



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?)



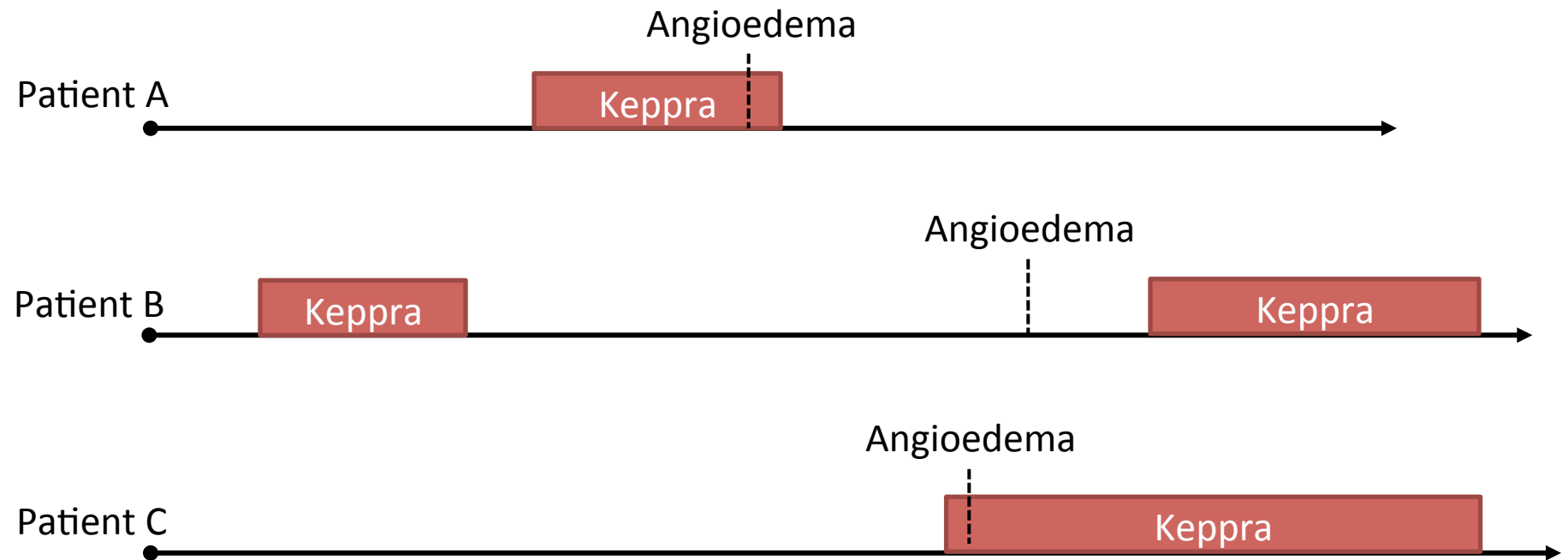
Methods library

- 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?





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

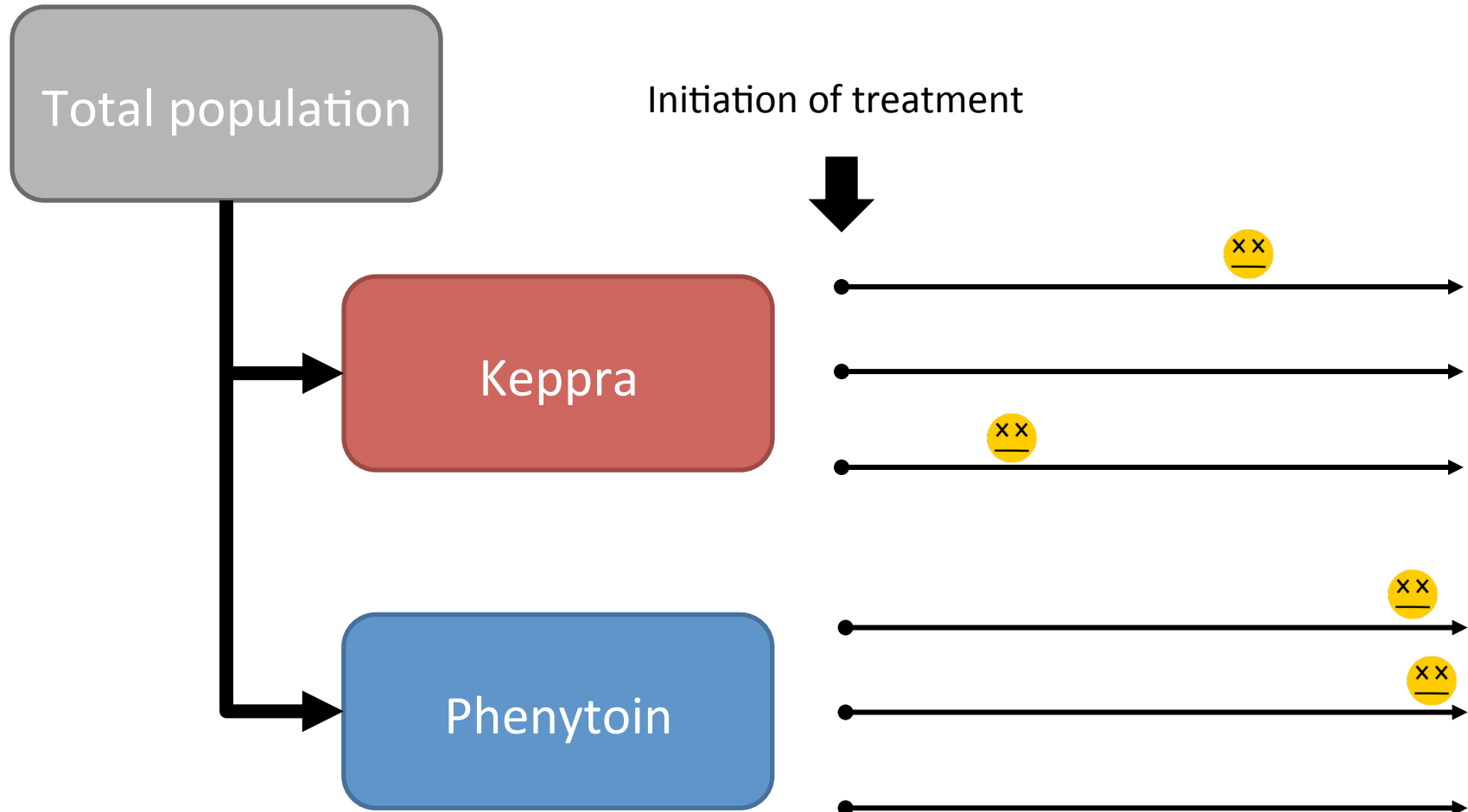
when compared to exposure to

Phenytoin

(and how much?)

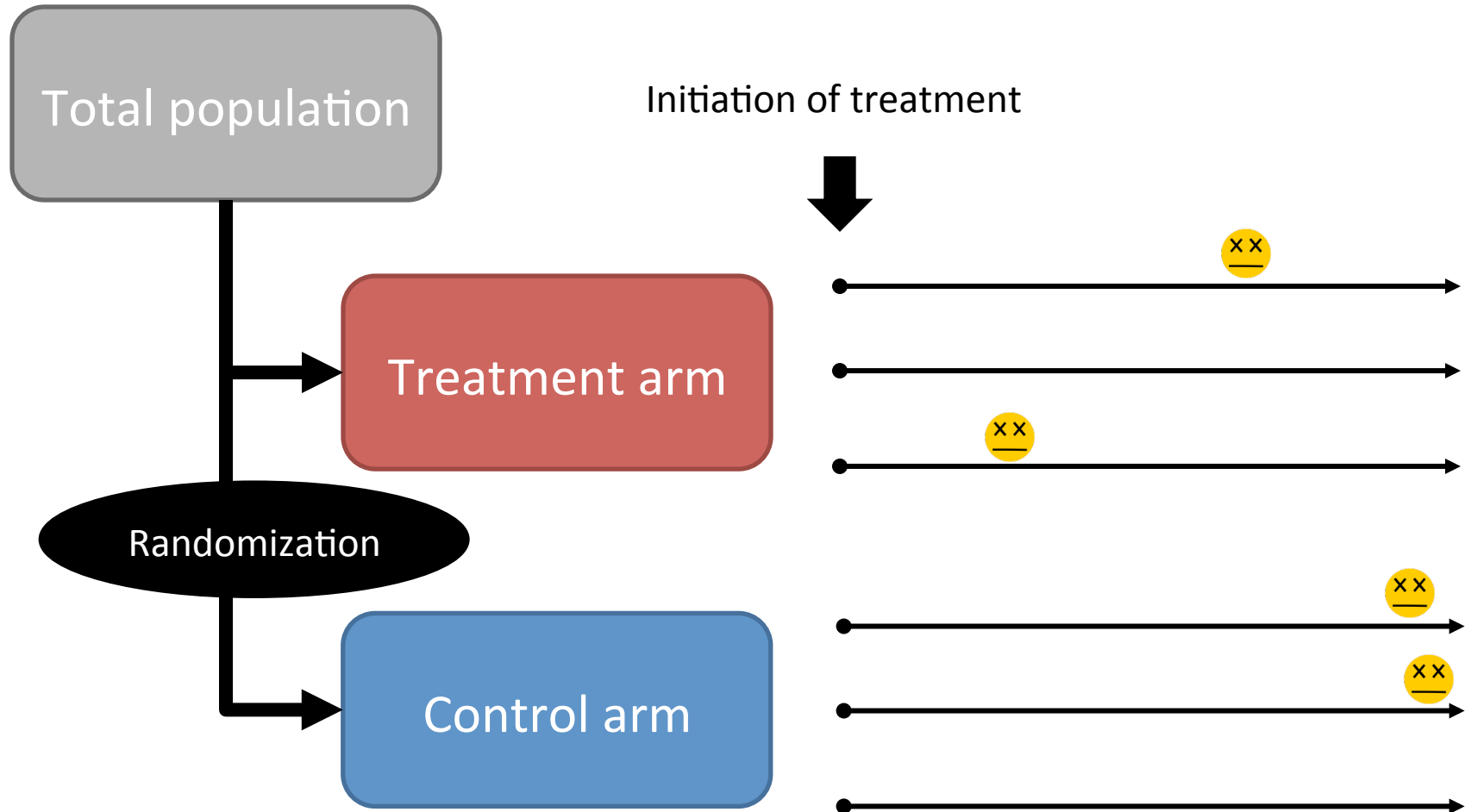


New-user cohort design





Randomized controlled trial





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

Phenytoin





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, 1))  
})
```

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

Development status

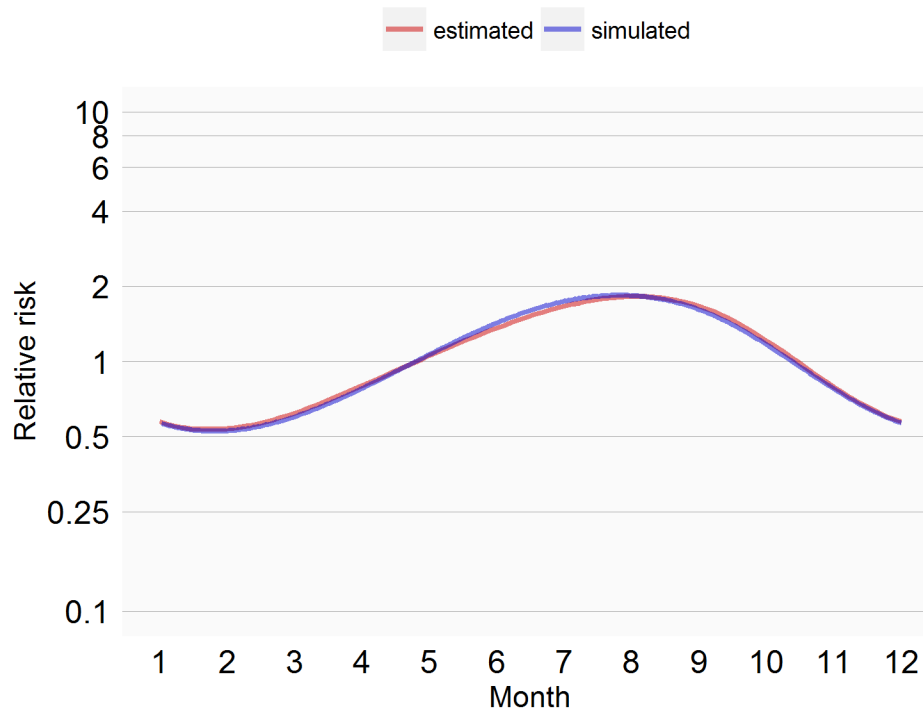
build passing Codecov 49%



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

Estimation methods

Cohort Method

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


Self-Controlled Case Series

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

Self-Controlled Cohort

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

IC Temporal Pattern Disc.

A self-controlled design, but using temporal patterns around other exposures and outcomes to correct for time-varying confounding. 

Prediction methods

Feature Extraction

Automatically extract large sets of features for user-specified cohorts using data in the CDM.

Patient Level Prediction


Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.

Method characterization

Empirical Calibration

Use negative control exposure-outcome pairs (where relative risk is assumed to be 1) to profile and calibrate a particular analysis design.

Method Evaluation

Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods. 

Supporting packages

Database Connector

Connect directly to a wide range of database platforms, including SQL Server, Oracle, and PostgreSQL.

Sql Render

Generate SQL on the fly for the various SQL dialects.

Cyclops

Highly efficient implementation of regularized logistic, Poisson and Cox regression.

Ohdsi R Tools

Support tools that didn't fit other categories, including tools for maintaining R libraries.



Under construction



CohortMethod package

```
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
```



CohortMethod package

```
cmd <- getDbCohortMethodData(connectionDe
                                cdmDatabas
                                targetId =
                                comparator
                                outcomeId
                                washoutPer
                                firstExpos
                                removeDupl
                                excludeDru
                                covariates
                                addEx

studyPop <- createStudyPopulation(cohor
                                outco
                                remov
                                minDa
                                riskv
                                riskv
                                addEx

ps <- createPs(cmd, studyPop)
plotPs(ps)
stratPop <- matchOnPs(ps,
                    caliper = 0.25,
                    caliperScale = "
                    maxRatio = 1)

plotPs(stratPop, ps)
balance <- computeCovariateBalance(strat
plotCovariateBalanceScatterPlot(balance
plotCovariateBalanceOfTopVariables(bala
outcomeModel <- fitOutcomeModel(stratPop
                                cmd
                                useCovariates = TRUE,
                                modelType = "cox",
                                stratified = TRUE)

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

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

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

Specify the two exposure groups and the outcome.

Typically using Circe-defined cohorts

```
plotPs(ps)  
stratPop <- matchOnPs(ps,  
                     caliper = 0.25,  
                     caliperscale = "standard",  
                     maxRatio = 1)  
  
plotPs(stratPop, ps)  
balance <- computeCovariateBalance(stratPop, ps)  
plotCovariateBalanceScatterPlot(balance)  
plotCovariateBalanceOfTopVariables(balance)  
outcomeModel <- fitOutcomeModel(stratPop, ps, cmd)  
useCovariateModelType("stratified")  
plotKaplanMeier(stratPop, includeZero = FALSE)  
drawAttritionDiagram(stratPop)  
outcomeModel
```

Index Population for Study: keppra: T for jon duke's study of keppra vs. phenytoin and angioedema

Save Copy Delete Show SQL

Description:

Expression Concept Sets Print Friendly Raw JSON Generate

People having any of the following: Add Primary Event Filters...

a drug era of **levetiracetam** Add Filter... Delete Filter

for the first time in the person's history

with observation at least **183** days prior and **0** days after index

Limit primary events to: **earliest event** per person.

For people matching the Primary Events, include:

People having **all** of the following criteria: Add New Criteria...

with at least **1** using all occurrences of:

a condition occurrence of **epilepsy** Add Filter... Delete Filter

occurring between **All** days Before and **0** days Before index

Remove Additional Filters

Limit cohort expression results to: **earliest event** per person.

CohortMethod package

```
cmd <- getDbCohortMethodData(connectionDetails,
                              cdmDatabaseSchema = cdmSchema,
                              targetId = 1118084,
                              comparatorId = 1118085,
                              outcomeId = 1118086,
                              washoutPeriod = 180,
                              firstExposureWindow = 180,
                              removedDuplicateExposures = TRUE,
                              excludeDrugs = NULL,
                              covariates = NULL)

studyPop <- createStudyPopulation(cmd,
                                  outcomeId = 1118086,
                                  removedDuplicateExposures = TRUE,
                                  minDaysAtRisk = 180,
                                  riskWindow = 180,
                                  addExposure = TRUE)

ps <- createPs(cmd, studyPop)
plotPs(ps)
stratPop <- matchOnPs(ps,
                      caliper = 0.25,
                      caliperscale = "standardized")
```

```
plotPs(ps)
stratPop <- matchOnPs(ps,
                      caliper = 0.25,
                      caliperscale = "standardized")
```

HR = ...

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

Original cohorts:
Treated: n = 73768
Comparator: n = 279333

First exp. only & removed
subs in both cohorts & 180
days of obs. prior

No prior outcome

Have time at risk

Treated: n = 39356
Comparator: n = 98664

Treated: n = 1177
Comparator: n = 4265

Treated: n = 0
Comparator: n = 0

Treated: n = 3391
Comparator: n = 146560

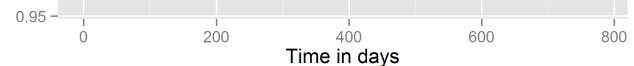


GROUP

difference of mean

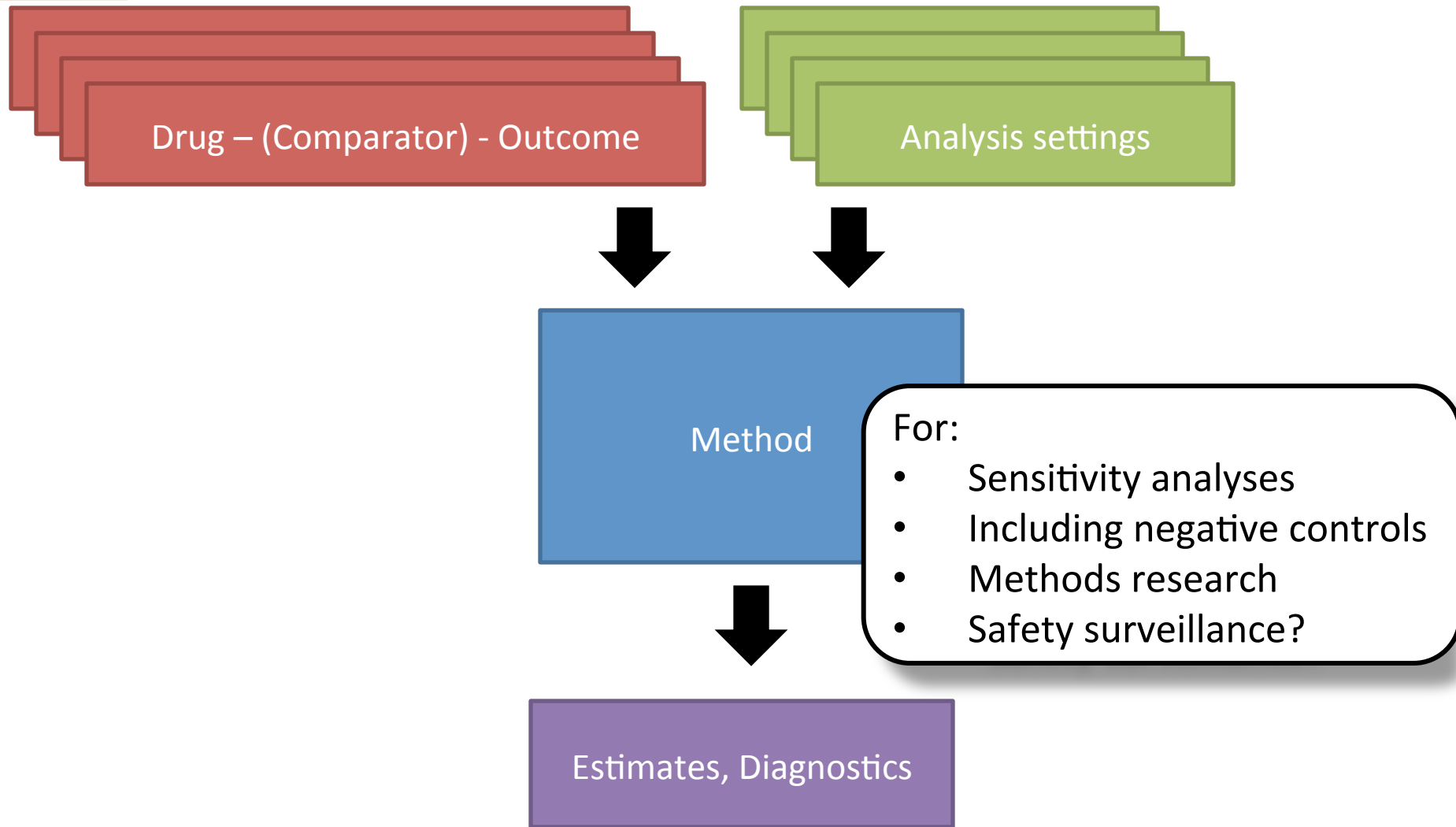
• before matching • after matching

Treated
Comparator





All-by-all support





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

