SENTINEL METHODS PROTOCOL

EVALUATION OF THREE SELF-CONTROLLED METHODS FOR SIGNAL DETECTION: TREESCAN, INFORMATION COMPONENT TEMPORAL PATTERN DISCOVERY, AND SEQUENCE SYMMETRY ANALYSIS

Prepared by: Judith C. Maro, PhD, MS,[1] Shirley V. Wang,PhD, ScM,[2] Inna Dashevsky, MS,[1] David Cole, BM,[1] Joshua J. Gagne, PharmD, ScD,[2] Sai Dharmarajan, PhD,[3] Esther H. Zhou, MD, PhD,[4] Sandra DeLuccia, MPH,[1] Ella Pestine, MPH,[1] Monica Munoz, PharmD, MS,[4] Danijela Stojanovic, PharmD, PhD,[4] Jesper Hallas, MD, DrMedSc,[5] G.Niklas Norén, PhD,[6] Martin Kulldorff, PhD,[2] Michael D. Nguyen, MD[4]

Author Affiliations: 1. Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA 2. Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 3. Office of Biostatistics, Office of Translational Sciences, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD 4. Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD 5. Clinical Pharmacology and Pharmacy, Department of Public Health, University of Southern Denmark, Odense, Denmark, 6. Uppsala Monitoring Centre, Uppsala, Sweden

Version 1: November 30, 2018
Version 2: April 19, 2019

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.

History of Modifications

Version	Date	Modification	By
V2	April 19, 2019	• The study period was altered to include only ICD-9-CM era codes to further reduce temporal variations. • The study data source was changed for logistic reasons. • Upon initial experimentation with the number of national drug codes (NDCs) required for a calibration study, it was determined that 100 was too low. Thus, the number of randomly selected NDCs was increased to 5000 for a more robust calibration exposure.	Judith C. Maro

Sentinel Methods Protocol

Evaluation of Three Self-Controlled Methods for Signal Detection: TreeScan, Information Component Temporal Pattern Discovery, and Sequence Symmetry Analysis

Table of Contents

I. INTRODUCTION

Post-market medical product signal detection – identifying unexpected potential associations between medical product exposures and health outcomes of interest in a real-world population -- is an important component of the U.S. Food and Drug Administration (FDA)'s mission to protect public health. There are multiple post-market signal detection methods that can be implemented in longitudinal healthcare databases to assess potential elevated frequencies of adverse events. These methods can support both cohort and self-controlled designs. Comparisons among methods implemented with stratified cohort designs have been performed, and the methods had similar findings in a small empirical dataset.[1] Three specific signal detection methods that can be utilized with a self-controlled risk interval design[2,3], which adjusts for time-invariant confounding, include a) tree-based scan statistics[4,5] operationalized in TreeScan™, b) Information Component Temporal Pattern Discovery[6–9] , and c) Sequence Symmetry Analysis[10–12]. Evaluations have been conducted on each of these methods using longitudinal data, however direct comparisons between these methods have not yet been performed. We propose to conduct an analysis using all three methods with the same empiric and simulated datasets to assess variation in findings.

We chose a clinical study problem to test the methods but emphasize that this project is not intended to be a safety assessment. We will evaluate whether there is an increased frequency of adverse events among new adult users of two oral anti-seizure treatments by comparing a window of time shortly after drug initiation (i.e., the risk window) to referent windows that may occur either prior to or after drug initiation (i.e., the comparison windows).

II. SPECIFIC AIMS

In a self-controlled risk interval design, only exposed individuals are included, and we will further limit our analysis to newly-exposed individuals. All analyses will treat exposure to the study drug of interest as a point exposure, meaning that any incident drug dispensing, irrespective of the number of days supplied, will be considered a qualifying exposure. Health outcomes of interest will be assessed in a designated risk window as well as several comparison windows occurring before or after the risk window. We will require enrollment during these observation windows to ensure an equal likelihood of observing health outcomes of interest.

A. EMPIRICAL COMPARISONS

We have identified two study exposures of interest and will evaluate these two exposures with all three methods. We will not name a primary analysis as this is a methods evaluation that will involve multiple comparisons varying a) the grouping of outcomes of interest, b) the timing of the referent comparison windows relative to exposure, and c) the length of such windows. A full list of analyses is listed later in this protocol in **Table 1**. The empiric evaluation provides a real-world assessment of the performance of each method when applied to drugs with well-characterized safety profiles.

B. SIMULATED DATA ASSESSMENTS

Using empirical data gathered for the two study exposures of interest, we will choose some outcomes and add investigator-inserted occurrence during the risk window according to a pre-specified simulated association. The occurrence of all other outcomes will be randomized according to their background rate occurrence. The resultant simulated dataset will then consist of "positive" outcomes with pre-specified associations and the rest will be "null outcomes." We will vary the strength of the simulated association and the prevalence of the outcomes with the simulated association to evaluate the statistical power of the three methods to detect the investigator-inserted elevated frequency. The simulation is important because it evaluates each method's statistical power in detecting known associations that have been artificially inserted at different elevated risk levels.

III. METHODS

A. SURVEILLANCE POPULATION, ENROLLMENT CRITERIA AND STUDY EXPOSURES

Data will be obtained from IBM® MarketScan® Research Databases from the years 2010 to September 30, 2015. We will include individuals who were dispensed the study exposures of interest on or after their 18th birthday during the available date range without further inclusion or exclusion criteria. There are two study exposures of interest: oral forms of levetiracetam and lamotrigine used in the adult population. Additionally, there is one "calibration" exposure that must be collected to implement one of the methods. The two study exposures were chosen because they have relatively well-characterized safety profiles and are used primarily for treatment for seizure disorders which are medical conditions that are not usually associated with rapid changes in overall clinical status (except for rate of seizures itself). Stability in the expected underlying clinical status is especially important when implementing self-controlled designs to isolate the effect of the exposure itself rather than other temporal confounding factors (e.g., generally declining or improving health state), a concept described in epidemiology as exchangeability.[13,14] Self-controlled designs compare time periods and it is only a fair comparison when factors outside of the incident dispensing of the study exposure are held as equal as possible through either design or analysis techniques. When the general risk of the health outcomes of interest is equivalent but for the study exposure, then the differences between the observed and expected rate beyond a predefined statistical threshold signifies a potential causal hypothesis requiring further investigation.

Although we will not assess outcomes in the period of time immediately preceding the incident dispensing of the study drug of interest, we still expect that both of these anti-seizure medications may show an increased frequency of seizure-related disorders in the pre-exposure comparison window. Nonetheless, we have chosen to include these outcomes within the groupings of outcomes being assessed. Lamotrigine also has other on- and off-label uses for mood disorders or pain that may create additional time-varying confounding by indication. We expect to triage any increased frequency of these adverse outcomes with these acknowledged limitations.

Only incident exposures of interest will be ascertained, and incidence will be defined as the first oral drug dispensing in 183 days. Patients will be required to be free of any form of the exposure of interest (e.g., intravenous, oral). Exposures of interest will be identified by National Drug Codes (NDCs) and Healthcare Common Procedure Coding System (HCPCS) codes listed in a separate code list document.

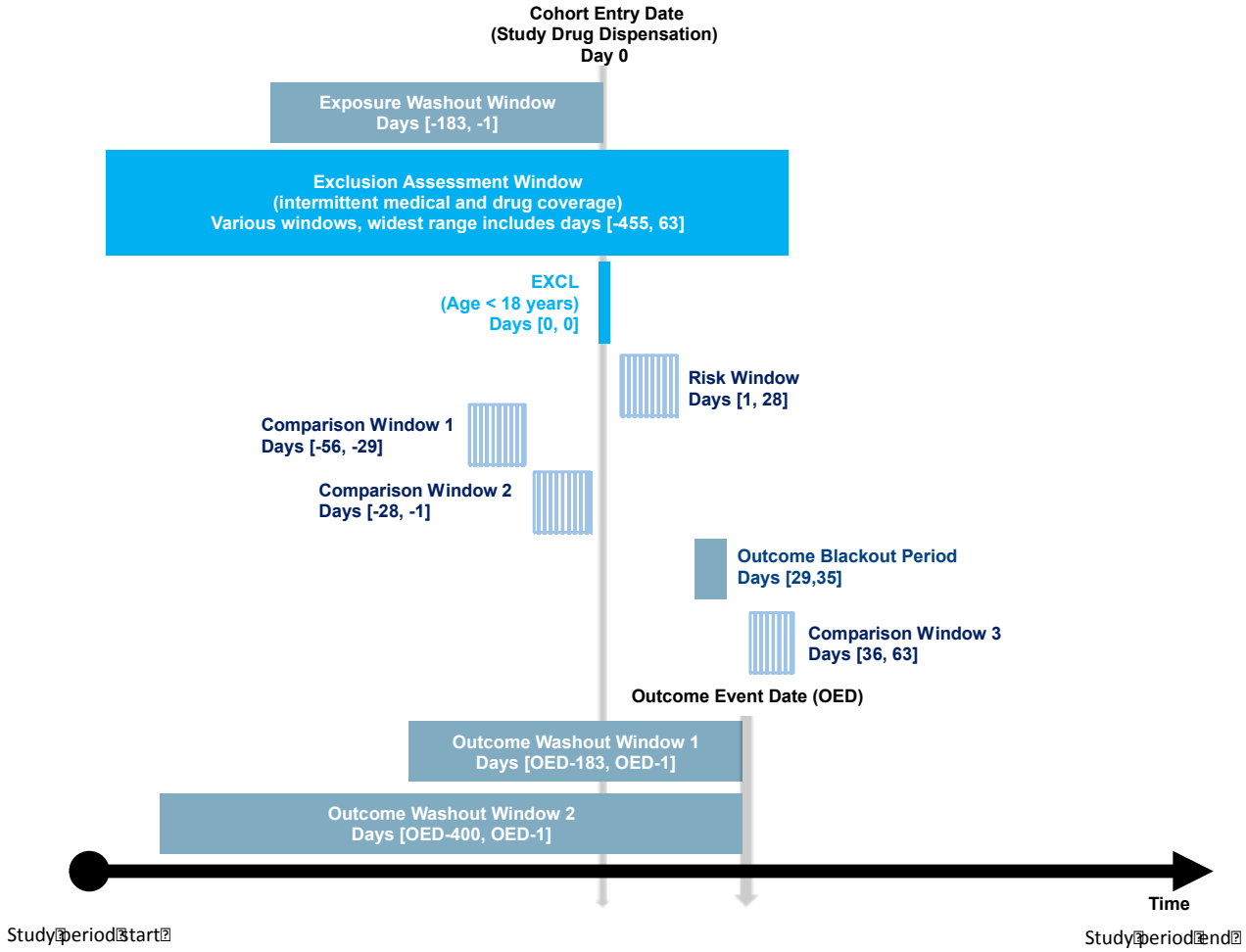
Enrollment is required to assess exposure incidence. Enrollment gaps of 45 days or less will be bridged, i.e., treated as continuously enrolled time.

Figure 1 is a visual guide to the various parameters that govern how the study cohort is created.

B. RISK AND COMPARISON WINDOWS

We are interested in examining multiple risk and comparison windows within the study period. A complete list of analyses is shown in **Table 1** below and illustrated in **Figure 1**. All windows are defined relative to the study exposure index date (i.e., Day 0). Inclusion of time in these windows includes whole weeks such that weekend or weekday days are equally represented.

Figure 1. Design Diagram

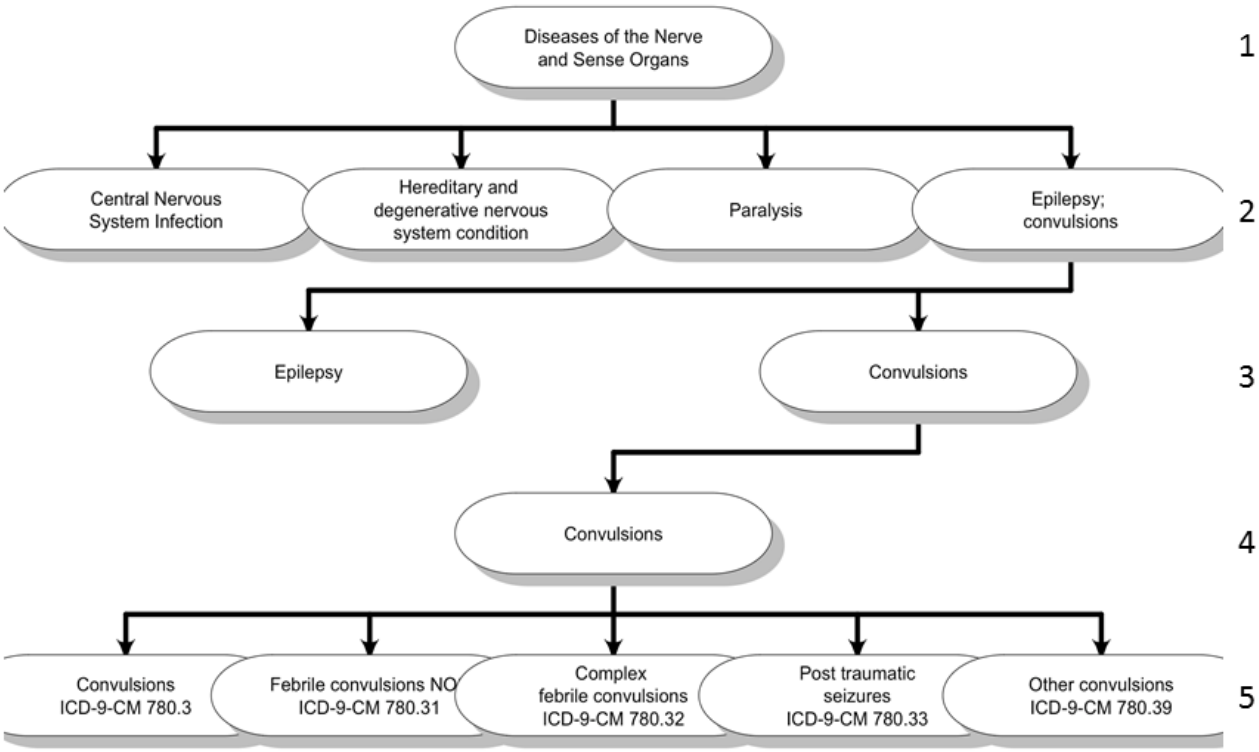


C. STUDY OUTCOMES IN A HIERARCHICAL DIAGNOSIS TREE

Use of a tree structure with hierarchical groupings of clinically related potential adverse events is inherent to TreeScan. The tree structure does two things: 1) it declares an outcome to be incident with respect to itself as well as clinically related groupings and 2) it eliminates the need for an investigator to prespecify how a clinician might record similar outcomes in the billing record (i.e., claims data) across thousands of different outcomes. Use of such trees is compatible with all three methods under consideration; we therefore will define outcomes based on their clinical grouping in the Multi-Level Clinical Classification System (MLCCS).

Outcomes will be identified using ICD-9-CM codes. All ICD-9-CM diagnoses are classified into a hierarchical tree structure defined by the MLCCS. The MLCCS is a product of the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project (http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp). It is a hierarchical system with four aggregate diagnosis levels, although on some branches there may only be two or three levels. The first and broadest level identifies 18 body systems, while the entries at the finest level contain one or multiple ICD-9-CM codes. For example, as can be seen in **Figure 2**, five different ICD-9-CM codes make-up the aggregate concept of "convulsions" which has the same grouping of outcomes at both the 3rd and 4th level of the MLCCS tree.

Figure 2. Example Section of the Multi-Level Clinical Classification Software Tree



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

- Some conditions unlikely to manifest themselves within the short follow-up time we are dealing with, such as cancer.
- History of pre-existing conditions (e.g., codes with nomenclature that refers to previously diagnoses medication conditions usually given as "History of…").
- Other outcomes very unlikely to be caused by exposure, such as well-care visits and normal delivery of an infant.

In general, these outcomes are not considered possibly exposure-associated and so we remove them from the tree to reduce multiple hypothesis testing.

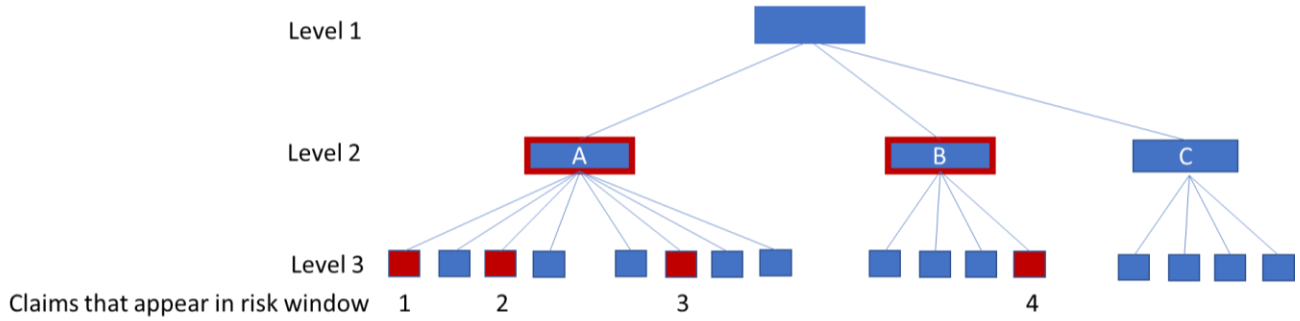
D. INCIDENT OUTCOMES OF INTEREST

This study will focus on incident outcomes that occur in either the risk or comparison windows, i.e., collectively the observation window. An incident outcome is defined as one that was observed in the pre-defined settings (see **Table 1**) during the observation window with no other outcomes in the same level and branch of the MLCCS tree in any setting during a pre-set number of prior days. We will perform analyses using both the fourth and fifth level, separately, as the incident outcome assessment level. The fourth level signifies that a never-before-seen ICD-9-CM code is not counted if a unique ICD-9-CM code belonging to the same level was observed during the outcome washout window. This prevents double counting of very closely related outcomes such as if someone were coded as having experienced both ICD-9-CM 780.3 (i.e., convulsions) and ICD-9-CM 780.31 (i.e., febrile convulsions) within a few days' time. **Figure 2** shows their relationship with each other with respect to the tree. Both of these codes are in the same 4th level node so only the first incident code within this grouping will be included in the analytic data set. Each member can contribute multiple incident outcomes in the risk and comparison windows as long as these outcomes are not part of the same branch of the MLCCS tree (**Figure 3**).

Figure 3. Counting Incidence Outcomes

Hypothetical example where incidence is defined at level 2

- Patient has 4 ICD9 diagnosis claims (leaf level) occurring in the exposure risk window
- At leaf level, patient contributes 2 outcomes, 1) the first occurring ICD9 code for A1, A2, A3 and 2) B4
- Patient contributes 2 outcomes at level 2, A and B



E. SUMMARY OF ANALYSES

Table 1 contains the list of analytic datasets that are needed to isolate the effect of particular study design choices, explained below. The risk window is constant and set at days 1-28 following incident study drug dispensing for all analyses.

1. Assess use of a pre-exposure vs. post-exposure comparison window to demonstrate the impact and trade-offs of using an unexposed referent window compared to a later post-exposure reference window. The timing of the windows is always relative to the day of the incident study exposure (i.e., Day 0).
2. Assess the impact of a 183-day outcome washout that seeks to preserve statistical power, compared to 400-day outcome washout that improves ability to detect truly incident outcomes. Outcome incidence is always assessed relative to the date of the outcome event.
3. Assess the incidence definition of the health outcome at either the 4th vs. 5th level of the MLCCS tree to improve that ability to detect only incident outcomes.
4. Assess the additional value of surveillance for outcomes that are primarily occurring in ambulatory encounter settings. Inclusion of such outcomes may allow for capture of conditions that will later become more serious and require hospitalization but may also add alerts that are not serious enough to warrant further consideration.

Table 1. Full List of Analytic Datasets for Two Study Exposures and One Calibration Exposure

Analytic Dataset ID	Study Exposure	Comparison Window	Outcome Incidence	Tree Level	Outcome Encounter Setting
1	1 (levetiracetam)	[-56,-29]	183	4	IP, ED
2	1 (levetiracetam)	[-28,-1]	183	4	IP, ED
3	1 (levetiracetam)	[36,63]	183	4	IP, ED
4	1 (levetiracetam)	[-56,-29]	400	4	IP, ED
5	1 (levetiracetam)	[-28,-1]	400	4	IP, ED
6	1 (levetiracetam)	[36,63]	400	4	IP, ED
7	1 (levetiracetam)	[-56,-29]	183	5	IP, ED
8	1 (levetiracetam)	[-28,-1]	183	5	IP, ED
9	1 (levetiracetam)	[36,63]	183	5	IP,ED
10	1 (levetiracetam)	[-56,-29]	400	5	IP,ED
11	1 (levetiracetam)	[-28,-1]	400	5	IP,ED
12	1 (levetiracetam)	[36,63]	400	5	IP,ED
13	1 (levetiracetam)	[-56,-29]	400	4	IP, ED, AV
14	1 (levetiracetam)	[-28,-1]	400	4	IP, ED, AV
15	1 (levetiracetam)	[36,63]	400	4	IP, ED, AV
16	2 (lamotrigine)	[-56,-29]	183	4	IP, ED
17	2 (lamotrigine)	[-28,-1]	183	4	IP, ED
18	2 (lamotrigine)	[36,63]	183	4	IP, ED
19	2 (lamotrigine)	[-56,-29]	400	4	IP, ED
20	2 (lamotrigine)	[-28,-1]	400	4	IP, ED
21	2 (lamotrigine)	[36,63]	400	4	IP, ED
22	2 (lamotrigine)	[-56,-29]	183	5	IP, ED
23	2 (lamotrigine)	[-28,-1]	183	5	IP, ED
24	2 (lamotrigine)	[36,63]	183	5	IP,ED
25	2 (lamotrigine)	[-56,-29]	400	5	IP,ED

Analytic Dataset ID	Study Exposure	Comparison Window	Outcome Incidence	Tree Level	Outcome Encounter Setting
26	2 (lamotrigine)	[-28,-1]	400	5	IP,ED
27	2 (lamotrigine)	[36,63]	400	5	IP,ED
28	2 (lamotrigine)	[-56,-29]	400	4	IP, ED, AV
29	2 (lamotrigine)	[-28,-1]	400	4	IP, ED, AV
30	2 (lamotrigine)	[36,63]	400	4	IP, ED, AV
31	3 (calibration)	[-56,-29]	183	4	IP, ED
32	3 (calibration)	[-28,-1]	183	4	IP, ED
33	3 (calibration)	[36,63]	183	4	IP, ED
34	3 (calibration)	[-56,-29]	400	4	IP, ED
35	3 (calibration)	[-28,-1]	400	4	IP, ED
36	3 (calibration)	[36,63]	400	4	IP, ED
37	3 (calibration)	[-56,-29]	183	5	IP, ED
38	3 (calibration)	[-28,-1]	183	5	IP, ED
39	3 (calibration)	[36,63]	183	5	IP, ED
40	3 (calibration)	[-56,-29]	400	5	IP, ED
41	3 (calibration)	[-28,-1]	400	5	IP, ED
42	3 (calibration)	[36,63]	400	5	IP, ED
43	3 (calibration)	[-56,-29]	400	4	IP, ED, AV
44	3 (calibration)	[-28,-1]	400	4	IP, ED, AV
45	3 (calibration)	[36,63]	400	4	IP, ED, AV

Notes: IP = Inpatient Setting; ED = Emergency Department Setting; AV = Ambulatory Care Setting. Risk Window is set to Days 1-28 for all analytic datasets.

An example analytic dataset is shown with simulated data in **Table 2.** The Analytic Dataset ID is directly mapped to the list above. The original diagnosis captures the incident ICD-9-CM code captured prior to mapping and the leaf level node represents the mapped value. The time of the outcome is relative to the incident study drug exposure that is always indexed as Day 0. The number of counts is the number of times this particular code was observed with this particular relative timing in the study.

Table 2. Sample Analytic Dataset with Simulated Data

Analytic Dataset ID	Original Diagnosis	Original Diagnosis Type	Leaf Level Node	Leaf Level Node Type	Relative Time of Outcome	Counts
1	250	9	250	9	-15	3

F. STATISTICAL METHODS

Most analytic datasets created will be analyzed with two versions of TreeScan, Information Component Temporal Pattern Discovery, and Sequence Symmetry Analysis. Sequence Symmetry Analysis is not intended to be used with a post-exposure comparison window and these scenarios will not include this analytic technique.

1. TreeScan

We will use both the conditional and unconditional Bernoulli tree-based scan statistic that detects elevated frequencies of outcomes in pre-specified time periods. We do not plan on testing the tree-temporal scan statistic although the data collected could be repurposed in a separate project. TreeScan™ (www.treescan.org) is a data mining method that scans electronic health data that are organized into clinically relevant groupings of medical product exposures or health outcomes.[4,5,15] With respect to the self-controlled risk interval design, TreeScan has been evaluated in simulation environments[16] as well as empirical environments.[17,18] For a pre-defined exposure, it adjusts for the multiple testing inherent in evaluating thousands of potential adverse events to assist with prioritization of alerts or statistically significant findings for further investigation.

Two versions of the Bernoulli test statistic are planned for this evaluation, and both rely on the computation of the log likelihood ratio for each outcome node *LLR(G).* The test statistic *T* is the maximum log likelihood ratio across all observed nodes. For hypothesis testing using *T*, the p-value is derived non-parametrically using Monte Carlo simulation. Under the null hypothesis, there is no association between the study exposure and any outcomes. If the null hypothesis is true, each event within a node is expected to occur in relationship to the length of the risk and comparison windows. The conditional version of the test statistic adjusts this probability by accounting for the overall pattern of outcomes in both windows (see equations below). In both versions, when performing Monte Carlo simulations, all nodes must contain the same total number of outcomes per node as observed in the original data. They are assigned to occur across the observation window per the null hypothesis to derive the test statistic used to generate an alert (i.e., a statistically significant departure from the null hypothesis).

$$Equation\ 1a\ (Unconditional)\text{: } LLR = \ln\left(\frac{\left(\frac{c_G}{c_G+n_G}\right)^{c_G}\left(\frac{n_G}{c_G+n_G}\right)^{n_G}}{(\mathrm{p})^{c_G}(1-\mathrm{p})^{n_G}}\right) I\left(\frac{c_G}{c_G+n_G} > \mathrm{p}\right)$$

$$Equation\ 1b\ (Conditional)\text{: } LLR = \ln\left(\frac{\left(\frac{c_G}{c_G+n_G}\right)^{c_G}\left(\frac{n_G}{c_G+n_G}\right)^{n_G}}{\left(\frac{C}{C+N}\right)^{c_G}\left(\frac{N}{C+N}\right)^{n_G}}\right) I\left(\frac{c_G}{c_G+n_G} > \frac{C}{C+N}\right)$$

$$\boldsymbol{Equation\ 2\text{: } T = \max_{G} LLR(G)}$$

Where: T = Bernoulli tree scan statistic

G = node of interest
c_G = number of outcomes in the risk window for a given node G
n_G = number of outcomes in the comparison window for a given node G
p = ratio of the risk window length to the entire observation window length
C = total number of outcomes in the risk window across ALL nodes
N = total number of outcomes in the comparison window across ALL nodes

$I()$ 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.

We will use an alerting threshold of p=0.01. In other words, we will evaluate nodes with log-likelihood ratios that fall into the top 1% of the maximal log-likelihood ratios observed using Monte Carlo. These nodes represent the most extreme departures from the null hypothesis.

2. Information Component Temporal Pattern Discovery

Information Component Temporal Pattern Discovery (ICTPD)[6–9] is another longitudinal signal detection method that can be applied to self-controlled designs. ICTPD uses a shrinkage estimator to reduce the number of false positives due to random variability (and then indirectly the multiple comparisons) or due to very low expected values for rare adverse events or exposures. It does not explicitly control for multiplicity. The information component with shrinkage IC_Δ can be calculated as shown below. The more general formula allows for accounting of censoring between the risk window and control window, but this study will require overall enrollment in both. There is a closed form solution for the credibility interval of the IC_Δ.[9]

$$IC_\Delta = \log_2\left(\frac{y_1 + \frac{1}{2}}{y_0\left(\frac{E_1}{E_0}\right) + \frac{1}{2}}\right)$$

Where: y_1 = number of outcomes in the risk window among study exposure of interest
y_0 = number of outcomes in the comparison window among study exposure of interest
E_1 = number of outcomes in the risk window among the external calibration exposure of interest
E_0 = number of outcomes in the comparison window among the external calibration exposure of interest

ICTPD makes use of an external calibration mechanism to control for global patterns of healthcare utilization that may account for systematic differences between pre- and post- exposure windows. It has a similar intention to the conditional Bernoulli version of the tree scan statistic described above but is implemented differently. The conditional Bernoulli version of the tree-based scan statistic utilizes the pattern of outcome occurrence in the entire MLCCS tree to control for global patterns of healthcare. The ICTPD uses an external exposure group – specifically, all exposures using all dispensing or prescription records – and then performs the adjustment in a node-specific fashion. Using node-specific notation similar to the equations above, the IC_Δ can be re-written below:

$$IC_\Delta = log_2\left(\frac{c_G + \frac{1}{2}}{n_G\left(\frac{C_G}{N_G}\right) + \frac{1}{2}}\right)$$

where: G = node of interest
c_G = number of cases in the risk window for a given node G among the study exposure of interest
n_G = number of cases in the comparison window for a given node G among the study exposure of interest
C_G = total number of cases in the risk window for a given node G among the calibration exposure population
N_G = total number of cases in the comparison window for a given node G among the calibration exposure population

In our study, access to all dispensing records (i.e., a "general" population) is not feasible so we will randomly select 5000 NDC and HCPCS codes and combine their data to create our "calibration exposure."

The traditional alerting threshold for the ICTPD method uses a 2-sided 95% credible interval, denoted $IC_{\Delta 025}$. An alert occurs only when this value is positive, implying that an alert is only raised for an unusual elevation in the frequency of observed outcomes in the risk window. The indication function in the test statistic for TreeScan performs the same function. For this analysis, we will observe the number of alerts when $IC_{\Delta 025} > 0$ and when $IC_{\Delta 01} > 0$.

3. Sequence Symmetry Analysis

Sequence Symmetry Analysis (SSA)[10–12] is a tool used for rapid detection of potential adverse drug events associated with newly marketed medicines utilizing computerized claims data. SSA is based on analyzing the sequences of medications and events; if an event occurs more frequently after exposure to a medication than before, it may be an indication of an adverse effect of the medication. The method uses a simple sequence-based algorithm (pre-post) and is relatively straightforward computationally. In patients with both exposure and outcome during the defined observation window, the crude sequence ratio = # with exposure before outcome/# with outcome before exposure. SSA does not control for multiple hypothesis tests across many outcomes. A prior study ranked SSA alerts based on magnitude of absolute difference in sequence orders and presented unadjusted p-values from chi-square tests.[12] SSA will not be performed with any of the scenarios that include a post-exposure comparison window.

Traditionally, SSA does not utilize a threshold for alerts, and instead examines a pre-specified metric – e.g., the ranked magnitude of absolute difference – across all outcomes. With thousands of outcomes, such an examination is impractical. Therefore, we propose to take the maximum total number of alerts arising from analogous TreeScan or ICTPD analyses and use that value as cut-off point for SSA outcomes of interest. For example, if a particular analytic dataset gives rise to 5 alerts with ICTPD and 7 alerts with TreeScan, then we will look at the top 7 magnitudes of absolute difference.

G. ALERT ASSESSMENT FRAMEWORK

This project is intended to be a methods evaluation rather than a regulatory safety analysis of these two anti-seizure medications. For each of the analytic datasets (i.e., scenarios) considered, alerts will be defined according to the criteria described above per method. Generally, we proposed to use the method as it has been used in practice with slight modifications.

At the most basic level, alerts will be triaged as known or expected based on the study exposure's label, known safety profile or temporal administration pattern. We will also record nodes that alert in any given method across several analytic scenarios (e.g., alerts that are robust to the choice of comparison windows).

For the empiric assessment, we will record the total alert load, the alignment with the known or expected alerts, and the influence of the various study design choices examined. For the simulation assessment, we will record the ability to detect pre-specified, investigator-inserted risks.

H. CHALLENGES TO INTERNAL AND EXTERNAL VALIDITY BASED ON STUDY DESIGN

1. Time-Varying Confounding Resulting from Changes in Likelihood of Exposure

When using a pre-exposure comparison window, it is possible that experiencing an adverse event prior to the potential study drug exposure alters the patient's likelihood of being exposed to the study drug of interest. Examining multiple comparison windows may allow further examination of this concern.

2. Limitations on Generalizability as a Results of Survival Requirements

These analyses require equal opportunities to experience the outcome of interest across the observation window. That requirement is operationalized by requiring patient enrollment throughout the observation window or follow-up period. Patients will need to survive these periods in order to contribute to the analysis. This assumption means that mortality is not an outcome of interest and that outcomes that may be associated with acute mortality (e.g., sudden cardiac death) are likely to be underestimated in terms of alerts that appear in the analysis.

3. Concomitant Exposures that are Time-Dependent

Individuals in the cohort may be concomitantly exposed to other medications on the same day as the incident drug dispensing, and these other medications may represent a time-dependent variable if they are not constant throughout the observation window. A statistical alert that appears may be associated with one or more of these concomitant exposures instead.

4. Concomitant Condition Evaluated at the Medical Visit

We will exclude Day 0 in the analysis to prevent capturing conditions that were present at the time of exposure. However, it is still possible that an individual has evidence of an incident exposure that co-occurs with evaluation of unrelated symptoms or medical conditions. Follow-up evaluation or treatment for the unrelated condition may then appear to be associated with the study drug.

5. Possible Exposure Misclassifications Due to Treatment of the Incident Study Dispensing as a Point Exposure

The incident dispensing of the study drug is treated as a point exposure without regard to the days supply. Therefore, the post-exposure comparison window will include people that are a) no longer actively being treated with the study drug and b) are actively being treated with the study drug. Thus, alerts may be underestimated for outcomes that sustained elevations of risk or longer induction periods.

IV. LIMITATIONS

There are limitations of this evaluation, which are either inherent to secondary-use observational data, or the nature of data-mining.

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.[19]

Second, we chose to consider observation windows in the months surrounding exposure, implying that we cannot detect drug-associated outcomes that occur several years after exposure. As discussed earlier in the protocol, a self-controlled design requires an equal opportunity to experience the outcomes

across the observation window had there been no exposure at all. This assumption becomes very difficult to maintain over long periods of observation.

Third, we chose a point exposure analysis as a simplifying measure in this data-mining study. Thus, a 1-day incident exposure is dealt with in exactly the same way as an incident exposure that might last 90 days.

Fourth, the analysis is performed using a single tree curated by clinicians that uses ICD-9-CM codes primarily. This method can be used with other trees, either singularly or multiple trees simultaneously. While we expect that most trees developed with clinical expertise will generate similar results, some trees could potentially miss alerts generated by other trees.

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

V. SUMMARY

Signal identification has traditionally been strongly driven by spontaneous reports, which lack population data to provide context. While these reports are the backbone of post-market surveillance, there is more that can be done to characterize the general safety profile of thousands of outcomes following exposures of interest. Data-mining, and the use of a self-controlled risk interval design to control for time-invariant confounding, may generate a complementary stream of new information on these exposures. Our study seeks to illuminate the strengths and weaknesses of various approaches in both an empirical and simulated data setting.

VI. ACKNOWLEDGEMENTS

We would like to acknowledge the participation of our fellow workgroup member: Rima Izem, PhD.

VII. APPENDIX – CODES FOR EXPOSURES OF INTEREST

Table 3. List of Generic and Brand Names Used to Define Exposures of Interest

Generic Name	Brand Name	Route
lamotrigine	lamotrigine	oral
lamotrigine	Lamictal	oral
lamotrigine	Lamictal Starter (Green) Kit	oral
lamotrigine	Lamictal Starter (Orange) Kit	oral
lamotrigine	Lamictal Starter (Blue) Kit	oral
lamotrigine	Lamictal XR	oral
lamotrigine	Lamictal XR Starter (Blue)	oral
lamotrigine	Lamictal XR Starter (Green)	oral
lamotrigine	Lamictal XR Starter (Orange)	oral
lamotrigine	Lamictal ODT	oral
lamotrigine	Lamictal ODT Starter (Orange)	oral
lamotrigine	Lamictal ODT Starter (Blue)	oral
lamotrigine	Lamictal ODT Starter (Green)	oral
lamotrigine	Subvenite	oral
lamotrigine	Subvenite Starter (Orange) Kit	oral
lamotrigine	Subvenite Starter (Blue) Kit	oral
lamotrigine	Subvenite Starter (Green) Kit	oral
levetiracetam	levetiracetam	oral
levetiracetam	Keppra	oral
levetiracetam	Keppra XR	oral
levetiracetam	Spritam	oral
levetiracetam	Roweepra	oral
levetiracetam	Roweepra XR	oral

VIII. 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. Greene, S. K. *et al.* Near Real-Time Surveillance for Influenza Vaccine Safety: Proof-of-Concept in the Vaccine Safety Datalink Project. *Am. J. Epidemiol.* **171**, 177–188 (2010).
3. Greene, S. K. *et al.* Risk of Confirmed Guillain-Barré Syndrome Following Receipt of Monovalent Inactivated Influenza A (H1N1) and Seasonal Influenza Vaccines in the Vaccine Safety Datalink Project, 2009–2010. *Am. J. Epidemiol.* **175**, 1100–1109 (2012).
4. Kulldorff, M., Fang, Z. & Walsh, S. J. A tree-based scan statistic for database disease surveillance. *Biometrics* **59**, 323–331 (2003).
5. Kulldorff, M. *et al.* Drug safety data mining with a tree-based scan statistic. *Pharmacoepidemiol Drug Saf* **22**, 517–523 (2013).
6. Hauben, M. *et al.* Signal Detection using Temporal Pattern Discovery in Electronic Health Records (EHRs) - Lessons from Statins and Rhabdomyolysis. in (2016).
7. Norén, G. N. *et al.* Empirical performance of the calibrated self-controlled cohort analysis within temporal pattern discovery: lessons for developing a risk identification and analysis system. *Drug Saf* **36 Suppl 1**, S107-121 (2013).
8. Norén, G. N., Hopstadius, J. & Bate, A. Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. *Stat Methods Med Res* **22**, 57–69 (2013).
9. Norén, N., Hopstadius, J., Bate, A., Star, K. & Edwards, I. Temporal pattern discovery in longitudinal electronic patient records. *Data Min. Knowl. Discov.* **20**, 361–387 (2010).
10. Wahab, I. A., Pratt, N. L., Ellett, L. K. & Roughead, E. E. Sequence Symmetry Analysis as a Signal Detection Tool for Potential Heart Failure Adverse Events in an Administrative Claims Database. *Drug Saf* **39**, 347–354 (2016).
11. Wahab, I. A., Pratt, N. L., Wiese, M. D., Kalisch, L. M. & Roughead, E. E. The validity of sequence symmetry analysis (SSA) for adverse drug reaction signal detection. *Pharmacoepidemiol Drug Saf* **22**, 496–502 (2013).
12. Hallas, J. *et al.* Hypothesis-free screening of large administrative databases for unsuspected drug-outcome associations. *Eur. J. Epidemiol.* (2018). doi:10.1007/s10654-018-0386-8
13. Greenland, S. & Robins, J. M. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol* **15**, 413–419 (1986).
14. Mittleman, M. A. & Mostofsky, E. Exchangeability in the case-crossover design. *Int J Epidemiol* **43**, 1645–1655 (2014).
15. Wang, S. V. *et al.* Data mining for adverse drug events with a propensity score matched tree-based scan statistic. *Epidemiology* (2018). doi:10.1097/EDE.0000000000000907
16. Maro, J. C., Dashevsky, I. & Kulldorff, M. Postlicensure medical product safety data-mining: Power calculations for Bernoulli data. *Sentinel Initiative* (2017). Available at: https://www.sentinelinitiative.org/sites/default/files/PRISM/TreeScanPower_FinalReport.pdf. (Accessed: 29th December 2017)

17. Yih, W. K. *et al.* Assessment of Quadrivalent Human Papillomavirus Vaccine Safety Using the Self-Controlled Tree-Temporal Scan Statistic Signal-Detection Method in the Sentinel System. *Am J Epidemiol* doi:10.1093/aje/kwy023

18. Li, R. *et al.* Meningococcal conjugate vaccine safety surveillance in the Vaccine Safety Datalink using a tree-temporal scan data mining method. *Pharmacoepidemiol Drug Saf* (2018). doi:10.1002/pds.4397

19. 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).