

Name:	Martijn Schuemie
Affiliation:	Janssen R&D
Email:	schuemie@ohdsi.org
Presentation type (s):	Poster

## Empirical evaluation of the OHDSI Methods Library

Martijn J. Schuemie<sup>1</sup>, M. Soledad Cepeda<sup>1</sup>, Marc A. Suchard<sup>2</sup>, Yuxi Tian<sup>2</sup>, Alejandro Schuler<sup>3</sup>, Patrick B. Ryan<sup>1</sup>, George Hripcsak<sup>4</sup>

<sup>1</sup>Janssen R&D, Titusville, New Jersey, USA; <sup>2</sup>University of California, Los Angeles, California, USA; <sup>3</sup>Stanford, Stanford, California, USA; <sup>4</sup>New York-Presbyterian Hospital, New York, New York, USA

### Abstract

*The OHDSI Methods Library is a set of R packages implementing most well-known observational analysis designs, such as the new-user cohort, self-controlled case series, case-control, self-controlled cohort, case-crossover, and case-time control designs. Here we evaluate a large number of variations of each design using the OHDSI Methods Benchmark on a US insurance claims database. 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.*

### Introduction

Evidence from non-randomized observational studies has become a critical component of the evidence base about the safety and effectiveness profile of a medicine after approval, but the degree to which the evidence from observational analyses is consistent with ‘truth’ is not well understood. For example, because of non-random error such as unmeasured confounding it is unclear how often the 95% confidence interval (CI) of estimated effect sizes contain the true effect size.

A large number of observational analysis designs have been evaluated in the OMOP experiment<sup>1</sup>, but that evaluation has several limitations. One limitation is the uncertainty surrounding magnitude of effect sizes for positive controls, a second limitation is the possible bias due to positive controls being known to physicians<sup>2</sup>. Another limitation is issues concerning the implementation of methods. For example, the cohort method did not include survival models, and the self-controlled cohort design did not restrict control time to the observation period.

We have conducted a new large-scale observational data experiment that aims to establish the operating characteristics of several standard observational analysis methods. Specifically, we focus on those methods implemented in the OHDSI Methods Library, including the new-user cohort (optionally using propensity score adjustment), self-controlled case series (SCCS), case-control, self-controlled cohort, case-crossover, and case-time control designs. Besides the lessons learned from the OMOP experiment little is known about the performance of these designs in general, and even less of the performance of these specific implementations of the designs. We use the OHDSI Methods Benchmark to evaluate performance.

### Methods Benchmark

The Benchmark contains 200 negative and 600 positive controls where the true effect size is known, allowing measurement of performance metrics such as coverage of the confidence interval, area under the receiver operator curve, mean squared error, bias distribution, and type I and II error. The Benchmark is design for evaluating methods aimed at *effect estimation* (estimation of the average effect of an exposure on an outcome relative to no exposure) as well as methods aimed at *comparative effect estimation* (estimation of the average effect of an exposure on an outcome relative to another exposure).

## Database

In this initial run of the experiment a single database was used: the Truven Market Scan 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 1 November 2016, MDCD contained 21.6 million patients with patient-level observations from January 2006 through December 2014. After the OHDSI Symposium the experiment will be extended to other partners in the OHDSI network.

## Estimation methods

We evaluated the following effect estimation methods:

- New-user cohort design (optionally using propensity score adjustment) using an active comparator representing the counterfactual of no effect.
- Self-Controlled Case Series
- Self-Controlled Cohort
- Case-Control
- Case-Crossover
- Case-Time-Control

We evaluated the following comparative effect estimation method:

- New-user cohort design (optionally using propensity score adjustment)

For each method, a large number of variations was included in the evaluation, for example using either propensity score matching or stratification.

## Results

The preliminary results from the execution of this experiment suggest interesting results that warrant further consideration. For example, the (nested) case-control design appears to have poor performance in general, the self-controlled cohort shows good performance in general, and one specific implementation of the case-time-control showed very strong bias towards the null.

## Conclusions

Our evaluation allows characterization of performance of observational study designs and analytic choices, and provide an understanding of how methods perform in general. It also provides a thorough validation of the implementation of these methods in the OHDSI Methods Library. As the Library is used in more and more observational studies, such validation is crucial

## References

1. Ryan PB, Stang PE, Overhage JM, et al. A comparison of the empirical performance of methods for a risk identification system. *Drug safety*. 2013; 36 Suppl 1: S143-58.
2. Noren GN, Caster O, Juhlin K and Lindquist M. Zoo or savannah? Choice of training ground for evidence-based pharmacovigilance. *Drug safety*. 2014; 37: 655-9.