Application of Multivariate Self-Controlled Case Series Method for Active Drug Safety Surveillance in a Signal Screening Framework: Leveraging THIN Data in the OMOP Common Data Model

Xiaofeng Zhou, PhD[1], Rongjun Shen, MS[1], Sundaresan Murugesan, MS[1], Andrew Bate, PhD[1] , [1] Epidemiology, Pfizer Inc, NY, USA

Abstract

Literature on the application of the Multivariate Self-Controlled Case Series (MSCCS) method and its feasibility for active surveillance in a signal screening framework is sparse. The objective of this study is to test the MSCCS method using THIN, an UK EMR database in the Observational Medical Outcome Partnership (OMOP) Common Data Model (CDM), and assess performance characteristics of MSCCS and compare with Univariate Self-Controlled Case Series (USCCS) method. Drug and health outcome of interests predefined by OMOP were used as exposures and outcomes. The OMOP reference set of 53 drug-outcome pairs (9 positive and 44 negative control pairs) was used as a "Ground Truth" for method assessment. The results indicated that MSCCS highlighted 7 out of 9 pairs of the true associations, 5 with statistical significance. Thirty six true negative pairs out of 44 negative controls were identified. MSCCS offered lower sensitivity (56 vs. 78%) but better specificity (82 vs. 59%) than USCCS. MSCCS provided with stronger positive predictive value (PPV) than USCCS while the negative predictive value (NPV) was similar. For both methods, the majority of false positive pairs (>=63%) had RR <2. Most of RR estimates in MSCCS are smaller than those in USCCS for the same drug-outcome pairs. Sensitivity and specificity of MSCCS method vary by choosing other parameter settings: most variability with changing the risk period, followed by hyperparameter setting and minimal variability with changing with risk period start date or condition type. MSCCS was more computational intensive (≥6 times to complete a run) than USCCS and had more stringent system requirements to ensure tractability. In conclusion, MSCCS scores, though imperfect, were more predictive of true associations than USCCS scores, with similar negative predictive values. There was some performance variability in MSCCS based on choice of parameters. For both methods, a RR <2 led to a higher risk for false positives. Further research to understand MSCCS' performance characteristics across other databases are needed.

Background

Drs. Madigan and Zorych of Columbia University, developed Univariate and Multivariate variants of the Self-Controlled Case Series method for OMOP. As an OMOP Extended Consortium member, we mapped the UK THIN EMR database into OMOP Common Data Model (CDM) and tested their SCCS method implementations.

Acknowledgement

We wish to thank Drs. Madigan and Zorych for their advice to the implementation of MSCCS method for this study

Objective

To test the MSCCS method on THIN CDM, to assess the performance characteristics of MSCCS and compare with Univariate Self-Controlled Case Series (USCCS) method

Table 1: Test Results of MSCCS and USCCS Methods

	Test Results		OMOP Reference Set
	MSCCS	USCCS	N
True Positive Pairs (n)	5	7	9
True Negative Pairs (n)	36	26	44
False Positive Pairs (n)	8	18	NA
False Negative Pairs (n)	4	2	NA

Figure 1. PPV and NPV Results

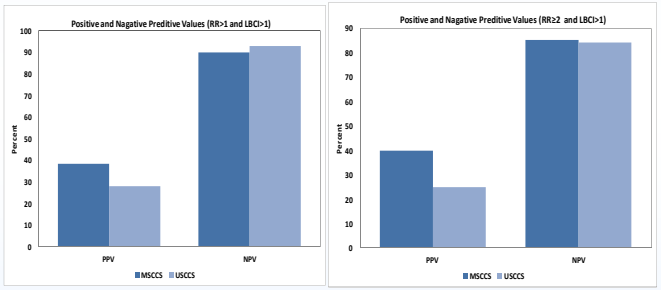
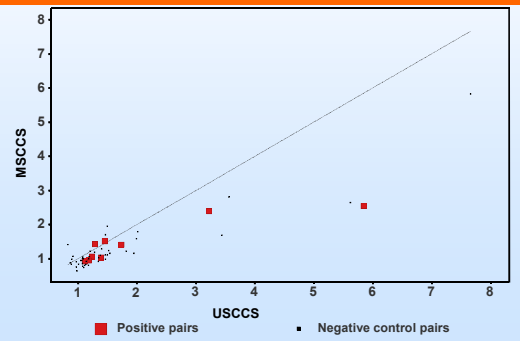


Figure 2: RR Estimates of 53 Drug and Outcome Pairs



Methods

- SCCS method only includes cases in the analysis. The MSCCS model accounts for the presence of multiple drugs and USCCS only considers one drug.
- 10 drug and 9 health outcome of interests predefined by OMOP were used as exposures and outcomes. The OMOP reference set of 53 drug-outcome pairs was used as a "Ground Truth" for method assessment. SAS codes and parameter settings provided by OMOP were applied.

Results

- MSCCS highlighted 7 out of 9 pairs of the true associations, 5 with statistical significance. Thirty six true negative pairs out of 44 negative controls were identified (Table 1)
- MSCCS offered lower sensitivity (56 vs. 78%) but better specificity (82 vs. 59%) than USCCS. For both methods, the majority of false positive pairs (>=63%) had RR <2.
- MSCCS provided with stronger positive predictive value (PPV) than USCCS while the negative predictive value (NPV) was similar: using the threshold (RR>1 and lower bond 95% CI (LBCI)>1), the PPV was stronger for MSCCS (38 vs. 28%) while NPV was slight weaker (90 vs. 93%) than USCCS; using RR>=2 and LBCI>1, the PPV was stronger for MSCCS (40 vs. 25%) while NPV was slight better (85 vs. 84%) than USCCS (Figure 1).
- Most of RR estimates in MSCCS are smaller than those in USCCS for the same drug-outcome pairs. (Figure 2).
- Sensitivity and specificity of MSCCS method vary by choosing other parameter settings, in particular with most variability with changing the risk period,
- MSCCS was more computational intensive (≥6 times to complete a run) than USCCS and had more stringent system requirements to ensure tractability.

Conclusions

- MSCCS scores, though imperfect, were more predictive of true associations than USCCS scores, with similar NPV. There was some performance variability in MSCCS based on choice of parameters. For both methods, a RR <2 led to a higher risk for false positives. Further research to understand MSCCS' performance characteristics across other databases are needed.
- The study limitations include that the results are focused on configuration selection optimizing performance in an exploratory framework in a single database and the current test cases are limited to the 53 pairs in OMOP reference set, which were well studied examples of medications marketed for many years.

Conflict of Interest: X. Zhou, R. Shen , S. Murugesan , and A. Bate are employees of Pfizer Inc.