Drug Saf (2013) 36 (Suppl 1):S107–S121
DOI 10.1007/s40264-013-0095-x

ORIGINAL RESEARCH ARTICLE

Empirical Performance of the Calibrated Self-Controlled Cohort Analysis Within Temporal Pattern Discovery: Lessons for Developing a Risk Identification and Analysis System

G. Niklas Norén · Tomas Bergvall · Patrick B. Ryan · Kristina Juhlin · Martijn J. Schuemie · David Madigan

© Springer International Publishing Switzerland 2013

Abstract

Background Observational healthcare data offer the potential to identify adverse drug reactions that may be missed by spontaneous reporting. The self-controlled cohort analysis within the Temporal Pattern Discovery framework compares the observed-to-expected ratio of medical outcomes during post-exposure surveillance periods with those during a set of distinct pre-exposure control periods in the same patients. It utilizes an external control group to account for systematic differences between the different time periods, thus combining within- and between-patient confounder adjustment in a single measure.

Objectives To evaluate the performance of the calibrated self-controlled cohort analysis within Temporal Pattern Discovery as a tool for risk identification in observational healthcare data.

Research Design Different implementations of the calibrated self-controlled cohort analysis were applied to 399 drug-outcome pairs (165 positive and 234 negative test cases across 4 health outcomes of interest) in 5 real observational databases (four with administrative claims and one with electronic health records).

Measures Performance was evaluated on real data through sensitivity/specificity, the area under receiver operator characteristics curve (AUC), and bias.

Results The calibrated self-controlled cohort analysis achieved good predictive accuracy across the outcomes and databases under study. The optimal design based on this reference set uses a 360 days surveillance period and a single control period 180 days prior to new prescriptions. It achieved an average AUC of 0.75 and AUC >0.70 in all

The OMOP research used data from Truven Health Analytics (formerly the Health Business of Thomson Reuters), and includes MarketScan® Research Databases, represented with MarketScan Lab Supplemental (MSLR, 1.2 m persons), MarketScan Medicare Supplemental Beneficiaries (MDCR, 4.6 m persons), MarketScan Multi-State Medicaid (MDCD, 10.8 m persons), MarketScan Commercial Claims and Encounters (CCAE, 46.5 m persons). Data also provided by Quintiles® Practice Research Database (formerly General Electric's Electronic Health Record, 11.2 m persons) database. GE is an electronic health record database while the other four databases contain administrative claims data.

G. N. Norén (✉) · T. Bergvall · K. Juhlin
Uppsala Monitoring Centre, WHO Collaborating Centre for International Drug Monitoring, Uppsala, Sweden
e-mail: niklas.noren@who-umc.org

G. N. Norén
Department of Mathematics, Stockholm University, Stockholm, Sweden

P. B. Ryan
Janssen Research and Development LLC, Titusville, NJ, USA

M. J. Schuemie
Department of Medical Informatics, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

D. Madigan
Department of Statistics, Columbia University, New York, NY, USA

P. B. Ryan · M. J. Schuemie · D. Madigan
Observational Medical Outcomes Partnership, Foundation for the National Institutes of Health, Bethesda, MD, USA

but one scenario. A design with three separate control periods performed better for the electronic health records database and for acute renal failure across all data sets. The estimates for negative test cases were generally unbiased, but a minor negative bias of up to 0.2 on the RR-scale was observed with the configurations using multiple control periods, for acute liver injury and upper gastrointestinal bleeding.

Conclusions The calibrated self-controlled cohort analysis within Temporal Pattern Discovery shows promise as a tool for risk identification; it performs well at discriminating positive from negative test cases. The optimal parameter configuration may vary with the data set and medical outcome of interest.

1 Background

Medicinal products must be monitored throughout their lifecycle for adverse drug reactions and other threats to patient safety. When new medicines are made available to patients for routine use, their safety profiles are only known in part. Pre-marketing clinical trials feature randomization and careful experimental design but are limited in size (so lack power to detect rare reactions), are restricted to patients without certain co-morbidities and co-medications, and reflect a carefully controlled setting that corresponds poorly to real-world clinical practice.

The aim of post-marketing surveillance is to better determine the balance between desired and undesired effects of medicines and specifically to highlight previously unknown risks to patients. Currently, safety signal detection is largely based on individual case reports of suspected harm to patients from medicines (otherwise known as spontaneous reports). This system relies on the active analysis of data in reports by health professionals and patients. Such reports ideally contain just the right information to assess the likelihood of a causal effect in a specific case [1]. They are particularly effective for identifying and describing adverse drug reactions with characteristic temporal relationships to exposure, with low background rates, or that tend to resolve soon after termination of drug treatment [2, 3]. They are less good for adverse events with high background rates in the population or that have long latency [4]. A constant challenge is the variable amount of information on these reports, and the multitude of factors that affect what is reported as a suspected adverse reaction. Longitudinal observational databases are important sources of information for evaluation of suspected causal associations between medicines and adverse events in pharmacoepidemiology. Though they are currently not used for initial safety signal detection, there has been increased interest over recent years in their use. As an example, the Food and Drug Administration (FDA) Amendment Act passed by U.S. Congress in 2007 called for the establishment of a post-marketing risk identification and analysis system with access to observational data from 100 million patients by 2012 [5]. One of the challenges is to develop statistical methods that can identify suspected adverse drug reactions or other safety signals related to drug therapy in longitudinal data [6].

Temporal Pattern Discovery is an analytical framework for exploratory analysis of longitudinal observational data for iterative and interactive use [7, 8]. Its methodology covers the identification of temporal associations between discrete events in longitudinal data, statistical graphical visualization of temporal association patterns, and the combination of summary statistics with patient-level data; the focus of this study is the identification of temporal associations between incident prescriptions of drugs and registration of medical outcomes. In Temporal Pattern Discovery, this is achieved through a self-controlled cohort analysis contrasting post-exposure surveillance periods to a set of pre-exposure control periods. The analysis also incorporates calibration relative to an external control group of incident prescriptions of other drugs; this is to adjust for systematic variability in the registration of medical events in different time periods, such as the tendency of doctor visits and recorded medical events to cluster in time. The idea is analogous to the case–time–control design [9], which extends the (self-controlled) case crossover design with adjustment for systematic variation between time periods in the rate of registration of the outcome, using external controls to do so.

Any method proposed for broad risk identification of a wide variety of drugs and medical events without adaptation to each specific issue will be imperfect. Given the current interest in screening observational data for emerging safety signals, an important challenge is to determine the operational performance of different methods, and their areas of strength and weakness. The Observational Medical Outcomes Partnership (OMOP) is a collaborative undertaking aiming to characterize the operational performance of all available risk identification methods against a broad range of longitudinal observational databases. OMOP's first empirical evaluation showed an operational performance of the calibrated self-controlled cohort analysis across eight data sets that is moderate in absolute terms, but strong relative another eight methods evaluated in the same study [10, 11]. At the same time, evaluation against simulated data have highlighted that the use of all other medicines as the external control group can lead to failure to properly identify potential confounding by underlying disease, [12] with resulting possible false positives.

The aim of the study at hand is to evaluate the predictive ability of the calibrated self-controlled cohort analysis across five observational databases, for a reference set of

165 true positive associations and 234 considered to be true negative test cases. The following four medical outcomes are covered by the reference set: acute liver injury, acute myocardial infarction, acute kidney injury, and upper gastrointestinal bleeding [13]. A secondary aim is to identify which parameter configurations of the methodology perform best in different scenarios.

2 Methods

The capability to reduce or eliminate confounding is a major aim of epidemiological design and analysis. A common approach is to identify 'equivalent' unexposed patients against which to compare the rate in the exposed, as outlined in Fig. 1. By design, one may utilize an active comparator treatment, and in the analysis one may adjust for suspected confounders through stratification or regression. Propensity scores have proven to be a valuable component of such approaches in allowing a large number of covariates to be projected onto a single dimension thereby simplifying design and analysis [14]. A main limitation with the use of external comparators is the difficulty of knowing whether the control patients are indeed at equivalent risk of the outcome as the exposed except for the attributable risk from the treatment; even when the active comparator is matched on indication for treatment, disease severity may be unknown and variable.

As an alternative, self-controlled designs have recently been proposed and utilized in certain settings. They attempt to identify 'equivalent' periods of unexposed time within the same patients, against which to compare the same patients' exposed time, as outlined in Fig. 2. In principle, this has the potential to eliminate time-constant confounding. Self-controlled designs require exchangeability between exposed and unexposed time periods, which has restricted their use in epidemiology to specific scenarios, such as those with intermittent exposure and acute onset of the outcome.

The calibrated self-controlled cohort analysis proposed here combines these two approaches to confounder adjustment. Its primary adjustment is within the cohort through comparison to separate control periods prior to treatment in the same group of patients. As a complement, it utilizes calibration by an external control group to adjust for systematic differences between these time periods, as outlined in Fig. 3. It is similar in spirit to the case-time control design, in its combination of within- and between-patient confounder adjustment.

2.1 Calibrated Self-Controlled Cohort Analysis

Temporal Pattern Discovery uses a calibrated self-controlled cohort analysis to identify temporal associations between medical events and first prescriptions of medicines for further review. The primary comparison is *within* the cohort of patients exposed to the medicine of interest: one or more post-exposure surveillance periods is contrasted to a set of pre-exposure control periods as shown in Fig. 4.

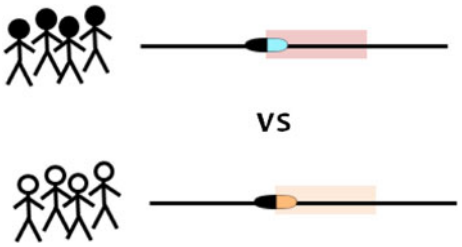

Fig. 1 Confounder adjustment through comparison to an external control group of patients exposed to another drug; important covariates are matched on or adjusted for in the analysis, but the approach is sensitive to residual systematic differences between the exposed and comparator groups

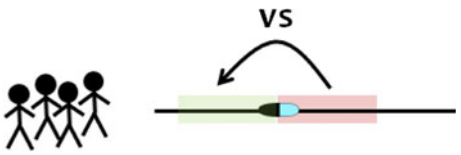

Fig. 2 Confounder adjustment through comparison to unexposed time in the same patients; time-constant covariates are automatically adjusted for, but the approach is sensitive to residual systematic differences between the surveillance and control periods

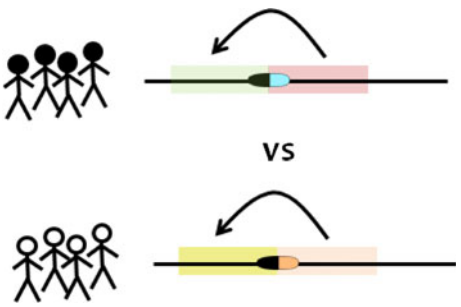

Fig. 3 Confounder adjustment through comparison to unexposed time in the same patients, with calibration for systematic differences between time periods in the external control group; this approach combines the two modes of confounder adjustment, but is still sensitive to systematic differences between the surveillance and control periods *unique to the exposed*, for example from risk minimization or protopathic bias

Assume a single surveillance period indexed 1 and a single control period indexed 0. The proportion of patients who have been exposed to the drug of interest with the medical event registered at least once in the surveillance period (y_1/n_1) is contrasted with the corresponding proportion in a control period (y_0/n_0). y_t denotes the number of patients with at least one event of interest in period t, and n_t denotes the number of patients followed up in period t. Due

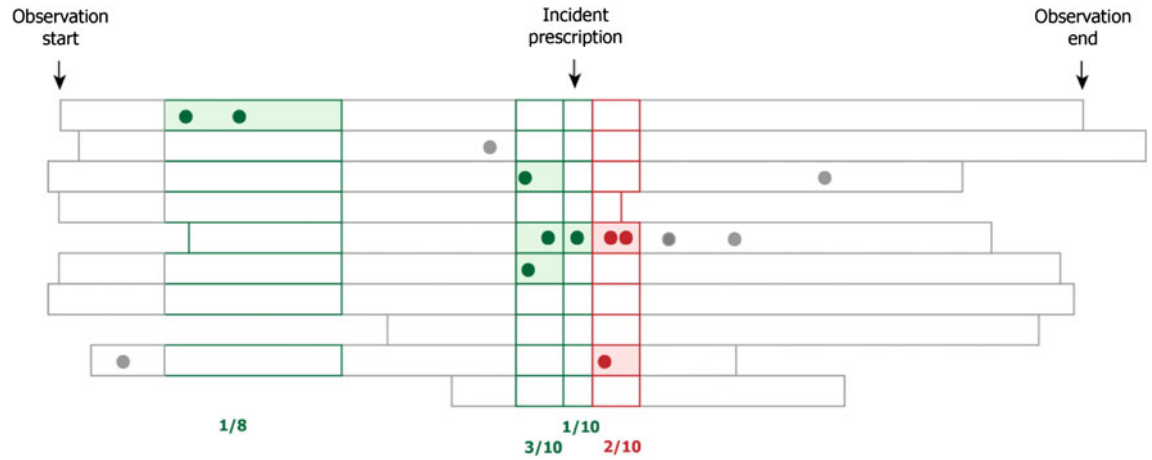

Fig. 4 Self-controlled cohort design with three separate control periods, and 30 days surveillance period, in the spirit of the configuration described in Norén et al. [7]. *Each band* represents a patient with a corresponding observation period. *Solid circles* represent registrations of the outcome of interest and the *arrow marks* incident first prescription. The surveillance period is marked in *red* and the three simultaneous control periods in *green*

to differential censoring, n_1 can differ from n_0, but will usually be of similar magnitude. To account for general differences in the ascertainment of the medical event in the surveillance and control periods, we calibrate with the corresponding proportions for an external control group of first prescriptions of other medicines. Let Y_1/N_1 denote the proportion of first prescriptions in the external control group with the event of interest in the surveillance period and let Y_0/N_0 denote the corresponding proportion for the control period. Again, N_1 can differ from N_0 due to censoring.

A simple observed-to-expected ratio for the surveillance period is:

$$\mathrm{OE}_1 = \frac{y_1}{n_1 \cdot Y_1/N_1} = \frac{y_1}{E_1}$$

This is a normalized relative risk. It indicates whether the event is more likely to occur soon after new prescriptions of the medicine of interest compared to soon after first prescription in general. We do not propose to use this measure for screening. It is a naïve measure of association in that its expected count E_1 accounts for none of the systematic variability between patients taking the medicine of interest and patients in the external control group. Thus, it does not distinguish true temporal associations from general differences in susceptibility to the event. However, this simple observed-to-expected ratio provides a basis for the measure of temporal association described in more detail below. Observed-to-expected ratios such as this are also the basis for the visualization of temporal patterns in chronographs (Figs. 8, 9), which is the core of the Temporal Pattern Discovery framework.

A corresponding observed-to-expected ratio for the control period is:

$$\mathrm{OE}_0 = \frac{y_0}{n_0 \cdot Y_0/N_0} = \frac{y_0}{E_0}$$

As our basis for identifying temporal associations in longitudinal observational databases, we use the ratio between the two observed-to-expected ratios above:

$$\mathrm{OE}_\Delta = \frac{\mathrm{OE}_1}{\mathrm{OE}_0} = \frac{y_1}{y_0 \cdot E_1/E_0}$$

This provides a measure of association that is primarily driven by the contrast between different time periods within the cohort of exposed patients in the rate of the event (y_1 versus y_0). It is corrected for tendencies of the event to be registered more often in either time period for patients in general, as measured by E_1/E_0. This is intended to eliminate the influence of differences in period length as well as the reliability with which events are ascertained in different time periods (for example due to variation in the frequency of patients' use of healthcare). It also goes beyond the naïve estimate in that between-patient differences that are fixed over the time periods studied, are eliminated.

The calibrated self-controlled cohort analysis has similarities with the self-control cohort design implemented in observational screening, [15] and to other self-controlled designs such as sequence symmetry analysis, [16] the self-controlled case series, [17], case-crossover analysis [18], case-time control analysis, [9] and prior event rate ratio analysis [19]. A summary of some similarities and distinctions is provided in Table 1.

2.2 Simple Shrinkage and Credibility Intervals

Observed-to-expected ratios are volatile for rare events when the observed and/or expected counts are near zero.

Table 1 Summary of similarities and distinctions of self-controlled designs

Method	Design analogy	Matching	Required data	Measure	Miscellaneous
Calibrated self-controlled cohort analysis (Temporal Pattern Discovery)	Cohort	Cohort	All exposed patients and a set of controls	RR	Simultaneous control and surveillance periods Calibration for systematic differences between surveillance and control periods
Self-controlled cohort analysis (observational screening)	Cohort	Cohort	All exposed patients	IRR	
Self-controlled case series (as implemented in OMOP)	Cohort	Individual	Cases with exposure	IRR	Covariate adjustment through regression
Prior event rate ratio	Cohort	Cohort	All exposed patients and a set of controls	IRR	Calibration for systematic differences between surveillance and control period
Sequence symmetry analysis	None[a]	Individual[a]	Cases with exposure	IRR	Adjustment for time trends in exposure and event
Case-cross over	Case–control	Individual	Cases with exposure	RR	
Case–time–control	Case–control	Individual	Cases and controls with exposure	RR	Adjustment for time trends in exposure

RR relative risk, *IRR* incidence rate ratio

[a] Sequence symmetry analysis treats both exposures and outcomes as random and uses a binomial test for their order; it does not explicitly match different time periods

For a more stable measure, we apply a simple shrinkage transformation that moderates the ratio towards one by adding a constant to each of the nominator and the denominator.

$$\mathrm{OE}'_{\Delta} = \frac{y_1 + 1/2}{y_0 \cdot E_1/E_0 + 1/2}$$

The base 2 logarithm of this shrunk observed-to-expected ratio is referred to as IC_{Δ} and is our basis for screening:

$$\mathrm{IC}_{\Delta} = \log_2 \frac{y_1 + 1/2}{y_0 \cdot E_1/E_0 + 1/2}$$

Its sign indicates the direction of association and its magnitude measures the strength of positive and negative associations on a common scale. Formally, OE'_{Δ} can be viewed as the Bayesian posterior mean of a parameter μ under the assumption that y_1 is Poisson $\mathrm{Po}(\mu \cdot y_0 \cdot E_1/E_0)$-distributed with a Gamma G(1/2,1/2) prior distribution for μ. A range of likely values as represented by the Bayesian credibility interval can be computed as the inverse of the Gamma $\mathrm{G}(y_1 + 1/2,\ y_0 \cdot E_1/E_0 + 1/2)$ cumulative distribution function for appropriate percentiles. The base 2 logarithm of the lower 95 % credibility interval limit for OE'_{Δ} is referred to as $\mathrm{IC}_{\Delta 025}$. A common threshold to identify temporal associations in Temporal Pattern Discovery is [7]

$$\mathrm{IC}_{\Delta 025} > 0$$

2.3 Surveillance Periods

Confirmatory studies are designed so that the surveillance period matches the expected time-at-risk of the medical outcome as an adverse reaction to the treatment of interest. For broad surveillance, there is less opportunity for customization, and generic choices must be made. Table 2 lists the surveillance periods available in the open-source implementation of Temporal Pattern Discovery available at http://omop.org/MethodsLibrary, but in principle any fixed length time periods can be used. When simultaneous surveillance periods are used, $\mathrm{IC}_{\Delta 025}$ is maximized across these, in an attempt to ensure sensitivity to different risk scenarios.

In addition to censoring both control and surveillance periods at the beginning and end of observation period for the patient, one can censor the surveillance period at the estimated end of treatment with the medicine of interest. With censoring of the surveillance period at end of treatment, only those patients for which the outcome occurs within the treatment duration (and within the surveillance period) are counted, and expected counts are adjusted accordingly. The advantage is an increased probability that patients are actually exposed to the medicine at the time of the adverse event. The disadvantage is the potential inaccuracy of the duration of treatment estimation that leads to failure to capture some exposed cases.

2.4 Control Periods

Temporal Pattern Discovery can incorporate a set of distinct control periods, requiring the observed-to-expected ratio in the surveillance period(s) to exceed those in each of the control periods. The motivation for considering simultaneous control periods separately instead of grouping together all unexposed time is that there may be variation

Table 2 Parameterization of six analytical design choices

Design choice	Considerations	Options
Surveillance periods	In what periods relative to start of exposure shall medical events be ascertained as potential adverse reactions? Surveillance periods are used in parallel and selected independently of one another; at least one must be used	1–30 days 31–90 days 91–180 days 1–360 days 721–1080 days
Censor at end of treatment	Shall the surveillance period be censored at the end of treatment with the drug of interest?	True/false
Primary control period	What period prior to the start of exposure shall be used as the primary control period in the self-controlled cohort analysis?	Customizable
Secondary control period	Shall the 30 days immediately prior to start of exposure be used as a secondary control period?	True/false
Tertiary control period	Shall the day of the start of exposure be used as a tertiary control period?	True/false
External control group	Shall the external control group include all medicines, or be restricted to other medicines with the same indication?	All / Same indication

in the rate of the outcome over the unexposed patient time. In the case of confounding by underlying disease, the most robust control period might be immediately prior to the first prescription since it best matches the health state of the patient just after initiation of treatment. As an example, the rate of suicidal ideation in a depressed patient can be expected to be elevated around the time of the first prescription of an antidepressant, whereas a control period far before the first antidepressant prescription will not capture this. In the case of a contraindication, a control period long before the first prescription may on the other hand be preferred in order to reduce distortion from influence of the event on future prescriptions. As an example, upper GI bleeding may be relatively infrequent just before omeprazole prescriptions, thanks to attempts at risk minimization, rendering the month just before new prescriptions an unrepresentative control period. A control period positioned some time after the end of exposure may be preferred in order to eliminate distortion by reversed causality, but may be subject to carry-over effects and has not been evaluated in the study reported here. Self-controlled case series analysis often uses pre- and post-exposure control periods together, but the calibrated self-controlled cohort design would allow for a pure post-exposure control period to be used in parallel with the set of pre-exposure control periods.

With simultaneous control periods, $IC_{\Delta 025}$ is computed separately for each control period, and the minimal value selected (or equivalently, $IC_{\Delta 025}$ is computed based on the control period with maximal OE_0). The open-source implementation of Temporal Pattern Discovery available at http://omop.org/MethodsLibrary allows for a primary control period with arbitrary length and position combined with the option to include a secondary control period 30–1 days prior to first prescription and/or a tertiary control period on the day of the prescription, but in principle any set of control periods is possible.

Figure 4 illustrates four different control period configurations: two with a set of three simultaneous control periods in the spirit of the implementation described earlier [7, 8] and two with a single control period 180 days prior to first exposure identified as optimal in the OMOP 2010 evaluation [10, 11].

2.5 External Control Group

A basic implementation of Temporal Pattern Discovery uses first prescriptions of any medicine as its external control group. This is computationally efficient but analytically naïve. As an alternative, the external control group can be restricted to prescriptions of other drugs with the same indication, approximating an active comparator better matched to the treated patients. One benefit of this is reduced (though not necessarily eliminated) negative influence of time-varying confounding by the underlying disease. Another benefit is that a more narrow control group is less likely to be influenced by outlier drug-outcome pairs with strong temporal associations that can distort the expected value [11, 20]. For example, the inclusion of acetylsalicylic acid in the external control group for myocardial infarction may inflate the expected registration of myocardial infarction in the month prior to prescription, because of the tendency to prescribe acetylsalicylic acid to patients who have just suffered a myocardial infarction.

A restriction of the external control group to drugs with the same indication on the other hand increases computational complexity and may yield unstable or

Table 3 Overview of design choices evaluated in the experiment

Design choice	Explanation	Options
Surveillance period	In what period relative to start of exposure shall medical events be ascertained as potential adverse reactions	1–30 days 1–60 days 1–360 days
Control period	What period prior to the start of exposure shall be used as the primary control period in the self-controlled cohort analysis	30–1 days 180–1 days 810–361 days 1080–361 days
Secondary control	Shall the 30 days immediately prior to start of exposure be used as a secondary control period	True/false
Tertiary control	Shall the day of the start of exposure be used as a tertiary control period?	True/false

unrepresentative expected counts. Moreover, drugs with multiple possible indications are challenging to manage.

2.6 Implementation

Within the open-source software implementation of the design entitled 'IC Temporal Pattern Discovery', publicly available at http://omop.org/MethodsLibrary, six analytical design choices are parameterized, as shown in Table 2.

2.7 Evaluation Against Real-World Observational Data

Temporal Pattern Discovery was executed against 399 drug-outcome pairs described above. The criteria for the 165 positive test cases were that the corresponding Medical Dictionary for Regulatory Activities (MedDRA) Preferred Term or one of its Lower-Level Terms be listed in the Boxed warning or Warnings/Precautions section of the active U.S. FDA Structured Product Label for the drug, that the medicine be listed as 'causative agent' for the event in Tisdale et al. [21] and that a broad literature review identified no powered studies refuting the effect. The criteria for its 234 negative test cases were that the event not be listed anywhere in U.S. FDA Structured Product Label for the medicine, that the drug not be listed as 'causative agent' in Tisdale et al. [21] and that literature review identified no powered study showing an effect. The full reference set and its construction are described elsewhere [13]. In the evaluation reported here, the broadest definition of each outcome was used and available in the results set publicly available at http://75.101.131.161/download/loadfile.php?docname=OMOP%202011-2012%20Experiment%20Method%20Results. The definitions are denoted *OMOP Acute Myocardial Infarction 1*, *OMOP Acute Renal Failure 1*, *OMOP Acute Liver Failure 1*, and *OMOP Upper GI Ulcer Hospitalization 3*.

Four of the analytical design choices in Table 2 were implemented at the time that this study was initiated and are included in its evaluation, as specified in Table 3. Parallel surveillance periods, censoring at end of treatment and restriction to a more narrow external control group were not considered. A total of 42 unique configurations were considered (six configurations with control period equal to 30–1 days and tertiary control TRUE are redundant, and were excluded on this basis).

The study was conducted against five observational healthcare databases to allow evaluation of performance across different populations and data capture processes: MarketScan® Lab Supplemental (MSLR, 1.2 million persons), MarketScan® Medicare Supplemental Beneficiaries (MDCR, 4.6 million persons), MarketScan® Multi-State Medicaid (MDCD, 10.8 million persons), MarketScan® Commercial Claims and Encounters (CCAE, 46.5 million persons), and the General Electric Centricity™ (GE, 11.2 million persons) database. GE is an electronic health record (EHR) database, whereas the other four databases contain administrative claims data. For every database, the analysis was restricted to those drug-outcome pairs with power greater than or equal to 80 % to detect a relative risk of 1.25, based on the age-by-gender-stratified drug and outcome prevalence estimates [22].

For each database-outcome scenario and parameter configuration, an IC_{Δ} measure of temporal association with associated lower percentile $IC_{\Delta 025}$ was computed. To gain insight into the ability of the method to distinguish between positive and negative test cases, sensitivity and specificity were computed for the fixed threshold $IC_{\Delta 025} > 0$ with different parameter configurations (corresponding to a relative risk exceeding 1). As a complement, the area under the receiver operating characteristics (ROC) curve (AUC) [23] was computed for each database-outcome scenario. It measures predictive accuracy based on the sensitivity and specificity for varying thresholds $IC_{\Delta 025} > T$. The AUC indicates the probability with which a randomly selected positive test case is ranked higher than a randomly selected negative test case. An AUC of 1 reflects perfection whereas an AUC of 0.5 is equivalent to random ordering. An overall AUC was computed as the average of the AUC's for each

Table 4 The four representative configurations evaluated in detail in this study

Configuration	Control periods	Surveillance period (days)	Symbol[a]
OMOP360	180–1 days	1–360	▲
OMOP30	180–1 days	1–30	Δ
DMKD360	1080–361 days Secondary and tertiary	1–360	■
DMKD30	1080–361 days Secondary and tertiary	1–30	□

[a] Symbols used to denote the four parameter configurations in Figs. 6 and 7

database-outcome scenario. The estimates and associated standard errors for all of the analyses are available for download at: http://omop.org/Research.

To appraise to what extent the IC_{Δ} measure is subject to bias, we computed IC_{Δ} for all properly powered negative test cases and considered the empirical distributions of the IC_{Δ} measures under the assumption that the true relative risk of negative test cases be 1.

3 Results

In our analysis we focus especially on the four parameter configurations that are listed in Table 4, visualized in Fig. 5, and are marked with squares and triangles in Figs. 6 and 7. OMOP360 refers to a configuration that defines the surveillance period as 360 days from exposure start, and uses the 180 days immediately prior to incident prescriptions as its single control period; it achieved the highest average AUC in this evaluation. OMOP30 is a similar configuration differing only in its choice of a 30 day surveillance period; it was identified as the optimal setting in a random effects meta-analysis across eight different databases of the OMOP 2010 test cases [10, 11]. Configurations DMKD360 and DMKD30 are similar in spirit to the analysis proposed and evaluated in earlier research on Temporal Pattern Discovery [7, 8] with 30 and 360 days surveillance periods, respectively.

3.1 Predictive Accuracy Across Parameter Configurations

Figure 6 displays the operationally critical sensitivity and specificity of each parameter configuration at the natural threshold of RR = 1, implemented as $IC_{\Delta 025} > 0$. There is considerable spread in both specificity and sensitivity across parameter configurations within each single outcome-database scenario. For the four configurations that we focus on, shorter surveillance periods improve specificity at the expense of sensitivity, and so does the use of three simultaneous control periods compared to the use of a single control period. The average sensitivity and specificity of OMOP360 were 0.77 and 0.52 compared to 0.57 and 0.74 for DMKD360. For the two configurations with 30 days surveillance period, the average sensitivity and specificity were 0.46 and 0.76 for OMOP30 versus 0.25 and 0.88 for DMKD30.

Figure 7 shows predictive accuracy as measured by AUC for all Temporal Pattern Discovery configurations across the 4 outcomes and 5 databases. The OMOP360 configuration achieves the highest average AUC of 0.75 of all 42 configurations evaluated. Its AUC exceeds 0.70 in all but one scenario. Figure 7 shows that it outperforms OMOP30 with the 30 days surveillance period, for all outcome-database scenarios except acute renal failure and upper GI bleeding in the GE database. DMKD360 on the other hand achieves higher AUC than OMOP360 for all outcomes in the GE database and for acute renal failure across all databases, but lower AUC for the other three outcomes across the other four databases. It uses the same 360 days surveillance period but employs three separate and simultaneous control periods. Configuration DMKD30 with 30 day surveillance period and three distinct control periods achieves high specificity but poor sensitivity and relatively low AUC, with the exception of in the GE database.

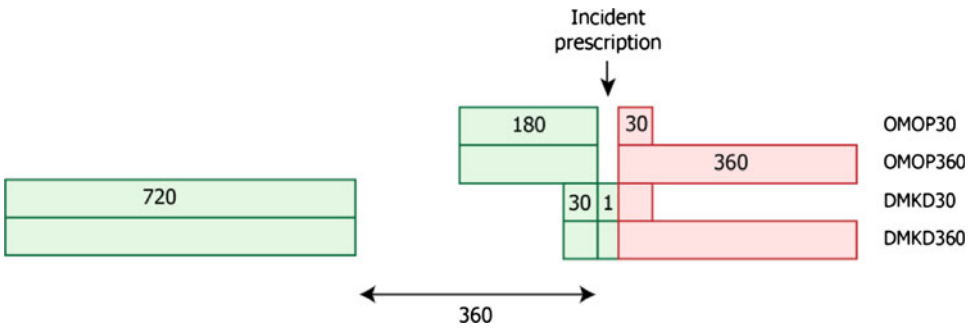

Fig. 5 Four configurations of the self-controlled cohort design in temporal pattern discovery evaluated in detail in this study. Surveillance periods are marked in *red* and control periods in *green*

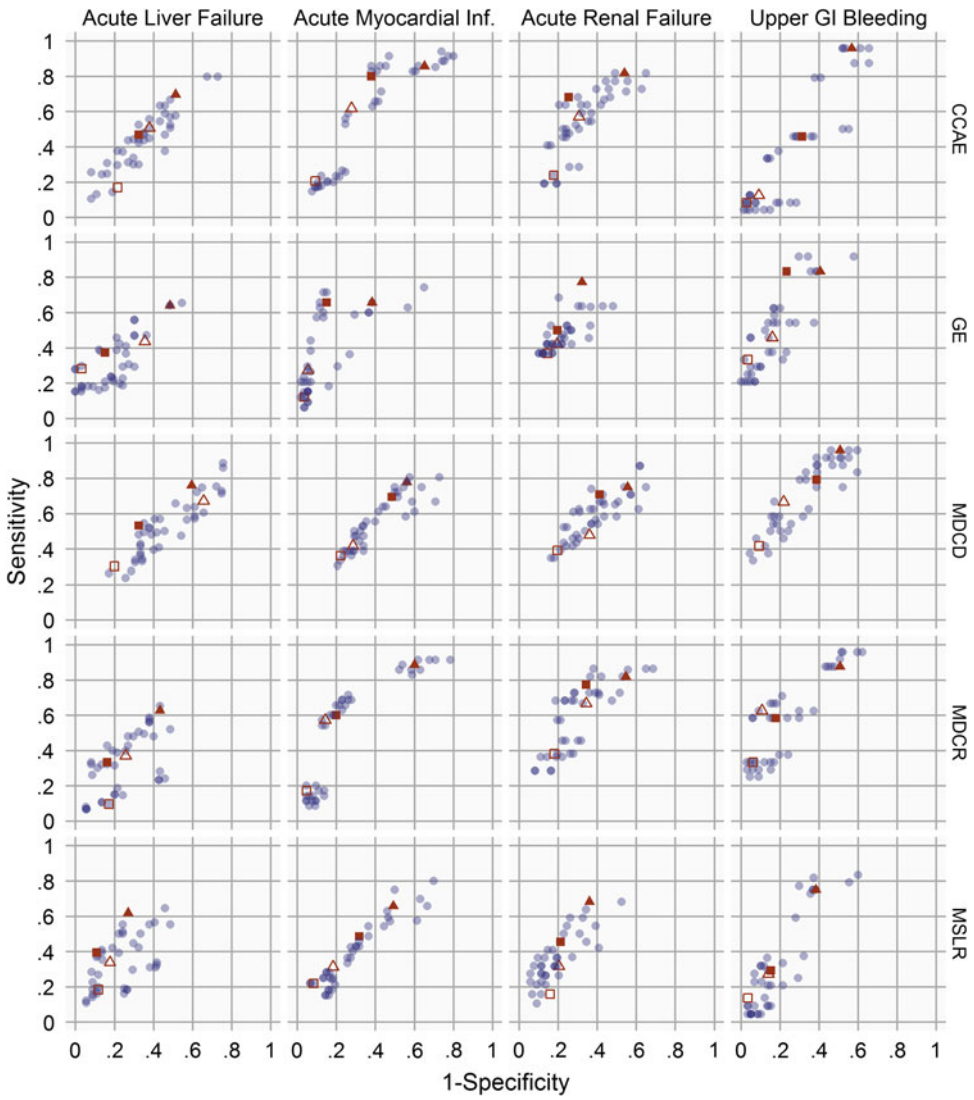

Fig. 6 Scatter plots of sensitivity and specificity of $IC_{\Delta 25} > 0$ for different Temporal Pattern Discovery parameter configurations by outcome and database

3.2 An Illustrative More Detailed Review: Acute Liver Injury in GE Database

We discuss six specific test cases for acute liver injury whose temporal patterns based on the GE database are shown in Figs. 8 and 9. The statistical graphical representations used in these figures are chronographs [7, 8]. The chronograph's bottom panel shows the observed and expected numbers of patients with the outcome of interest in different time periods relative to first prescriptions of the drug of interest. The top panel shows IC values as defined above (essentially the base 2 logarithm of the observed-to-expected ratio). Each bin represents a 30 day period with the exception of the central asterisk which represents the day of the first prescription.

The sharp elevation in the expected registration of acute liver injury on the day of and month after incident prescriptions in the GE database (visible in all six chronographs) indicates that on the whole patients are much more likely to have an event interpreted as acute liver injury registered on the day of or in the month after any new prescription than in the month prior. This increase could conceivably represent a discovery bias, due to incident drug prescriptions being accompanied by liver function tests, thereby making registrations of events interpreted as acute liver injury more likely in post- than pre-exposure time, also for drugs with no such causal effect.

Indeed, the increased registration of acute liver injury after incident sitagliptin prescriptions is well-matched by the increase in its expected value, as shown in Fig. 8d. Acute liver injury is consequently not highlighted to be unexpectedly frequent after sitagliptin with the calibrated self-controlled cohort analysis. Sitagliptin has not been associated with acute liver injury and represents a true negative finding in our analysis.

Isoniazid on the other hand is known to be associated with acute liver injury [24]. The chronograph in Fig. 8a shows a rate of registration that clearly exceeds that

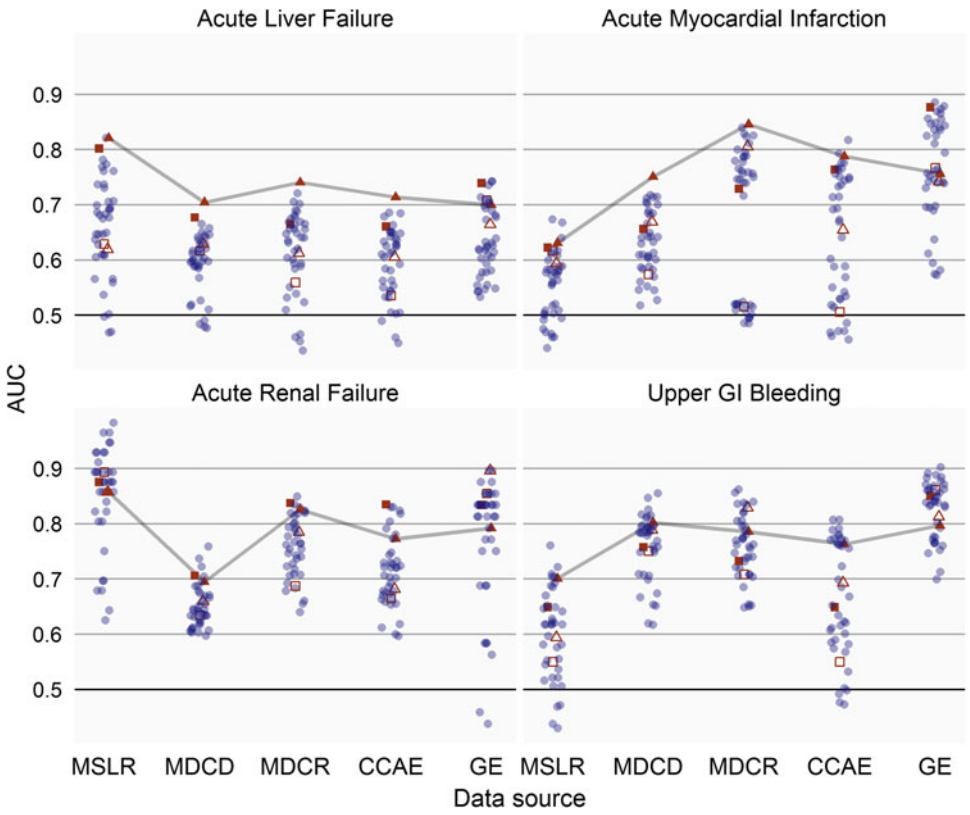

Fig. 7 Area under ROC curve (AUC) for different Temporal Pattern Discovery parameter configurations by outcome and database. *Each dot* represents a parameter configuration, four of which are marked in *red*. The *two red triangles* use a single control period 180 days prior to prescription (see Fig. 5). The *two red squares* use three simultaneous control periods (on the day, 30 days prior, and 3–1 year prior; see Fig. 4). *Open red symbols* denote 30 days surveillance periods and *solid red symbols* denote 360 days surveillance periods. The *solid red triangle* configuration achieved the highest average AUC in this study, and the *open red triangle* configuration was identified as optimal in the OMOP 2010 study. The *two red square* configurations use three simultaneous control periods in the spirit of the analyses in Norén et al. [7]. *MSLR* MarketScan Lab Supplemental, *MDCD* MarketScan Multi-state Medicaid, *MDCR* MarketScan Medicare Supplemental Beneficiaries, *CCAE* MarketScan Commercial Claims and Encounters, *GE* GE Centricity

expected during the first 4–5 months after exposure, representing a true positive finding according to our analysis.

In contrast, erythromycin is another positive test case for acute liver injury with evidence of a true association [25] for which our analysis lends no empirical support, rendering it a false negative finding. The chronograph in Fig. 8b indicates that patients exposed to erythromycin are less likely than the population in general to suffer from acute liver injury across the entire observation period. Since erythromycin is known to cause acute liver injury, this may reflect risk minimization. No increased registration of acute liver injury beyond the expected is observed after new erythromycin prescriptions. However, a relative increase can be noted just prior, perhaps associated with a severe infection for which the drug is often prescribed, particularly since altered liver function tests are included in this definition of acute liver injury.

Primidone is an anticonvulsant that has not previously been associated with acute liver injury, and represents a negative test case. However, a statistically significant (if weak) temporal association is detected in GE data, as shown in Fig. 8c. A plausible explanation for this false positive finding is that primidone is commonly co-prescribed with other anticonvulsant therapies, including carbamazepine and valproate, with known risks of acute liver injury. The calibrated self-controlled cohort analysis does not adjust for confounding by co-therapy in its current form, so events observed during concomitant use may be inappropriately attributed to innocent bystander drugs.

Figure 9 displays two associations with acute liver injury in the GE database that were highlighted by OMOP2012 but not by the DMKD2012 configuration. They exemplify these configurations' different balances between sensitivity and specificity and give insight into their different operational characteristics. Salmeterol is a negative test case and lisinopril is a positive test case. For salmeterol, the registration of acute liver injury is close to the expected across the patient histories, with the exception of immediately prior to the first prescription, where it is lower. This explains why this association is falsely

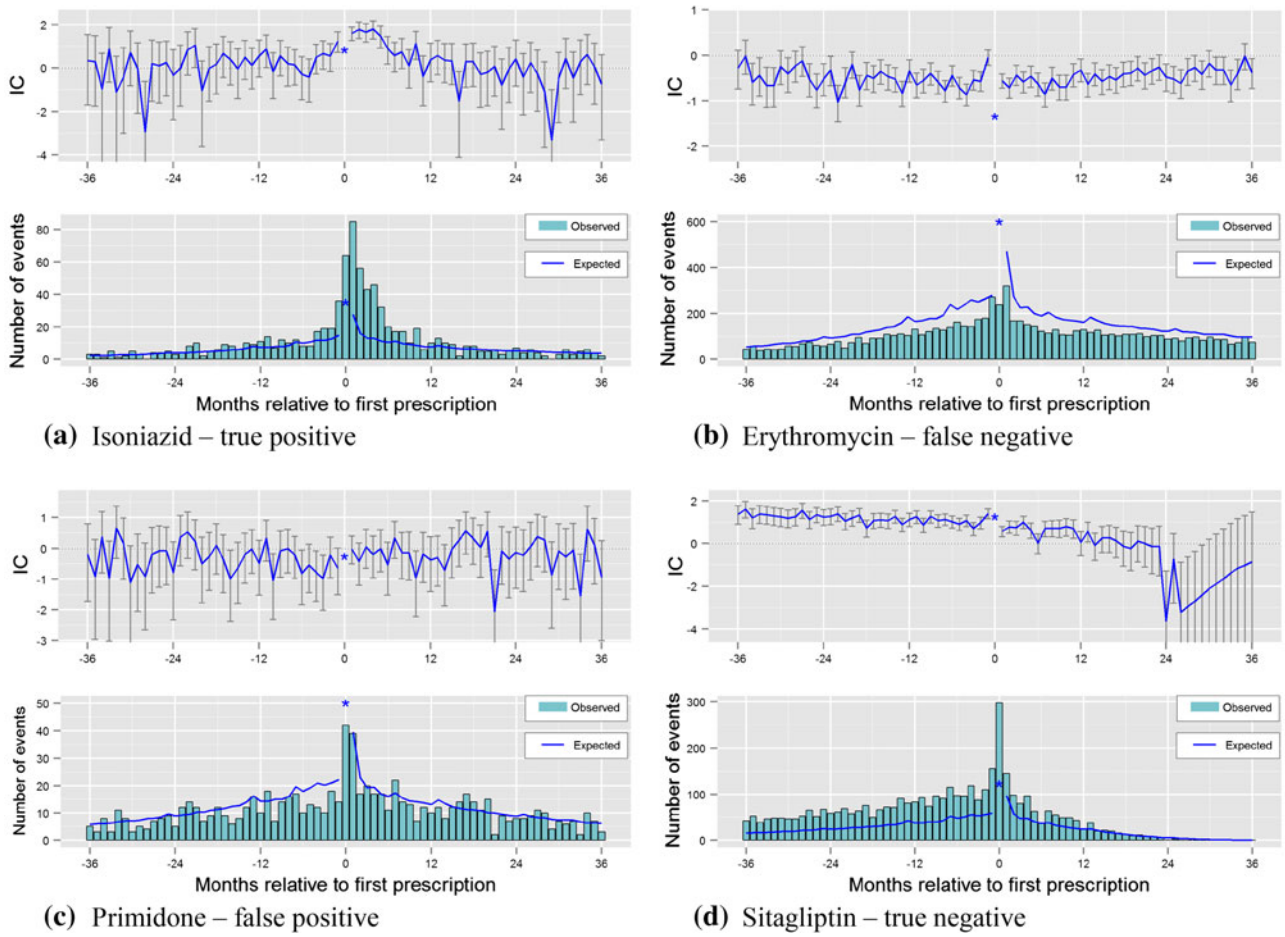

Fig. 8 Chronographs displaying temporal association of acute liver injury with four drugs in the GE database. The x-axis marks 30 days periods relative to first prescriptions of the drug (with the exception of time zero which represents the day of prescription). The *bars* in the *bottom panel* represent the number of patients with a recorded acute liver injury event in each time frame, and the *line* indicates the corresponding expected values. The *upper panel* displays the base 2 logarithm of a shrinkage observed-to-expected ratio ('IC') with 95 % credibility intervals

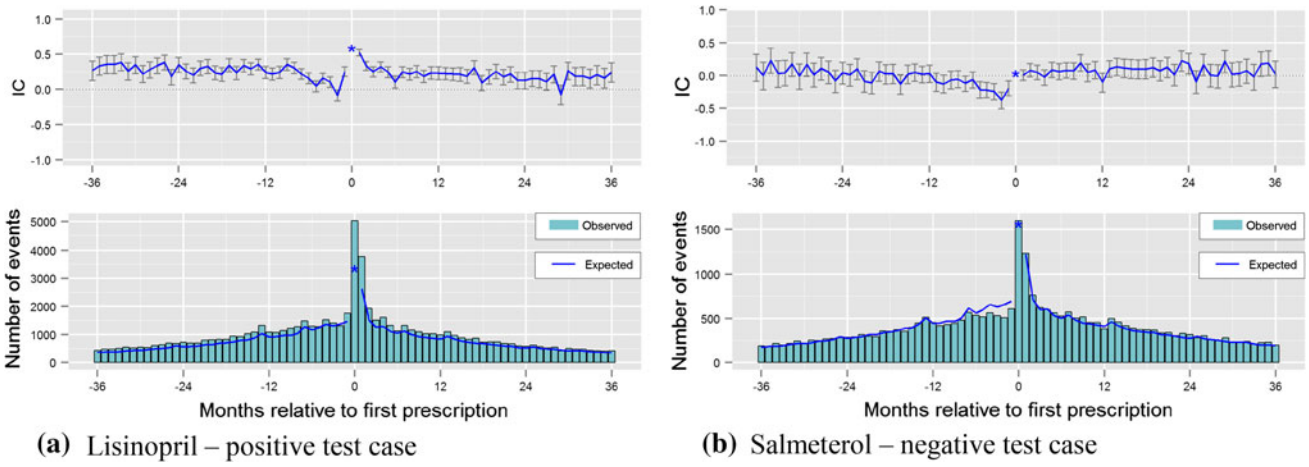

Fig. 9 Chronographs displaying the temporal association for acute liver damage with lisinopril (positive test case) and salmeterol (negative test case) in GE database. Both associations are highlighted by the OMOP360 parameter configuration but not by the DMKD360 parameter configuration

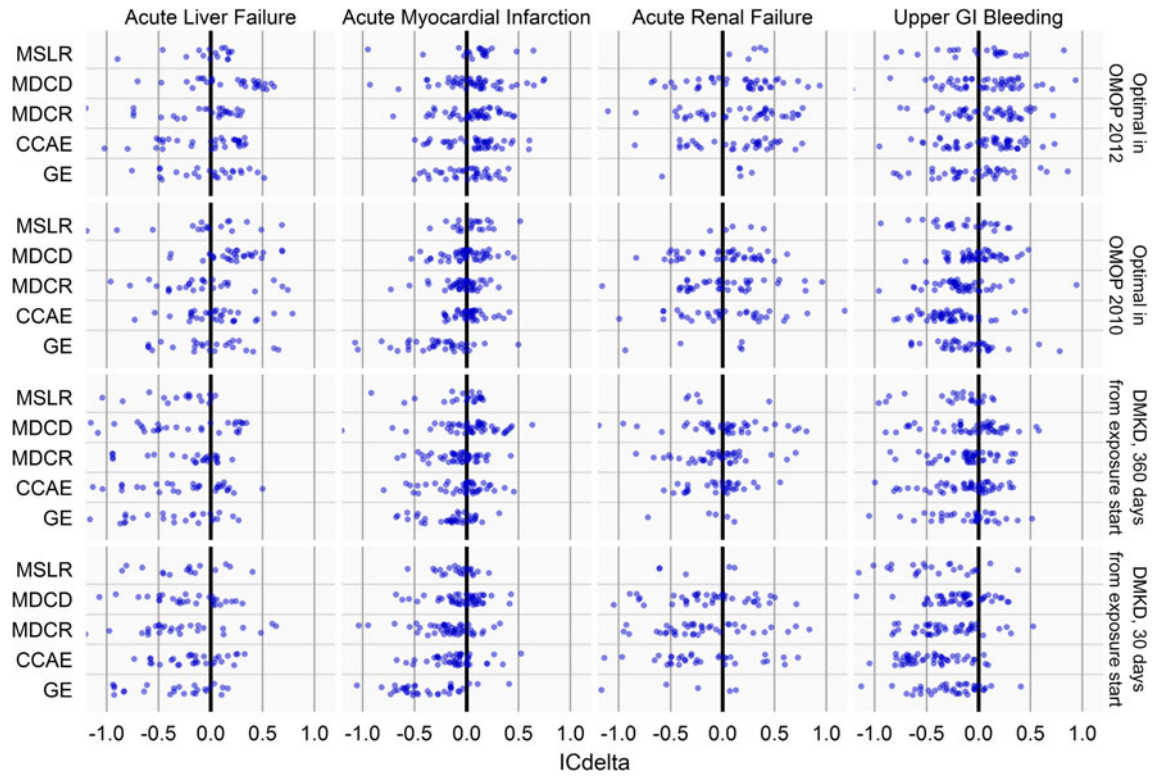

Fig. 10 Distribution of IC_{Δ} for negative controls across outcomes and databases (true value assumed to be zero corresponding to an RR of one). *MSLR* MarketScan Lab Supplemental, *MDCD* MarketScan Multi-state Medicaid, *MDCR* MarketScan Medicare Supplemental Beneficiaries, *CCAE* MarketScan Commercial Claims and Encounters, *GE* GE Centricity, *IC* information component, *DMKD* Data mining and knowledge discovery

highlighted by OMOP2012 that uses a single control period 180 days prior to the first salmeterol prescription, but not by DMKD2012 which also employs a separate control period around 2 years prior. For lisinopril, the registration of acute liver injury is elevated across the entire observation period, except in the few months immediately prior to the prescription. Beyond that, there is a marked relative increase on the day of prescription and in the first month after. This true positive test case is correctly highlighted by OMOP2012 but not by DMKD2012 on account of the latter's consideration of the day of new prescriptions as a separate control period. The elevation on the day could reflect more extensive use of liver function tests for patients prescribed lisinopril compared to for patients prescribed other medicines (in which case a similar bias may affect the first months after lisinopril prescriptions as well).

3.3 Bias

Figure 10 shows the estimated RRs for negative test cases across database-outcome scenarios for the four configurations of particular interest. Across all four outcomes and all 5 databases, ICΔ based on the OMOP30 configurations does not appear to be significantly biased. That is, its expected value for negative test cases is near 0. The OMOP360 configuration may have a slight positive bias for acute renal failure, whereas the DMKD360 configuration exhibits a limited negative bias (around −0.3 on log RR scale or −0.2 on RR scale) for acute liver injury, and DMKD30 a negative bias of similar magnitude for acute liver injury and upper GI bleeding.

4 Discussion

The calibrated self-controlled cohort analysis of the Temporal Pattern Discovery framework achieves good predictive accuracy for the 399 drug-outcome pairs across five observational databases, against the OMOP 2012 set of test cases. Its optimal configuration achieved an average AUC of 0.75, a sensitivity of 0.77 and specificity of 0.52.

The calibrated self-controlled cohort analysis combines within- and between-patient confounder adjustment. We have presented it as a self-controlled design calibrated with an external control group, because its adjustment within the cohort is better developed than its calibration against the external control group. The former includes the use of distinct and simultaneous control periods, whereas the latter is currently basic. As described above, a simplified

active comparator design is available in the open source implementation, but could not be included in the evaluation reported here. In an earlier analysis against the OMOP 2010 test cases, restriction of the external control group to other drugs with the same indication did improve AUC, but the computational burden was nearly prohibitive (results not shown here). In terms of future opportunities, the self-controlled cohort analysis lends itself naturally to stratification and so could be nested in a more sophisticated cohort design, e.g. stratified by propensity scores. Such a design might equally well be described as a comparator-adjusted analysis calibrated for residual confounding based on the rates of the outcome in the control period(s) for the exposed and the comparators, respectively; the two modes of confounder adjustment are applied on equal footing, and one does not inherently take precedence over the other.

The calibrated self-controlled cohort analysis utilizes shrinkage observed-to-expected ratios to rank temporal associations. These measure the strength of association adjusted for the support in data. Shrinkage observed-to-expected rations are used also in basic disproportionality analysis for longitudinal data [26], but then in what amounts to an unadjusted incident user design, with corresponding performance.

Out of the designs in Table 1, the self-controlled case series and case-crossover designs have the strongest theoretical foundation and broadest epidemiological use. Their main advantages over the self-controlled cohort designs include the matching of time periods within individual patients, as opposed to at the cohort level. Within-patient matching yields more efficient confounder-adjustment, and with shrunk odds ratios [27] as the basic measure of association, it would amount to ignoring patients with registrations of the outcome in both control and surveillance periods, in the analysis. Conceptually, this would be preferable, but it should be noted that methods that incorporate within-patient matching have not yielded superior performance in the empirical evaluations to date [10]. Another advantage of the self-controlled case series and case-crossover designs is that they require data on exposed cases only. This can be a strength when reliable data is hard to ascertain, but may have more limited impact on broad screening based on readily available data such as is considered here. The major advantage of the calibrated self-controlled cohort analysis (shared with the case–time–control design for a case–control type analysis) is that it offers the opportunity to utilize an external control group to reduce the impact of systematic differences across exposed and unexposed patient time; an example of such a systematic difference is the tendency of acute liver injury to be registered more often after than before prescriptions in the GE database: five of the six drugs evaluated above have higher rates of acute liver injury in the month after than in the month before treatment initiation, including two of the three negative test cases. Calibration by the external control group is intended to reduce the bias of the IC_{Δ} measure and provide a natural threshold at zero, corresponding to a relative risk of one, based on which temporal associations can be identified. It should provide a neutral ground based on which different outcomes can be rank-ordered by strength of association. This would be important in real-world risk identification, but does not impact the performance evaluation in this phase of OMOP, which is threshold-independent and stratified by outcome.

The self-controlled cohort analysis as implemented here compares proportions of patients with the event in specific time frames, rather than rates of events per unit time. This sets it apart from the other self-controlled designs in Table 1. The original motivation for this was to ensure robustness to data quality issues associated with repeated registrations related to the same event [7, 8]. Considering the drawbacks, it reduces sensitivity to increased rates of common events, and allows only crude adjustment for loss to follow-up: partial left- or right-censoring of control and surveillance periods is currently ignored. In the OMOP experiment, the challenge of repeat registration of the same event is also met by focusing exclusively on first-in-era medical events. This on the other hand introduces an immortal time bias, unless explicitly accounted for in the computation. The adaptation of Temporal Pattern Discovery to event rates is straightforward (the primary challenge being to ensure a computationally efficient modification of the time-at-risk), and the empirical evaluation of such a design is an interesting topic for future research.

The broad empirical evaluation against the OMOP 2012 test cases lends only partial support to the use of simultaneous control periods. For 360 days surveillance, the configuration with simultaneous control periods achieved higher AUC in the GE database as well as for acute renal failure across all databases, but lower AUC in other scenarios. By definition, simultaneous control periods increase specificity at the expense of decreased sensitivity (the estimated temporal association is always lower than or equal to that based on the individual control periods), for a fixed threshold. The use of a control period long before the prescription eliminated the false positive association between salmeterol and acute liver injury. On the other hand, the increased rate of acute liver injury on the day of lisinopril prescriptions led to this association not being highlighted by DMKD360; a false negative finding according to the reference set. Simultaneous control periods appear to work best for the GE electronic health records database. The Temporal Pattern Discovery methodology was originally developed for EHR databases, [7, 8] and it is conceivable that the control period long before

first prescriptions works less well in claims databases, for which patient follow-up tends to be shorter.

A major strength of the OMOP 2012 study compared to the OMOP 2010 study is the significantly expanded reference set of 399 test cases, which makes findings more robust to random variability. At the same time, the test cases are now restricted to four outcomes, which make results difficult to generalize to other medical conditions. The use of a 360 day surveillance period increased the chance that the self-controlled cohort analysis would rank positive test cases over negative test cases (as measured by AUC) against the 2012 reference set. In contrast, the configuration with highest average AUC in the OMOP 2010 experiment utilized a 30 day surveillance period, [10, 11] which achieved a lower AUC here. A possible explanation of this discrepancy is that the four outcomes all have insidious onset. It emphasizes the difficulty of identifying a one-size-fits-all optimal methodology for risk identification in longitudinal observational databases. One limitation of this and other reference sets considered to date is their focus on established safety issues. The patterns for well-known risks to medicines are likely to differ in important ways from those of unknown risks. In particular, the patterns for known safety hazards are likely to be affected by risk minimization activities, including contraindications and closer surveillance. The former will decrease the event rate prior to prescriptions of the drug of interest thus inflating the ratio of the rate of events in the exposed to unexposed time. It should be noted that this introduces an unfair advantage for self-controlled designs in the performance evaluation reported here that would not carry over to prospective risk identification. It might also favor reliance on control periods immediately prior to new prescriptions, although we did not find examples of this among the limited set of test cases reviewed in-depth here. Future performance evaluation should ideally focus on emerging safety issues, which might be achieved by using historical data restricted to the time point up until the initial safety signal. This will further limit the available data, but may be feasible within the on-going national and international collaborations with access to large database networks.

The performance evaluation should also be interpreted in light of its restriction to scenarios powered to detect a relative risk of 1.25. This effectively excludes drug-ADR combinations with low to moderate support in data. In particular, the use of a fixed threshold on a detectible relative risk may place undue emphasis on outcomes with high background rates in the population. For a given relative risk, power does increase with the background rate of the event: the more common the event, the greater the chance to detect a true relative risk of that magnitude. On the other hand, the lower the background rate of the event, the more likely it is that the true relative risk attains a certain magnitude. As an example, it seems likely that a restriction to detectible RR = 1.25 would exclude too many drugs for rare reactions such as agranulocytosis, for which true relative risks are likely to be higher than for events with more complex causal pathways such as myocardial infarction. Ideally, the threshold on the minimally detectible RR should account for the nature of the outcome under study and be based on empirical evidence that a revised threshold still yields adequate predictive accuracy.

As exemplified by the in-depth review of test cases for acute liver injury in the GE database, broad exploratory analysis as made possible by the visualization in chronographs can provide useful clinical and epidemiological perspectives and insights that cannot be distilled from univariate measures of temporal association. In this case, the chronographs showed a marked increase in the registration of acute liver injury after any treatment initiation, conceivably associated with an associated increase in liver function testing. Without such insights into the nature of data at hand, one may easily misinterpret observed patterns—in particular if the available information is limited to univariate measures of association that alone will not allow for effective risk identification in observational databases.

Given the necessity to rely on generic analytics for first pass screening, highlighted issues will require clinical and statistical review. To this end, guidelines and best practice need to be developed for the scrutiny of highlighted associations, a topic currently being investigated in the Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium (PROTECT) project (http://www.imi-protect.eu/).

5 Conclusions

The calibrated self-controlled cohort analysis in Temporal Pattern Discovery combines within- and between-patient confounder adjustment. Its unique aspects compared to other self-controlled designs are the use of simultaneous control periods and the calibration by an external control group. It shows promise as a tool for risk identification, performing well at discriminating positive from negative test cases, but its optimal parameter configuration may vary with the data set and medical outcome of interest.

Acknowledgments The Observational Medical Outcomes Partnership is funded by the Foundation for the National Institutes of Health (FNIH) through generous contributions from the following: Abbott, Amgen Inc., AstraZeneca, Bayer Healthcare Pharmaceuticals, Inc., Biogen Idec, Bristol-Myers Squibb, Eli Lilly & Company, GlaxoSmithKline, Janssen Research and Development, Lundbeck, Inc., Merck & Co., Inc., Novartis Pharmaceuticals Corporation, Pfizer Inc, Pharmaceutical Research Manufacturers of America (PhRMA),

Roche, Sanofi-aventis, Schering-Plough Corporation, and Takeda. Dr. Ryan is an employee of Janssen Research and Development, and a past employee of GlaxoSmithKline, but does not receive compensation for his work with OMOP. Dr. Schuemie has become an employee of Janssen Research and Development since completing the work described here. Dr. Schuemie received a fellowship from the Office of Medical Policy, Center for Drug Evaluation and Research, Food and Drug Administration. Drs. Schuemie and Madigan receive funding from FNIH. Dr. Norén, Mr. Bergvall and Ms. Juhlin were partially funded by FNIH grant NOREN11OMOP. The authors appreciate the thoughtful feedback from I. Ralph Edwards at the Uppsala Monitoring Centre, the OMOP Research Investigators, Executive Board, and Advisory Boards. No compensation was given for the reviews.

This article was published in a supplement sponsored by the Foundation for the National Institutes of Health (FNIH). The supplement was guest edited by Stephen J.W. Evans. It was peer reviewed by Olaf H. Klungel who received a small honorarium to cover out-of-pocket expenses. S.J.W.E has received travel funding from the FNIH to travel to the OMOP symposium and received a fee from FNIH for the review of a protocol for OMOP. O.H.K has received funding for the IMI-PROTECT project from the Innovative Medicines Initiative Joint Undertaking (http://www.imi.europa.eu) under Grant Agreement no 115004, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007–2013) and EFPIA companies' in kind contribution.

References

1. Edwards IR, Lindquist M, Wiholm BE, Napke E. Quality criteria for early signals of possible adverse drug reactions. Lancet. 1990;336(8708):156–8.
2. Rawlins MD. Spontaneous reporting of adverse drug reactions. I: the data. Brit J Clin Pharmacol. 1988;26(1):1–5.
3. Meyboom RH, Lindquist M, Egberts AC. An ABC of drug-related problems. Drug Saf. 2000;22(6):415–23.
4. Harpaz R, DuMouchel W, LePendu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. Clin Pharmacol Ther. 2013;93(6):539–46.
5. Public Law 110-85: Food and Drug Administration Amendments Act of 2007. 2007.
6. Woodcock J, Behrman RE, Dal Pan GJ. Role of postmarketing surveillance in contemporary medicine. Annu Rev Med. 2011; 62:1–10.
7. Norén GN, Hopstadius J, Bate A, Star K, Edwards IR. Temporal pattern discovery in longitudinal electronic patient records. Data Min Knowl Discov. 2010;20(3):361–87.
8. Norén GN, Bate A, Hopstadius J, Star K, Edwards IR. Temporal pattern discovery for trends and transient effects: its application to patient records. In: ACM SIGKDD international conference on knowledge discovery and data mining, KDD '08. Las Vegas: ACM; 2008. p. 963–71.
9. Suissa S. The case–time–control design. Epidemiology. 1995; 6(3):248–53.
10. Ryan PB, Madigan D, Stang PE, Marc Overhage J, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012;31(30):4401–15.
11. Norén GN, Hopstadius J, Bate A, Edwards IR. Safety surveillance of longitudinal databases: results on real-world data. Pharmacoepidemiol Drug Saf. 2012;21(6):673–5.
12. Schuemie MJ. Safety surveillance of longitudinal databases: further methodological considerations. Pharmacoepidemiol Drug Saf. 2012;21(6):670–2.
13. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. Drug Saf. 2013 (in this supplement issue). doi:10.1007/s40264-013-0097-8.
14. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.
15. Ryan PB, Schuemie MJ, Madigan D. Empirical performance of a self-controlled cohort method: lessons for developing a risk identification and analysis system. Drug Saf. 2013 (in this supplement issue). doi:10.1007/s40264-013-0101-3.
16. Hallas J. Evidence of depression provoked by cardiovascular medication: a prescription sequence symmetry analysis. Epidemiology. 1996;7:5.
17. Farrington CP, Nash J, Miller E. Case series analysis of adverse reactions to vaccines: a comparative evaluation. Am J Epidemiol. 1996;143(11):1165–73.
18. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. Am J Epidemiol. 1991;133(2):144–53.
19. Tannen RL, Weiner MG, Xie D. Use of primary care electronic medical record database in drug efficacy research on cardiovascular outcomes: comparison of database and randomised controlled trial findings. BMJ. 2009;338:b81.
20. Schuemie MJ. Safety surveillance of longitudinal databases: further methodological considerations. Pharmacoepidemiol Drug Saf. 2012;21(6):670–2. doi:10.1002/pds.3259.
21. Tisdale J, Miller D. Drug-induced diseases: prevention, detection, and management. 2nd ed. USA: American Society of Health-System Pharmacists; 2010.
22. Armstrong B. A simple estimator of minimum detectable relative risk, sample size, or power in cohort studies. Am J Epidemiol. 1987;126(2):356–8.
23. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology. 1983;148(3):839–43.
24. Smith BM, Schwartzman K, Bartlett G, Menzies D. Adverse events associated with treatment of latent tuberculosis in the general population. CMAJ. 2011;183(3):E173–9.
25. Carson JL, Strom BL, Duff A, Gupta A, Shaw M, Lundin FE, et al. Acute liver disease associated with erythromycins, sulfonamides, and tetracyclines. Ann Intern Med. 1993;119(7 Pt 1):576–83.
26. Zorych I, Madigan D, Ryan P, Bate A. Disproportionality methods for pharmacovigilance in longitudinal observational databases. Stat Methods Med Res. 2013;22(1):39–56.
27. Norén GN, Hopstadius J, Bate A. Shrinkage observed-to-expected ratios for robust and transparent large-scale pattern discovery. Stat Methods Med Res. 2013;22(1):57–69.