Empirical evaluation of the OHDSI Methods Library

Martijn J. Schuemie[1], M. Soledad Cepeda[1], Marc A. Suchard[2], Yuxi Tian[2], Alejandro Schuler[3], Patrick B. Ryan[1,4], George Hripcsak[4,5]

[1]Janssen R&D, Titusville, NJ, USA; [2]University of California, Los Angeles, CA, USA; [3]Stanford, Stanford, CA, USA; [4]Columbia University, New York, NY, USA; [5]NewYork-Presbyterian Hospital, New York, NY, USA

Background

The OHDSI Methods Library is a set of R packages implementing most well-known observational analyses. Here we evaluate several variations of each design using the OHDSI Population-Level Estimation Benchmark. The results not only inform on the operating characteristics of the designs in general, they are also a validation of the software implementation in the Library.

The Benchmark allows evaluation on two distinct tasks: effect estimation and comparative effect estimation. Currently only the cohort method is suited for comparative analyses, and can also be used for effect estimation if the comparator is believed to not cause the outcome.

OMOP experiments

Previously, a large number of observational analysis designs have been evaluated in the OMOP experiments. Here we aim to improve on these evaluations in several ways:

- Positive controls now have known effect sizes.
- Positive controls are not known to physicians and therefore cannot bias the performance evaluation.
- Improvements in the implementation of methods. For example, the cohort method now includes survival models, and the self-controlled cohort restricts control time to the observation period.

Database

In this initial run of the experiment a single database was used: the Truven MarketScan Multi-state Medicaid (MDCD) database. MDCD is an administrative health claims database for the pooled healthcare experience of Medicaid enrollees from multiple states. As of 5 July 2017, MDCD contained 30 million patients with patient-level observations from January 2006 through December 2016.

Challenges

Some exposures and outcomes of interest are highly prevalent (e.g. there are over 1 million ciprofloxacin users, and over 1 million cases of otitis media in the database) requiring methods to be adapted to handle large numbers.

Next steps

- Expand the number of variations per method, especially looking at combinations of several options.
- Run the evaluation on multiple databases across the OHDSI network.
- Further analysis of results using advanced statistics.

Cohort Method
New-user cohort studies using large-scale regression for propensity and outcome models.

SCCS
Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.

Case-Control
Case-control studies, matching controls on age, gender, provider, and visit date. Allows nesting in another cohort.

Self-Controlled Cohort
A self-controlled cohort design, where time preceding exposure is used as control.

Case-Crossover
Case-crossover design including the option to adjust for time-trends in exposures (so-called case-time-control).

Metrics computed using controls with MDRR < 1.25 (139 negative and 348 positive controls)

Method	Analysis choices	AUC	Coverage of 95% CI	Mean precision	MSE	Type 1 error	Type 2 error	Missing
Case-control	Matching on age and gender, 2 controls per case	0.92	0.12	1812.92	0.6	0.81	0.01	0.01
Case-control	Matching on age and gender, 10 controls per case	0.91	0.1	3303.4	0.58	0.84	0.01	0.01
Case-control	Matching on age and gender, nesting in indication, 2 controls per case	0.9	0.3	1344.33	0.48	0.64	0.04	0.01
Case-control	Matching on age and gender, nesting in indication, 10 controls per case	0.91	0.25	2189.06	0.55	0.7	0.03	0.01
Case-crossover	Simple case-crossover	0.85	0.35	486.51	0.76	0.7	0.07	0
Case-crossover	Nested case-crossover	0.85	0.43	284.12	1.34	0.59	0.11	0
Case-crossover	Nested case-time-control, matching on age and gender	0.82	0.61	117.27	1.5	0.44	0.19	0.01
Cohort method	No matching, simple outcome model	0.76	0.42	131.74	1.17	0.49	0.18	0.04
Cohort method	Matching plus simple outcome model	0.82	0.61	85.66	0.58	0.26	0.23	0.11
Cohort method	Stratification plus stratified outcome model	0.86	0.68	104.05	1.46	0.19	0.23	0.06
Cohort method	Matching plus stratified outcome model	0.8	0.82	39.54	0.43	0.08	0.35	0.13
Cohort method	Matching plus full outcome model	0.77	0.86	25.22	0.42	0.01	0.54	0.49
SCCS	Simple SCCS	0.9	0.28	1958.69	0.45	0.71	0.02	0
SCCS	Using pre-exposure window	0.89	0.26	1871.1	0.48	0.75	0.03	0
SCCS	Using age and season	0.91	0.28	1913.83	0.45	0.7	0.01	0
SCCS	Using event-dependent observation	0.88	0.25	1906.17	0.5	0.7	0.02	0
SCCS	Using all other exposures	0.9	0.41	962.33	0.39	0.55	0.03	0
Self-controlled cohort	Length of exposure, index date in exposure window	0.9	0.32	1418.27	0.3	0.55	0.09	0.01
Self-controlled cohort	30 days of each exposure, index date in exposure window	0.91	0.52	466.84	0.08	0.49	0.11	0
Self-controlled cohort	Length of exposure, index date in exposure window, require full obs	0.91	0.34	1217.81	0.29	0.51	0.09	0.01
Self-controlled cohort	30 days of each exposure, index date in exposure window, require full obs	0.91	0.52	466.84	0.08	0.49	0.11	0
Self-controlled cohort	Length of exposure, index date ignored	0.94	0.36	1392.35	0.18	0.5	0.1	0.01
Self-controlled cohort	30 days of each exposure, index date ignored	0.93	0.55	438.31	0.09	0.26	0.14	0
Self-controlled cohort	Length of exposure, index date ignored, require full obs	0.94	0.39	1187.46	0.17	0.44	0.1	0.01
Self-controlled cohort	30 days of each exposure, index date ignored, require full obs	0.93	0.55	438.31	0.09	0.26	0.14	0

AUC: Area under the ROC curve for classifying all positive controls vs. all negative controls
Coverage: Coverage of the 95% confidence interval
Mean precision: Precision = $1/SE^2$; higher precision means narrower confidence intervals
MSE: Mean squared error between effect size (point) estimate and the true effect size
Type 1 error: For all negative controls, how often was the null rejected (at alpha = 0.05)
Type 2 error: For all positive controls, how often was the null not rejected (at alpha = 0.05)
Missing: For how many of the controls was the method unable to produce an estimate

The good

- Cohort method using PS matching has high coverage but relatively low power (but only cohort method can be used for comparative effect estimation).
- Self-controlled cohort has high AUC and low MSE, but coverage is lacking.

The bad

- (Nested) case-control has low coverage, often overestimating the true effect size with very narrow confidence intervals.
- Cohort method with full outcome models often fails to produce estimates due to lack of power.

The full set of results can be explored interactively at

data.ohdsi.org/MethodEvalViewer/