OHDSI Methods Library

Martijn Schuemie, Marc Suchard, Patrick Ryan, Tomas Bergvall
OHDSI Methods Library

• What is it for?
• How does it work?
• How can I use it?
What is it for?

Does exposure to $T$ increase/decrease the risk of $O$?

(and how much?)
What is it for?

Does exposure to $T$ increase/decrease the risk of $O$ when compared to exposure to $C$ (and how much?)
What is it for?

Does exposure to Keppra increase/decrease the risk of Angioedema

(and how much?)
Methods library

- Population-level estimation analysis designs
  - Self-Controlled Case Series
  - New-user cohort method using propensity scores
  - Self-Controlled Cohort
  - IC Temporal Pattern Discovery

- Implemented as open source R packages

- Run against the CDM

- In (almost) any environment
  - Windows, Linux, Mac
  - PostgreSQL, Oracle, SQL Server, Amazon RedShift, Microsoft APS

- Lots of flexibility

- Validated
Self-Controlled Case Series

Is the outcome more likely during exposed time compared to non-exposed time?

Patient A

Keppra

Angioedema

Patient B

Keppra

Patient C

Keppra
Self- Controlled Case Series

By design, adjusted for

• Patient characteristics constant over time

Additionally adjust for

• Age
• Season
• Contra-indication
• (All) other exposures (MSCCS)
• Outcome-dependent censoring
Self-Controlled Case Series

Weakness

• Sensitive to time-varying confounding

Is the time when people take Keppra really comparable to when they don’t?
Cohort Method

Does exposure to Keppra increase/decrease the risk of Angioedema

(and how much?)
Cohort Method

Does exposure to Keppra increase/decrease the risk of Angioedema when compared to exposure to Phenytoin (and how much?)
New-user cohort design

Total population

Keppra

Phenytoin

Initiation of treatment
Randomized controlled trial

Total population

Randomization

Treatment arm

Initiation of treatment

Control arm
New-user cohort design

Total population

Initiation of treatment

Treatment assignment is not random!

Doctors have reasons why they prescribe Keppra to some patients and phenytoin to others
Propensity model

• Statistical model (logistic regression) of why patients get one treatment or the other
• Using (all) information prior to initiation of treatment
• Used to adjust for differences through
  – Trimming
  – Matching
  – Stratification
Outcome model

After trimming / matching / stratification on the propensity score

Estimate effect of treatment on outcome using

• Cox
• Poisson
• Logistic

Include same information prior to initiation of treatment
Under the hood

• **R**
  – Easy to deploy highly optimized code
  – Flexibility (80-20 rule)
  – Can be mastered by epidemiologists

• **DatabaseConnector + RenderSql** for interacting with various database platforms

• **Cyclops** for fitting large regression models
  – $L_1$ and $L_2$ Regularization
  – Cox, Poisson, logistic
  – Conditioned and unconditioned

• Exploiting ‘embarrassingly parallelizability’ for increased performance
Validation

• Unit tests
• Simulations
Unit tests

A unit test is a piece of code that tests a function:

test_that("Simple 1-on-1 matching", {  
  rowId <- 1:5  
  treatment <- c(1, 0, 1, 0, 1)  
  propensityScore <- c(0, 0.1, 0.3, 0.4, 1)  
  data <- data.frame(rowId = rowId,  
                      treatment = treatment,  
                      propensityScore = propensityScore)  
  result <- matchOnPs(data, caliper = 0, maxRatio = 1)  
  expect_equal(result$stratumId, c(0, 0, 1, 1))
})

All unit tests are executed every time a change is made to the package:

Development status

[build passing Codecov 49%]
Simulation

Using simulation for more complicated functionality

E.g: SCCS seasonality modeling:
How can I use it?

https://github.com/OHDSI
### Methods Library R packages

<table>
<thead>
<tr>
<th>Estimation methods</th>
<th>Cohort Method</th>
<th>Self-Controlled Case Series</th>
<th>Self-Controlled Cohort</th>
<th>IC Temporal Pattern Disc.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>New-user cohort studies using large-scale regression for propensity and outcome models</td>
<td>Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.</td>
<td>A self-controlled cohort design, where time preceding exposure is used as control.</td>
<td>A self-controlled design, but using temporal patterns around other exposures and outcomes to correct for time-varying confounding.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prediction methods</th>
<th>Feature Extraction</th>
<th>Patient Level Prediction</th>
<th>Empirical Calibration</th>
<th>Method Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Automatically extract large sets of features for user-specified cohorts using data in the CDM.</td>
<td>Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.</td>
<td>Use negative control exposure-outcome pairs (where relative risk is assumed to be 1) to profile and calibrate a particular analysis design.</td>
<td>Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method characterization</th>
<th>Database Connector</th>
<th>Sql Render</th>
<th>Cyclops</th>
<th>Ohdsi R Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Connect directly to a wide range of database platforms, including SQL Server, Oracle, and PostgreSQL.</td>
<td>Generate SQL on the fly for the various SQL dialects.</td>
<td>Highly efficient implementation of regularized logistic, Poisson and Cox regression.</td>
<td>Support tools that didn’t fit other categories, including tools for maintaining R libraries.</td>
</tr>
</tbody>
</table>

**Supporting packages**

- **Cohort Method**
- **Self-Controlled Case Series**
- **Self-Controlled Cohort**
- **IC Temporal Pattern Disc.**
- **Feature Extraction**
- **Patient Level Prediction**
- **Empirical Calibration**
- **Method Evaluation**
- **Database Connector**
- **Sql Render**
- **Cyclops**
- **Ohdsi R Tools**

**Under construction**
CohortMethod package

```r
cmd <- getDbCohortMethodData(connectionDetails, 
    cdmDatabaseSchema = cdmSchema, 
    targetId = 1118084, 
    comparatorId = 1124300, 
    outcomeId = 192671, 
    washoutPeriod = 183, 
    firstExposureOnly = TRUE, 
    removeDuplicateSubjects = TRUE, 
    excludeDrugsFromCovariates = TRUE, 
    covariateSettings = createCovariateSettings())

studyPop <- createStudyPopulation(cohortMethodData = cmd, 
    outcomeId = 192671, 
    removeSubjectsWithPriorOutcome = TRUE, 
    minDaysAtRisk = 1, 
    riskWindowStart = 0, 
    riskWindowEnd = 30, 
    addExposureDaysToEnd = TRUE)

ps <- createPs(cmd, studyPop)
plotPs(ps)
stratPop <- matchOnPs(ps, 
    caliper = 0.25, 
    caliperScale = "standardized", 
    maxRatio = 1)
plotPs(stratPop, ps)
balance <- computeCovariateBalance(strata, cmd)
plotCovariateBalanceScatterPlot(balance)
plotCovariateBalanceOfTopVariables(balance)
outcomeModel <- fitOutcomeModel(stratPop, 
    useCovariates = TRUE, 
    modelType = "cox", 
    stratified = TRUE)
plotKaplanMeier(stratPop, includeZero = FALSE)
drawAttritionDiagram(stratPop)
outcomeModel
```
CohortMethod package

These 13 statements

- Implement a full study
- Interact directly with database in CDM
- Many covariates constructed
  - Demographics
  - Every drug (+ class)
  - Every condition (+ group)
  - Every procedure
  - Every observation
  - Charleston, CHAD2, etc.
- Propensity model using LASSO
- 1-on-1 matching on propensity score
- Fitting a Cox model
  - including same covariates used in PS model
CohortMethod package

```r
cmd <- getDbCohortMethodData(connectionDetails,
  cdmDatabaseSchema = cdmSchema,
  targetId = 1118084,
  comparatorId = 1124300,
  outcomeId = 192671,
  washoutPeriod = 183,
  firstExposureOnly = TRUE,
  removeDuplicateSubjects = TRUE,
  excludeDrugsFromCovariates = TRUE,
  covariateSettings = createCovariateSettings())

studyPop <- createStudyPopulation(cohortMethodData = cmd,
  outcomeId = 192671,
  removeSubjectsWithPriorOutcome = TRUE,
  minDaysAtRisk = 1,
  riskWindowStart = 0,
  riskWindowEnd = 30,
  addExposureDaysToEnd = TRUE)

ps <- createPs(cmd, studyPop)
plotPs(ps)

stratPop <- matchOnPs(ps,
  caliper = 0.25,
  caliperScale = "standardized",
  maxRatio = 1)

plotPs(stratPop, ps)
balance <- computeCovariateBalance(strata, cmd)
plot Covariate Balance Scatter Plot(balance)
plot Covariate Balance Of Top Variables(balance)
outcomeModel <- fitOutcomeModel(stratPop,
  cmd
  useCovariates = TRUE,
  modelType = "cox",
  stratified = TRUE)

plotKaplanMeier(stratPop, includeZero = FALSE)
drawAttritionDiagram(stratPop)
outcomeModel
```

Specify the two exposure groups and the outcome.

Typically using Circe-defined cohorts
CohortMethod	
  
package

```r
cmd <- getDbCohortMethodData(connectionDetails, cdmDatabaseSchema = cdmSchema,
  targetId = 1118084, comparatorId = 1124300, outcomeId = 192671,
  washoutPeriod = 183, firstExposureOnly = TRUE, removeDuplicateSubjects = TRUE,
  excludeDrugsFromCovariates = TRUE, covariateSettings = createCovariateSettings())

studyPop <- createStudyPopulation(cohortMethodData = cmd, outcomeId = 192671,
  removeSubjectsWithPriorOutcome = TRUE, minDaysAtRisk = 1, riskWindowStart = 0,
  riskWindowEnd = 30, addExposureDaysToEnd = TRUE)

ps <- createPs(cmd, studyPop)
plotPs(ps)
stratPop <- matchOnPs(ps, caliper = 0.25, caliperScale = "standardized",
  maxRatio = 1)
plotPs(stratPop, ps)
balance <- computeCovariateBalance(strata, cmd)
plotCovariateBalanceScatterPlot(balance)
plotCovariateBalanceOfTopVariables(balance)
outcomeModel <- fitOutcomeModel(stratPop, cmd, useCovariates = TRUE,
  modelType = "cox", stratified = TRUE)
plotKaplanMeier(stratPop, includeZero = FALSE)
drawAttritionDiagram(stratPop)
outcomeModel

HR = ...
```
All-by-all support

For:
- Sensitivity analyses
- Including negative controls
- Methods research
- Safety surveillance?
Learning more...

• All packages have manuals describing the functions
• Most packages have vignettes that describe its use
• Marc and Martijn are organizing a travelling roadshow to teach CohortMethod
  – One stop: OHDSI Symposium