. That is member must have **drug** coverage (i.e., medical coverage not required) in [-*D*,-1] days prior to exposure when exposure is assigned as new day 0. They may have *Y* membership gaps during that time.

    a. For Angioedema and Achilles tendon rupture, *D*=127 days.

In **Figure B 7**, new dispensings / exposures have to have sufficient pre-exposure enrollment to assess exposure incidence.

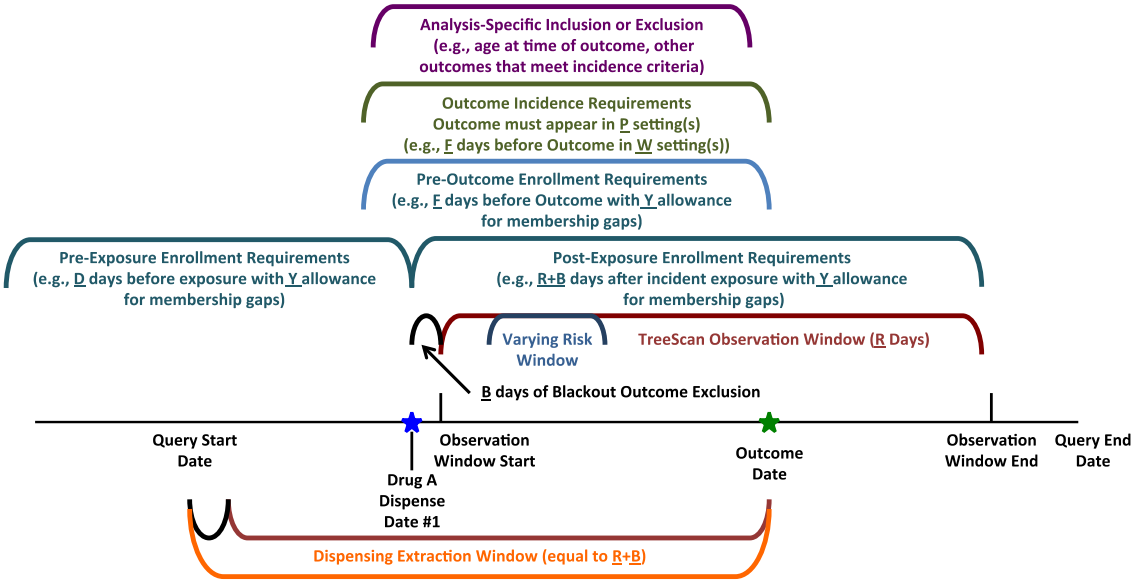
Figure B 7. Example step illustrating the pre-exposure enrollment requirements.



10. Check that exposure has enough post-exposure enrollment to satisfy the null hypothesis. That is, member must have **medical** coverage (i.e., drug not required) in [1, *R+B*] days post-exposure when exposure is assigned as new day 0. They may have Y membership gaps during that time.

In **Figure B 8**, dispensings / exposures have to be assessed for post-exposure enrollment requirements.

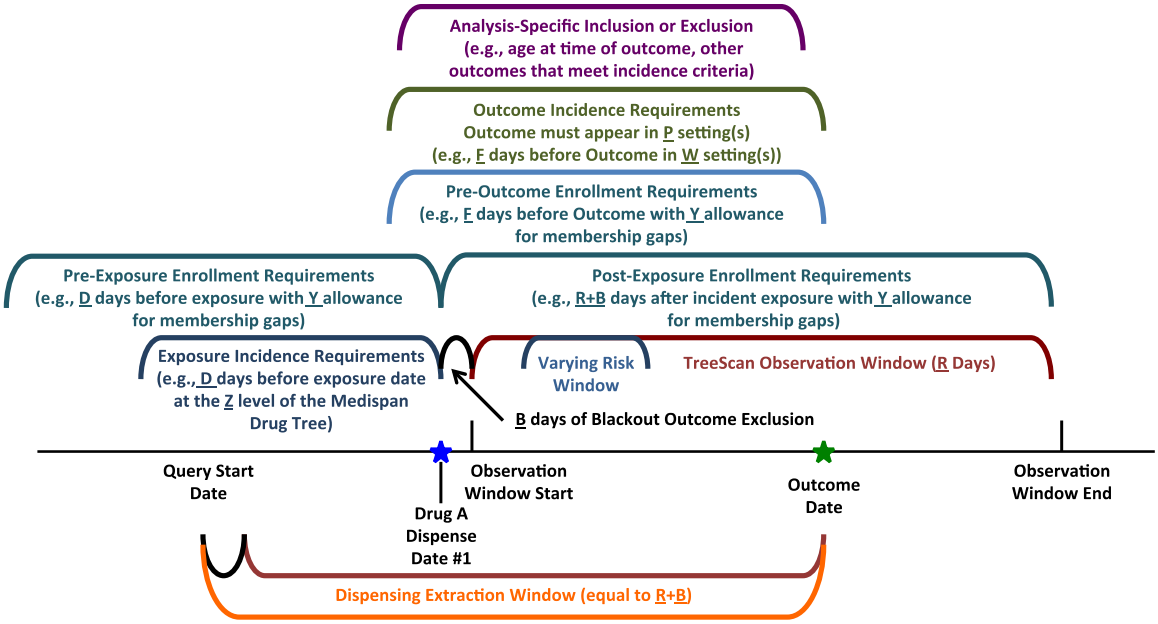
Figure B 8. Example step illustrating post-exposure enrollment requirements.



11. Check that exposure is incident. Exposure is first that appears at *Z* level of the Medi-Span Drug Tree in the *D* days prior to the exposure date.
    a. For Angioedema and Achilles tendon rupture, we will try multiple *Z* levels – 3rd and 4th.

In **Figure B 9**, the dispensings / exposures are assessed for incidence.

Figure B 9. Example step illustrating exposure incidence requirements.



12. For same-day incident dispensings at the *Z* level of the Medi-Span drug tree, perform tie-breaking procedures based on tie-breaker rules.
    a. For same-day incident dispensings, choose drug with longest days supplied
    b. For same-day incident dispensings with equal number of days supplied, choose drug with least frequent number of unique users among the age-appropriate cohort.
13. Output time-to-event data for incident drug dispensings by node.

XI. APPENDIX C – ADDITIONAL RESULTS FOR INCIDENCE AT THE 4TH LEVEL

A. ANGIOEDEMA

Table 12 reports the results of the conditional tree-temporal scan statistic with a varying risk window analysis applied to the angioedema dataset when incidence was set to the 4th level. There were 20 unique alerting nodes located in 15 branches of the tree.

Table 12. Data-Mining Results of the Conditional Tree-Temporal Scan Statistic with a Varying Risk Window Analysis in the Angioedema Dataset with Incidence set to the 4th Level.

Node name	Tree level	Node cases	Risk window	Cases in window	Expected cases	Test statistic	P-value
Prednisone	4	7934	1-7	2420	1485.3	250.4	<0.001
Epinephrine	4	3640	1-15	1925	1253.9	156.0	<0.001
Famotidine	4	799	1-6	305	130.1	85.0	<0.001
Hydroxyzine	4	2203	1-4	474	260.0	70.8	<0.001
..Hydroxyzine HCl	5	1956	1-4	413	230.9	58.2	<0.001
..Hydroxyzine Pamoate	5	247	1-10	106	61.1	13.5	0.003
Methylprednisolone	4	4481	1-8	1317	934.5	70.0	<0.001
..Methylprednisolone	5	4479	1-8	1316	934.1	69.8	<0.001
Sulfamethoxazole-Trimethoprim	5	1812	3-13	641	431.7	44.3	<0.001
Ranitidine	4	968	1-5	227	135.3	25.8	<0.001
Diphenhydramine	4	160	1-3	50	15.0	25.3	<0.001
..Diphenhydramine HCl	5	159	1-6	70	25.9	25.5	<0.001
Lisinopril	4	2210	2-8	507	379.8	19.3	<0.001
Minocycline	4	231	18-19	29	7.3	18.2	<0.001
Levofloxacin	4	1240	1-3	177	116.1	13.8	0.002
..Levofloxacin	5	1239	1-3	177	116.0	13.8	0.002
Bupropion	4	390	19-30	122	73.1	13.6	0.002
Simvastatin	4	915	42-54	211	147.2	12.2	0.012
Triazolam	4	35	51-52	8	0.8	11.5	0.025
Clindamycin	4	770	1-3	116	72.1	11.3	0.033

Table 13 reports the results of the conditional Bernoulli scan statistic with a fixed risk window analysis applied to the angioedema dataset when incidence was set to the 4th level. There were 16 unique alerting nodes located in 11 branches of the tree.

Table 13. Data-Mining Results of the Conditional Bernoulli Scan Statistic with a Fixed Risk Window Analysis in the Angioedema Dataset with Incidence set to the 4th Level.

Node name	Tree level	Node cases	Risk window	Cases in window	Expected cases	Test statistic	P-value
Prednisone	4	7934	1-28	5508	4370.9	364.9	<0.001
Epinephrine	4	3640	1-28	2656	2005.3	254.5	<0.001
Methylprednisolone	4	4481	1-28	2949	2468.6	110.6	<0.001
..Methylprednisolone	5	4479	1-28	2947	2467.5	110.3	<0.001
Hydroxyzine	4	2203	1-28	1515	1213.6	87.7	<0.001
..Hydroxyzine HCl	5	1956	1-28	1336	1077.6	72.4	<0.001
..Hydroxyzine Pamoate	5	247	1-28	179	136.1	15.8	<0.001
Famotidine	4	799	1-28	616	440.2	84.3	<0.001
Ranitidine	4	968	1-28	655	533.3	32.1	<0.001
Cimetidine	4	201	1-28	147	110.7	13.9	<0.001
..Cimetidine	5	200	1-28	146	110.2	13.6	<0.001
Sulfamethoxazole-Trimethoprim	5	1812	1-28	1169	998.2	33.7	<0.001
Diphenhydramine	4	160	1-28	129	88.1	23.1	<0.001
..Diphenhydramine HCl	5	159	1-28	128	87.6	22.7	<0.001
Lisinopril	4	2210	1-28	1363	1217.5	20.0	<0.001
Clopidogrel	4	419	1-28	270	230.8	7.6	0.049

B. ACHILLES TENDON RUPTURE

Table 14 reports the results of the conditional tree-temporal scan statistic with a varying risk window analysis applied to the Achilles tendon rupture dataset when incidence was set to the 4th level. There were 11 alerting nodes located in 9 branches of the tree.

Table 14. Data-Mining Results of the Conditional Tree-Temporal Scan Statistic with a Varying Risk Window Analysis in the Achilles Tendon Rupture Dataset with Incidence set to the 4th level.

Node name	Tree level	Node cases	Risk window	Cases in window	Expected cases	Test statistic	P-value
Hydrocodone Combination - Two Ingredient	4	1389	1-9	851	335.5	287.1	<0.001
..Hydrocodone-Acetaminophen	5	1366	1-9	844	330.0	289.1	<0.001
Opioid Combination - Two Ingredient	4	746	1-7	508	160.9	241.6	<0.001
..Oxycodone w/ Acetaminophen	5	745	1-7	507	160.7	240.9	<0.001
Ibuprofen	4	445	1-14	268	128.6	58.2	<0.001
Promethazine	4	185	1-8	113	42.2	40.6	<0.001
Oxycodone	4	99	1-7	66	21.4	29.9	<0.001
Cephalexin	4	310	1-7	136	66.9	27.6	<0.001
Enoxaparin	4	66	1-10	49	16.6	20.7	<0.001
Ketorolac	4	49	1-8	34	11.2	15.0	<0.001
Cefadroxil	4	35	1-8	25	8.0	11.5	0.025

Table 15 reports the results of the conditional Bernoulli scan statistic with a fixed risk window analysis applied to the Achilles tendon rupture dataset when incidence was set to the 4th level. There were 15 unique alerting nodes located in 13 branches of the tree.

Table 15. Data-Mining Results of the Conditional Bernoulli Scan Statistic with a Fixed Risk Window Analysis in the Achilles Tendon Rupture Dataset with Incidence set to the 4th level.

Node name	Tree level	Node cases	Risk window	Cases in window	Expected cases	Test statistic	P-value
Hydrocodone Combination - Two Ingredient	4	1389	1-28	1024	535.1	397.8	<0.001
..Hydrocodone-Acetaminophen	5	1366	1-28	1010	526.3	395.3	<0.001
Opioid Combination - Two Ingredient	4	746	1-28	624	287.4	341.6	<0.001
..Oxycodone w/ Acetaminophen	5	745	1-28	623	287.0	340.8	<0.001
Ibuprofen	4	445	1-28	307	171.4	87.5	<0.001
Oxycodone	4	99	1-28	83	38.1	43.5	<0.001
Promethazine	4	185	1-28	133	71.3	42.9	<0.001
Enoxaparin	4	66	1-28	55	25.4	28.2	<0.001
Cephalexin	4	310	1-28	184	119.4	28.1	<0.001
Ketorolac	4	49	1-28	40	18.9	19.2	<0.001
Cefadroxil	4	35	1-28	30	13.5	16.7	<0.001
Ondansetron	4	75	1-28	50	28.9	12.2	0.000
..Ondansetron HCl	5	37	1-28	30	14.3	14.1	0.000
Hydroxyzine Pamoate	5	26	1-28	21	10.0	9.8	0.002
Meperidine	4	22	1-28	18	8.5	8.7	0.005
Naproxen	4	328	1-28	162	126.4	8.2	0.011

Table 16 reports the results of the conditional Bernoulli scan statistic with a fixed risk window analysis applied to the Achilles tendon rupture dataset when incidence was set to the 4th level. There were 13 unique alerting nodes located in 11 branches of the tree.

Table 16. Data-Mining Results of the Conditional Bernoulli Scan Statistic with a Fixed Risk Window Analysis in the Achilles Tendon Rupture Dataset with Incidence set to the 4th level.

Node name	Tree level	Node cases	Risk window	Cases in window	Expected cases	Test statistic	P-value
Hydrocodone Combination - Two Ingredient	4	1389	1-63	1172	847.5	200.2	<0.001
..Hydrocodone-Acetaminophen	5	1366	1-63	1153	833.4	197.1	<0.001
Opioid Combination - Two Ingredient	4	746	1-63	670	455.2	164.8	<0.001
..Oxycodone w/ Acetaminophen	5	745	1-63	669	454.5	164.4	<0.001
Ibuprofen	4	445	1-63	366	271.5	48.6	<0.001
Oxycodone	4	99	1-63	93	60.4	29.1	<0.001
Promethazine	4	185	1-63	150	112.9	17.6	<0.001
Enoxaparin	4	66	1-63	60	40.3	15.3	<0.001
Cephalexin	4	310	1-63	234	189.1	14.9	<0.001
Ketorolac	4	49	1-63	45	29.9	12.2	<0.001
Cefadroxil	4	35	1-63	32	21.4	8.4	0.006
Ondansetron HCl	5	37	1-63	33	22.6	7.4	0.026
Meperidine	4	22	1-63	21	13.4	7.3	0.028