[Note: All submissions must be in PDF format. Failure adhere to the format requirements may result in rejection of your submission without review]

Name:	Martijn J. Schuemie PhD
Affiliation:	Janssen R&D
Email:	schuemie@ohdsi.org
Presentation type (select one):	Collaborator Demonstration

Demonstration of the OHDSI Methods Library

Martijn J. Schuemie, PhD[1], Patrick B. Ryan[1], PhD, Marc A. Suchard[2], MD, PhD
[1]Janssen Research and Development, Raritan, NJ, USA [2]Dept. of Biomathematics & Dept. of Human Genetics, David Geffen School of Medicine, Uni. of California, Los Angeles, CA, USA

Abstract

The OHDSI methods library is a set of open source R packages for population-level estimation implementing several observational study designs: the new-user cohort design with propensity scores, self-controlled case series, case-control, self-controlled cohort, and IC temporal pattern discovery. The packages are actively used in studies into the average effect of exposures on outcomes within the OHDSI community.
The packages interact directly with observational data in the Common Data Model, and are designed to support both large datasets and large numbers of analyses (e.g. for testing many hypotheses including control hypotheses, and testing many analyses design variations). The tools support best practices for use of observational data as learned from previous and ongoing research, such as transparency, reproducibility, as well as measuring of the operating characteristics of methods in a particular context and subsequent empirical calibration of estimates produced by the methods.

Technology

The methods library is implemented as a set of R packages, some of which rely on C++ for speed, or on Java for special features such as database connectivity.

Database connectivity

Each method directly interacts with the data in the CDM model through a JDBC interface. Currently supported platforms include Microsoft SQL Server, Oracle, PostgreSQL, Microsoft PDW, and Amazon RedShift. The SqlRender package allows SQL to be generated on the fly for the various SQL dialects.

Handling large data

Handling of large data inside the methods packages is done using the public ff and ffbase packages, allowing manipulation of datasets that can be much larger than RAM.

Large scale regularized regression

Most OHDSI methods rely on large scale regularized regression, which has been implemented in Cyclops. Cyclops used cyclic coordinate descent for optimization, and supports (conditional) logistic, (conditional) Poisson, and (conditional) Cox regression. Picking the optimal hyperparameter can be done automatically through cross-validation.

Validation

All OHDSI packages rely on unit tests to guarantee the validity of the functionality on a continuous basis. In addition, most methods are tested using extensive simulation.
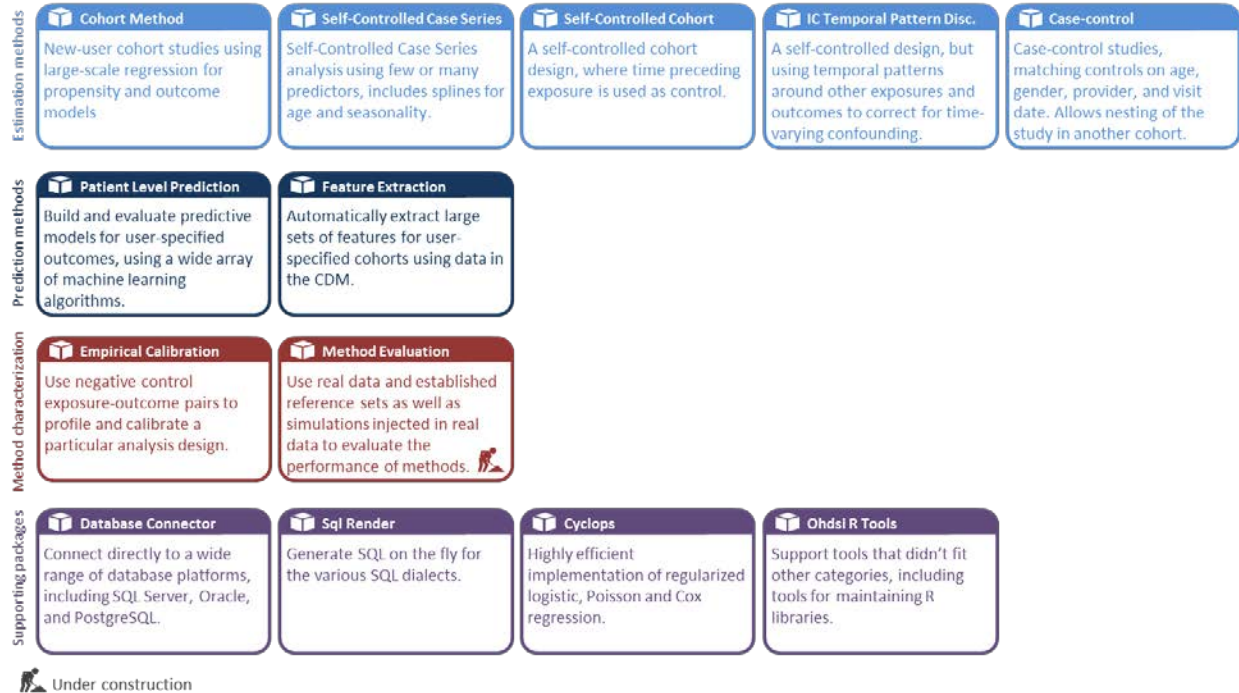
Library packages

	Cohort Method	Self-Controlled Case Series	Self-Controlled Cohort	IC Temporal Pattern Disc.	Case-control
Estimation methods	New-user cohort studies using large-scale regression for propensity and outcome models	Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.	A self-controlled cohort design, where time preceding exposure is used as control.	A self-controlled design, but using temporal patterns around other exposures and outcomes to correct for time-varying confounding.	Case-control studies, matching controls on age, gender, provider, and visit date. Allows nesting of the study in another cohort.

	Patient Level Prediction	Feature Extraction
Prediction methods	Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.	Automatically extract large sets of features for user-specified cohorts using data in the CDM.

	Empirical Calibration	Method Evaluation
Method characterization	Use negative control exposure-outcome pairs to profile and calibrate a particular analysis design.	Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods.

	Database Connector	Sql Render	Cyclops	Ohdsi R Tools
Supporting packages	Connect directly to a wide range of database platforms, including SQL Server, Oracle, and PostgreSQL.	Generate SQL on the fly for the various SQL dialects.	Highly efficient implementation of regularized logistic, Poisson and Cox regression.	Support tools that didn't fit other categories, including tools for maintaining R libraries.

Under construction

Figure 1. Overview of the methods library R packages.

Using the OHDSI Methods Library

All packages in the OHDSI methods library can be found in the OHDSI GitHub repository: https://github.com/ohdsi.
All packages have a package manual describing all the functions and datasets in the package. Most packages also have vignettes which provide step-by-step instructions on how to use the package.

For questions on how to use the OHDSI methods library you can turn to the OHDSI Forums: http://forums.ohdsi.org, or send an e-mail to schuemie@ohdsi.org.