

# Package ‘MethodEvaluation’

August 27, 2019

**Type** Package

**Title** Package for evaluation of estimation methods

**Version** 1.1.0

**Date** 2019-08-27

**Maintainer** Martijn J. Schuemie <schuemie@ohdsi.org>

**Description** This package contains resources for the evaluation of the performance of methods that aim to estimate the magnitude (relative risk) of the effect of a drug on an outcome. These resources include reference sets for evaluating methods on real data, as well as functions for inserting simulated effects in real data based on negative control drug-outcome pairs. Further included are functions for the computation of the minimum detectable relative risks and functions for computing performance statistics such as predictive accuracy, error and bias.

**License** Apache License 2.0

**VignetteBuilder** knitr

**URL** <https://github.com/OHDSI/MethodEvaluation>

**BugReports** <https://github.com/OHDSI/MethodEvaluation/issues>

**Depends** R (>= 3.2.0),  
DatabaseConnector (>= 2.0.0),  
FeatureExtraction (>= 2.0.0),  
Cyclops (>= 1.2.2)

**Imports** ff,  
ffbase (>= 0.12.1),  
SqlRender (>= 1.6.2),  
pROC,  
ggplot2,  
ParallelLogger,  
methods,  
EmpiricalCalibration,  
rmarkdown

**Suggests** testthat,  
DT,  
shiny,  
knitr

**RoxygenNote** 6.1.1

**Encoding** UTF-8

## R topics documented:

checkCohortSourceCodes . . . . .	2
computeMdr . . . . .	3
computeMetrics . . . . .	4
computeOhdsiBenchmarkMetrics . . . . .	5
createReferenceSetCohorts . . . . .	6
euadrReferenceSet . . . . .	7
findOrphanSourceCodes . . . . .	8
injectSignals . . . . .	8
launchMethodEvaluationApp . . . . .	11
MethodEvaluation . . . . .	12
ohdsiNegativeControls . . . . .	12
omopReferenceSet . . . . .	13
packageOhdsiBenchmarkResults . . . . .	14
plotControls . . . . .	15
plotCoverageInjectedSignals . . . . .	15
plotRocsInjectedSignals . . . . .	16
synthesizePositiveControls . . . . .	17
<b>Index</b>	<b>22</b>

---

checkCohortSourceCodes

*Check source codes used in a cohort definition*

---

### Description

This function first extracts all concept sets used in a cohort definition. Then, for each concept set the concept found in the CDM database the contributing source codes are identified. Overall and per month counts are computed and shown for all concept sets, concepts, and source codes in a HTML table.

This function requires the rmarkdown package as well as PanDoc.

### Usage

```
checkCohortSourceCodes(connectionDetails, cdmDatabaseSchema,
  oracleTempSchema = NULL, cohortJson, cohortSql, outputFile)
```

### Arguments

connectionDetails

An object of type connectionDetails as created using the [createConnectionDetails](#) function in the DatabaseConnector package.

cdmDatabaseSchema

Schema name where your patient-level data in OMOP CDM format resides. Note that for SQL Server, this should include both the database and schema name, for example 'cdm\_data.dbo'.

oracleTempSchema

Should be used in Oracle to specify a schema where the user has write privileges for storing temporary tables.

cohortJson	A characteric string containing the JSON of a cohort definition.
cohortSql	The OHDSI SQL representation of the same cohort definition.
outputFile	The name of the HTML file to create.

---

computeMdr	<i>Compute minimal detectable relative risk (MDRR)</i>
------------	--

---

## Description

computeMdr computes the minimal detectable relative risk (MDRR) for drug-outcome pairs using a standard approach that stratifies by age and gender (Armstrong 1987).

## Usage

```
computeMdr(connectionDetails, cdmDatabaseSchema,
  oracleTempSchema = cdmDatabaseSchema, exposureOutcomePairs,
  exposureDatabaseSchema = cdmDatabaseSchema,
  exposureTable = "drug_era",
  outcomeDatabaseSchema = cdmDatabaseSchema,
  outcomeTable = "condition_era", cdmVersion = "5")
```

## Arguments

connectionDetails	An R object of type ConnectionDetails created using the function createConnectionDetails in the DatabaseConnector package.
cdmDatabaseSchema	Name of database schema that contains OMOP CDM and vocabulary.
oracleTempSchema	For Oracle only: the name of the database schema where you want all temporary tables to be managed. Requires create/insert permissions to this database.
exposureOutcomePairs	A data frame with at least two columns: <ul style="list-style-type: none"> <li>"exposureId" or "targetId" containing the drug_concept_ID or cohort_definition_id of the exposure variable</li> <li>"outcomeId" containing the condition_concept_ID or cohort_definition_id of the outcome variable</li> </ul>
exposureDatabaseSchema	The name of the database schema that is the location where the exposure data used to define the exposure cohorts is available. If exposureTable = DRUG_ERA, exposureDatabaseSchema is not used by assumed to be cdmSchema. Requires read permissions to this database.
exposureTable	The tablename that contains the exposure cohorts. If exposureTable <> DRUG_ERA, then expectation is exposureTable has format of COHORT table: COHORT_DEFINITION_ID, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.
outcomeDatabaseSchema	The name of the database schema that is the location where the data used to define the outcome cohorts is available. If exposureTable = CONDITION_ERA, exposureDatabaseSchema is not used by assumed to be cdmSchema. Requires read permissions to this database.

outcomeTable	The tablename that contains the outcome cohorts. If outcomeTable <> CONDITION_OCCURRENCE, then expectation is outcomeTable has format of COHORT table: COHORT_DEFINITION_ID, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.
cdmVersion	Define the OMOP CDM version used: currently support "4" and "5".

### Value

A data frame containing the MDRRs for the given exposure-outcome pairs.

### References

Armstrong B. A simple estimator of minimum detectable relative risk, sample size, or power in cohort studies. American journal of epidemiology. 1987; 126: 356-8.

### Examples

```
## Not run:
connectionDetails <- createConnectionDetails(dbms = "sql server",
                                             server = "RNDUSRDHIT07.jnj.com")
exposureOutcomePairs <- data.frame(exposureId = c(767410, 1314924, 907879),
                                   outcomeId = c(444382, 79106, 138825))
mdrrs <- computeMdrd(connectionDetails,
                    "cdm_truven_mdc",
                    exposureOutcomePairs,
                    outcomeTable = "condition_era")

## End(Not run)
```

---

computeMetrics	<i>Compute method performance metrics</i>
----------------	---

---

### Description

Compute method performance metrics

### Usage

```
computeMetrics(logRr, seLogRr = NULL, ci95Lb = NULL, ci95Ub = NULL,
               p = NULL, trueLogRr)
```

### Arguments

logRr	A numeric vector of effect estimates on the log scale.
seLogRr	The standard error of the log of the effect estimates. Hint: often the standard error = (log(<lower bound 95 percent confidence interval>) - log(<effect estimate>))/qnorm(0.025). If not provided the standard error will be inferred from the 95 percent confidence interval.
ci95Lb	The lower bound of the 95 percent confidence interval. IF not provided it will be inferred from the standard error.
ci95Ub	The upper bound of the 95 percent confidence interval. IF not provided it will be inferred from the standard error.

p	The two-sided p-value corresponding to the null hypothesis of no effect. IF not provided it will be inferred from the standard error.
trueLogRr	A vector of the true effect sizes

### Details

Compute the AUC, coverage, mean precision, MSE, type 1 error, type 2 error, and the fraction non-estimable..

### Examples

```
library(EmpiricalCalibration)
data <- simulateControls(n = 50 * 3, trueLogRr = log(c(1, 2, 4)))
computeMetrics(logRr = data$logRr, seLogRr = data$seLogRr, trueLogRr = data$trueLogRr)
```

---

```
computeOhdsiBenchmarkMetrics
```

*Generate performance metrics for the OHDSI Methods Benchmark*

---

### Description

Generate performance metrics for the OHDSI Methods Benchmark

### Usage

```
computeOhdsiBenchmarkMetrics(exportFolder, mdr = 1.25,
  stratum = "All", trueEffectSize = "Overall", calibrated = FALSE,
  comparative = FALSE)
```

### Arguments

exportFolder	The folder containing the CSV files created using the <a href="#">packageOhdsiBenchmarkResults</a> function. This folder can contain results from various methods, analyses, and databases.
mdr	The minimum detectable relative risk (MDRR). Only controls with this MDRR will be used to compute the performance metrics. Set to "All" to include all controls.
stratum	The stratum for which to compute the metrics, e.g. 'Acute Pancreatitis'. Set to 'All' to use all controls.
trueEffectSize	Should the analysis be limited to a specific true effect size? Set to "Overall" to include all.
calibrated	Should confidence intervals and p-values be empirically calibrated before computing the metrics?
comparative	Should the methods be evaluated on the task of comparative effect estimation? If FALSE, they will be evaluated on the task of effect estimation.

### Value

A data frame with the various metrics per method - analysisId - database combination.

---

```
createReferenceSetCohorts
```

*Create cohorts used in a reference set.*

---

## Description

Create cohorts used in a reference set.

## Usage

```
createReferenceSetCohorts(connectionDetails, oracleTempSchema = NULL,
  cdmDatabaseSchema, outcomeDatabaseSchema = cdmDatabaseSchema,
  outcomeTable = "outcomes", nestingDatabaseSchema = cdmDatabaseSchema,
  nestingTable = "nesting", referenceSet = "ohdsiMethodsBenchmark")
```

## Arguments

`connectionDetails`

An R object of type `ConnectionDetails` created using the function `createConnectionDetails` in the `DatabaseConnector` package.

`oracleTempSchema`

Should be used in Oracle to specify a schema where the user has write privileges for storing temporary tables.

`cdmDatabaseSchema`

A database schema containing health care data in the OMOP Common Data Model. Note that for SQL Server, both the database and schema should be specified, e.g. `'cdm_schema.dbo'`

`outcomeDatabaseSchema`

The database schema where the target outcome table is located. Note that for SQL Server, both the database and schema should be specified, e.g. `'cdm_schema.dbo'`

`outcomeTable`

The name of the table where the outcomes will be stored.

`nestingDatabaseSchema`

(For the OHDSI Methods Benchmark only) The database schema where the nesting outcome table is located. Note that for SQL Server, both the database and schema should be specified, e.g. `'cdm_schema.dbo'`.

`nestingTable`

(For the OHDSI Methods Benchmark only) The name of the table where the nesting cohorts will be stored.

`referenceSet`

The name of the reference set for which outcomes need to be created. Currently supported are `"omopReferenceSet"`, `"euadrReferenceSet"`, and `"ohdsiMethods-Benchmark"`.

## Details

This function will create the outcomes of interest and nesting cohorts referenced in the various reference sets. The outcomes of interest are derived using information like diagnoses, procedures, and drug prescriptions. The outcomes are stored in a table on the database server.

---

euadrReferenceSet	<i>The EU-ADR reference set A reference set of 43 drug-outcome pairs where we believe the drug causes the outcome ( positive controls) and 50 drug-outcome pairs where we believe the drug does not cause the outcome (negative controls). The controls involve 10 health outcomes of interest. Note that originally, there was an additional positive control (Nimesulide and acute liver injury), but Nimesulide is not in RxNorm, and is not available in many countries.</i>
-------------------	--

---

## Description

The EU-ADR reference set A reference set of 43 drug-outcome pairs where we believe the drug causes the outcome ( positive controls) and 50 drug-outcome pairs where we believe the drug does not cause the outcome (negative controls). The controls involve 10 health outcomes of interest. Note that originally, there was an additional positive control (Nimesulide and acute liver injury), but Nimesulide is not in RxNorm, and is not available in many countries.

## Usage

```
data(euadrReferenceSet)
```

## Format

A data frame with 399 rows and 10 variables:

**exposureId** Concept ID identifying the exposure

**exposureName** Name of the exposure

**outcomeId** Concept ID identifying the outcome

**outcomeName** Name of the outcome

**groundTruth** 0 = negative control, 1 = positive control

**indicationId** Concept Id identifying the (primary) indication of the drug. To be used when one wants to nest the analysis within the indication

**indicationName** Name of the indication

**comparatorId** Concept ID identifying a comparator drug that can be used as a counterfactual

**comparatorName** Name of the comparator drug

**comparatorType** How the comparator was selected

## References

Coloma PM, Avillach P, Salvo F, Schuemie MJ, Ferrajolo C, Pariente A, Fourier-Reglat A, Molokhia M, Patadia V, van der Lei J, Sturkenboom M, Trifiro G. A reference standard for evaluation of methods for drug safety signal detection using electronic healthcare record databases. *Drug Safety* 36(1):13-23, 2013

---

findOrphanSourceCodes	<i>Find source codes that do not roll up to their ancestor</i>
-----------------------	--

---

### Description

Find source codes that do not roll up to their ancestor

### Usage

```
findOrphanSourceCodes(connectionDetails, cdmDatabaseSchema,
  oracleTempSchema = NULL, conceptName, conceptSynonyms = NULL)
```

### Arguments

connectionDetails	An object of type connectionDetails as created using the <a href="#">createConnectionDetails</a> function in the DatabaseConnector package.
cdmDatabaseSchema	Schema name where your patient-level data in OMOP CDM format resides. Note that for SQL Server, this should include both the database and schema name, for example 'cdm_data.dbo'.
oracleTempSchema	Should be used in Oracle to specify a schema where the user has write privileges for storing temporary tables.
conceptName	The exact name of the parent concept.
conceptSynonyms	Synonyms by which the code might be described.

### Details

Searches for concepts where the name contains the concept name of interest, or any of its synonyms as provided by the user. Then it checks whether these concepts roll up to the main concept.

### Value

A data frame with orphan source codes, with counts per domain how often the code was encountered in the CDM.

---

injectSignals	<i>Inject signals in database</i>
---------------	-----------------------------------

---

### Description

Inject signals in database



**Usage**

```
injectSignals(connectionDetails, cdmDatabaseSchema,
  oracleTempSchema = cdmDatabaseSchema,
  exposureDatabaseSchema = cdmDatabaseSchema,
  exposureTable = "drug_era",
  outcomeDatabaseSchema = cdmDatabaseSchema, outcomeTable = "cohort",
  outputDatabaseSchema = outcomeDatabaseSchema,
  outputTable = outcomeTable, createOutputTable = FALSE,
  exposureOutcomePairs, modelType = "poisson",
  minOutcomeCountForModel = 100, minOutcomeCountForInjection = 25,
  covariateSettings = FeatureExtraction::createCovariateSettings(useDemographicsAgeGroup
    = TRUE, useDemographicsGender = TRUE, useDemographicsIndexYear = TRUE,
    useDemographicsIndexMonth = TRUE, useConditionGroupEraLongTerm = TRUE,
    useDrugGroupEraLongTerm = TRUE, useProcedureOccurrenceLongTerm = TRUE,
    useMeasurementLongTerm = TRUE, useObservationLongTerm = TRUE,
    useCharlsonIndex = TRUE, useDcsi = TRUE, useChads2Vasc = TRUE,
    longTermStartDays = 365, endDays = 0),
  prior = Cyclops::createPrior("laplace", exclude = 0, useCrossValidation
    = TRUE), control = Cyclops::createControl(cvType = "auto",
  startingVariance = 0.1, noiseLevel = "quiet", threads = 10),
  firstExposureOnly = FALSE, washoutPeriod = 183,
  riskWindowStart = 0, riskWindowEnd = 0,
  addExposureDaysToEnd = TRUE, addIntentToTreat = FALSE,
  firstOutcomeOnly = FALSE, removePeopleWithPriorOutcomes = FALSE,
  maxSubjectsForModel = 1e+05, effectSizes = c(1, 1.25, 1.5, 2, 4),
  precision = 0.01, outputIdOffset = 1000,
  workFolder = "./SignalInjectionTemp", cdmVersion = "5",
  modelThreads = 1, generationThreads = 1)
```

**Arguments**

**connectionDetails**

An R object of type `ConnectionDetails` created using the function `createConnectionDetails` in the `DatabaseConnector` package.

**cdmDatabaseSchema**

Name of database schema that contains OMOP CDM and vocabulary.

**oracleTempSchema**

For Oracle only: the name of the database schema where you want all temporary tables to be managed. Requires create/insert permissions to this database.

**exposureDatabaseSchema**

The name of the database schema that is the location where the exposure data used to define the exposure cohorts is available. If `exposureTable = DRUG_ERA`, `exposureDatabaseSchema` is not used by assumed to be `cdmSchema`. Requires read permissions to this database.

**exposureTable**

The table name that contains the exposure cohorts. If `exposureTable <> DRUG_ERA`, then expectation is `exposureTable` has format of COHORT table: `cohort_concept_id`, `SUBJECT_ID`, `COHORT_START_DATE`, `COHORT_END_DATE`.

**outcomeDatabaseSchema**

The name of the database schema that is the location where the data used to define the outcome cohorts is available. If `exposureTable = CONDITION_ERA`, `exposureDatabaseSchema` is not used by assumed to be `cdmSchema`. Requires read permissions to this database.

outcomeTable	The table name that contains the outcome cohorts. When the table name is not <code>CONDITION_ERA</code> This table is expected to have the same format as the <code>COHORT</code> table: <code>SUBJECT_ID</code> , <code>COHORT_START_DATE</code> , <code>COHORT_END_DATE</code> , <code>COHORT_CONCEPT_ID</code> (CDM v4) or <code>COHORT_DEFINITION_ID</code> (CDM v5 and higher).
outputDatabaseSchema	The name of the database schema that is the location of the tables containing the new outcomesRequires write permissions to this database.
outputTable	The name of the table names that will contain the generated outcome cohorts.
createOutputTable	Should the output table be created prior to inserting the outcomes? If <code>TRUE</code> and the tables already exists, it will first be deleted. If <code>FALSE</code> , the table is assumed to exist and the outcomes will be inserted. Any existing outcomes with the same IDs will first be deleted.
exposureOutcomePairs	A data frame with at least two columns: <ul style="list-style-type: none"> <li>"exposureId" containing the <code>drug_concept_ID</code> or <code>cohort_concept_id</code> of the exposure variable</li> <li>"outcomeId" containing the <code>condition_concept_ID</code> or <code>cohort_concept_id</code> of the outcome variable</li> </ul>
modelType	Can be either "poisson" or "survival"
minOutcomeCountForModel	Minimum number of outcome events required to build a model.
minOutcomeCountForInjection	Minimum number of outcome events required to inject a signal.
covariateSettings	An object of type <code>covariateSettings</code> as created using the <code>createCovariateSettings</code> function in the <code>FeatureExtraction</code> package.
prior	The prior used to fit the outcome model. See <a href="#">createPrior</a> for details.
control	The control object used to control the cross-validation used to determine the hyperparameters of the prior (if applicable). See <a href="#">createControl</a> for details.
firstExposureOnly	Should signals be injected only for the first exposure? (ie. assuming an acute effect)
washoutPeriod	Number of days at the start of observation for which no signals will be injected, but will be used to determine whether exposure or outcome is the first one, and for extracting covariates to build the outcome model.
riskWindowStart	The start of the risk window relative to the start of the exposure (in days). When 0, risk is assumed to start on the first day of exposure.
riskWindowEnd	The end of the risk window relative to the start of the exposure. Note that typically the length of exposure is added to this number (when the <code>addExposureDaysToEnd</code> parameter is set to <code>TRUE</code> ).
addExposureDaysToEnd	Should length of exposure be added to the risk window?
addIntentToTreat	If true, the signal will not only be injected in the primary time at risk, but also after the time at risk (up until the obseration period end). In both time periods,

the target effect size will be enforced. This allows the same positive control synthesis to be used in both on treatment and intent-to-treat analysis variants. However, this will preclude the controls to be used in self-controlled designs that consider the time after exposure. Requires `firstExposureOnly = TRUE`.

<code>firstOutcomeOnly</code>	Should only the first outcome per person be considered when modeling the outcome?
<code>removePeopleWithPriorOutcomes</code>	Remove people with prior outcomes?
<code>maxSubjectsForModel</code>	Maximum number of people used to fit an outcome model.
<code>effectSizes</code>	A numeric vector of effect sizes that should be inserted.
<code>precision</code>	The allowed ratio between target and injected signal size.
<code>outputIdOffset</code>	What should be the first new outcome ID that is to be created?
<code>workFolder</code>	Path to a folder where intermediate data will be stored.
<code>cdmVersion</code>	Define the OMOP CDM version used: currently support "4" and "5".
<code>modelThreads</code>	Number of parallel threads to use when fitting outcome models.
<code>generationThreads</code>	Number of parallel threads to use when generating outcomes.
	A data.frame listing all the drug-pairs in combination with requested effect sizes and the real inserted effect size (might be different from the requested effect size because of sampling error).

## Details

DEPRECATED. Use [synthesizePositiveControls](#) instead.

---

launchMethodEvaluationApp

*Launch the Method Evaluation Shiny app*

---

## Description

Launch the Method Evaluation Shiny app

## Usage

```
launchMethodEvaluationApp(exportFolder, launch.browser = TRUE)
```

## Arguments

<code>exportFolder</code>	A folder where the data files for the Method Evaluation app are stored. Use the <a href="#">packageOhdsiBenchmarkResults</a> function to populate this folder.
<code>launch.browser</code>	Should the app be launched in your default browser, or in a Shiny window. Note: copying to clipboard will not work in a Shiny window.

## Details

Launches a Shiny app that allows the user to explore the results.

MethodEvaluation	MethodEvaluation
<b>Description</b>	
MethodEvaluation	
ohdsiNegativeControls	<p><i>The OHDSI Method Evaluation Benchmark - Negative Controls A set of 200 negative controls, centered around four outcomes of interest (acute pancreatitis, GI bleeding, Stroke, and IBD), and 4 exposures of interest (diclofenac, ciprofloxacin, metformin, and sertraline), which 25 negative controls each. Each drug-outcome pair also includes a comparator drug (where the comparator is also a negative control), allowing for evaluation of comparative effect estimation, and a nesting cohort for evaluating methods such as the nested case-control design. The exposure, outcome, and nesting cohorts can be created using the <a href="#">createReferenceSetCohorts</a> function. These negative controls can form the basis to generate positive controls using the <a href="#">injectSignals</a> function.</i></p>

## Description

The OHDSI Method Evaluation Benchmark - Negative Controls A set of 200 negative controls, centered around four outcomes of interest (acute pancreatitis, GI bleeding, Stroke, and IBD), and 4 exposures of interest (diclofenac, ciprofloxacin, metformin, and sertraline), which 25 negative controls each. Each drug-outcome pair also includes a comparator drug (where the comparator is also a negative control), allowing for evaluation of comparative effect estimation, and a nesting cohort for evaluating methods such as the nested case-control design. The exposure, outcome, and nesting cohorts can be created using the [createReferenceSetCohorts](#) function. These negative controls can form the basis to generate positive controls using the [injectSignals](#) function.

## Usage

```
data(ohdsiNegativeControls)
```

## Format

A data frame with 200 rows and 9 variables:

**targetId** Cohort ID identifying the target exposure  
**targetName** Name of the target cohort  
**comparatorId** Cohort ID identifying the comparator exposure  
**comparatorName** Name of the comparator cohort  
**nestingId** Cohort ID identifying the nesting cohort  
**nestingName** Name of the nesting cohort  
**outcomeId** Cohort ID identifying the outcome  
**outcomeName** Name of the outcome  
**type** The type of control: exposure or outcome

---

omopReferenceSet	<i>The OMOP reference set A reference set of 165 drug-outcome pairs where we believe the drug causes the outcome ( positive controls) and 234 drug-outcome pairs where we believe the drug does not cause the outcome (negative controls). The controls involve 4 health outcomes of interest: acute liver injury, acute kidney injury, acute myocardial infarction, and GI bleeding.</i>
------------------	---

---

## Description

The OMOP reference set A reference set of 165 drug-outcome pairs where we believe the drug causes the outcome ( positive controls) and 234 drug-outcome pairs where we believe the drug does not cause the outcome (negative controls). The controls involve 4 health outcomes of interest: acute liver injury, acute kidney injury, acute myocardial infarction, and GI bleeding.

## Usage

```
data(omopReferenceSet)
```

## Format

A data frame with 399 rows and 10 variables:

**exposureId** Concept ID identifying the exposure

**exposureName** Name of the exposure

**outcomeId** Concept ID identifying the outcome

**outcomeName** Name of the outcome

**groundTruth** 0 = negative control, 1 = positive control

**indicationId** Concept Id identifying the (primary) indication of the drug. To be used when one wants to nest the analysis within the indication

**indicationName** Name of the indication

**comparatorId** Concept ID identifying a comparator drug that can be used as a counterfactual

**comparatorName** Name of the comparator drug

**comparatorType** How the comparator was selected

## References

Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. Drug Safety 36 Suppl 1:S33-47, 2013

---

packageOhdsiBenchmarkResults

*Package results of a method on the OHDSI Methods Benchmark*


---

## Description

Stores the results of a method on the OHDSI Methods Benchmark in a standardized format, for example for use in the Method Evaluation Shiny app.

## Usage

```
packageOhdsiBenchmarkResults(estimates, controlSummary, analysisRef,
                              databaseName, exportFolder)
```

## Arguments

estimates	A data frame containing the estimates. See details for required columns.
controlSummary	A data frame with the summary of the controls as generated by the <a href="#">synthesizePositiveControls</a> function.
analysisRef	A file describing the various analyses that were performed. See details for required columns.
databaseName	A character string to identify the database the method was executed on.
exportFolder	The folder where the output CSV files will written.

## Details

The estimates argument should have the following columns: "targetId", "outcomeId", "analysisId", "logRr", "seLogRr", "ci95Lb", "ci95Ub". The analysisRef argument should have the following columns: "analysisId", "method", "comparative", "nesting", "firstExposureOnly". The targetId and outcomeId fields identify the specific control, and should correspond to those in the controlSummary object. The analysisId field is an integer that identifies a specific variant of the method. For example, if the method is 'CohortMethod', analysisId = 1 could identify a set of settings using propensity score matching, and analysisId = 2 could identify a set of settings using stratification. logRr, seLogRr, ci95Lb, and ci95Ub correspond to the log of the effect estimate (e.g. the hazard ratio), the standard error, and the upper and lower bound of the effect size estimate, as produced by the method. method is a character string identifying the method (e.g. "CohortMethod"). comparative is a boolean indicating whether the analysis can also be considered to perform comparative effect estimation (comparing the target to the comparator). nesting is a boolean indicating whether the analysis is nested in the nesting cohorts identified in the gold standard. firstExposureOnly is a boolean indicating whether only the first exposure was used in the analysis.

---

plotControls	<i>Plot negative and positive control estimates.</i>
--------------	--

---

**Description**

Plot negative and positive control estimates.

**Usage**

```
plotControls(logRr, seLogRr = NULL, ci95Lb = NULL, ci95Ub = NULL,
             trueLogRr, estimateType = "relative risk", fileName = NULL, title)
```

**Arguments**

logRr	A numeric vector of effect estimates on the log scale.
seLogRr	The standard error of the log of the effect estimates. Hint: often the standard error = (log(<lower bound 95 percent confidence interval>) - log(<effect estimate>))/qnorm(0.025). If not provided the standard error will be inferred from the 95 percent confidence interval.
ci95Lb	The lower bound of the 95 percent confidence interval. IF not provided it will be inferred from the standard error.
ci95Ub	The upper bound of the 95 percent confidence interval. IF not provided it will be inferred from the standard error.
trueLogRr	A vector of the true effect sizes
estimateType	A character string to denote the effect size estimate type. Used for the x-axis and the true effect size labels.
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggsave in the ggplot2 package for supported file formats.
title	An optional title to display above the plot.

**Value**

A Ggplot object. Use the ggsave function to save to file.

---

plotCoverageInjectedSignals	<i>Plot the coverage</i>
-----------------------------	--------------------------

---

**Description**

Plot the coverage

**Usage**

```
plotCoverageInjectedSignals(logRr, seLogRr, trueLogRr, region = 0.95,
                             fileName = NULL)
```

**Arguments**

logRr	A numeric vector of effect estimates on the log scale
seLogRr	The standard error of the log of the effect estimates. Hint: often the standard error = $(\log(\text{lower bound 95 percent confidence interval}) - \log(\text{effect estimate})) / \text{qnorm}(0.025)$
trueLogRr	A vector of the true effect sizes
region	Size of the confidence interval. Default is .95 (95 percent).
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggsave in the ggplot2 package for supported file formats.

**Details**

Plot the fractions of estimates where the true effect size is below, above or within the confidence interval, for one or more true effect sizes.

---

plotRocsInjectedSignals

*Plot the ROC curves for various injected signal sizes*

---

**Description**

Plot the ROC curves for various injected signal sizes

**Usage**

```
plotRocsInjectedSignals(logRr, trueLogRr, showAucs, fileName = NULL)
```

**Arguments**

logRr	A vector containing the log of the relative risk as estimated by a method.
trueLogRr	A vector containing the injected log(relative risk) for each estimate.
showAucs	Should the AUCs be shown in the plot?
fileName	Name of the file where the plot should be saved, for example 'plot.png'. See the function ggsave in the ggplot2 package for supported file formats.

**Value**

A Ggplot object. Use the ggsave function to save to file.



---

 synthesizePositiveControls

*Synthesize positive controls based on negative controls*


---

## Description

Synthesize positive controls based on negative controls

Synthesize positive controls

## Usage

```
synthesizePositiveControls(connectionDetails, cdmDatabaseSchema,
  oracleTempSchema = cdmDatabaseSchema,
  exposureDatabaseSchema = cdmDatabaseSchema,
  exposureTable = "drug_era",
  outcomeDatabaseSchema = cdmDatabaseSchema, outcomeTable = "cohort",
  outputDatabaseSchema = outcomeDatabaseSchema,
  outputTable = outcomeTable, createOutputTable = FALSE,
  exposureOutcomePairs, modelType = "poisson",
  minOutcomeCountForModel = 100, minOutcomeCountForInjection = 25,
  covariateSettings = FeatureExtraction::createCovariateSettings(useDemographicsAgeGroup
    = TRUE, useDemographicsGender = TRUE, useDemographicsIndexYear = TRUE,
    useDemographicsIndexMonth = TRUE, useConditionGroupEraLongTerm = TRUE,
    useDrugGroupEraLongTerm = TRUE, useProcedureOccurrenceLongTerm = TRUE,
    useMeasurementLongTerm = TRUE, useObservationLongTerm = TRUE,
    useCharlsonIndex = TRUE, useDcsi = TRUE, useChads2Vasc = TRUE,
    longTermStartDays = 365, endDays = 0),
  prior = Cyclops::createPrior("laplace", exclude = 0, useCrossValidation
    = TRUE), control = Cyclops::createControl(cvType = "auto",
  startingVariance = 0.1, noiseLevel = "quiet", threads = 10),
  firstExposureOnly = FALSE, washoutPeriod = 183,
  riskWindowStart = 0, riskWindowEnd = 0, endAnchor = "cohort end",
  addIntentToTreat = FALSE, firstOutcomeOnly = FALSE,
  removePeopleWithPriorOutcomes = FALSE, maxSubjectsForModel = 1e+05,
  effectSizes = c(1, 1.25, 1.5, 2, 4), precision = 0.01,
  outputIdOffset = 1000, workFolder = "./SignalInjectionTemp",
  cdmVersion = "5", modelThreads = 1, generationThreads = 1)
```

```
synthesizePositiveControls(connectionDetails, cdmDatabaseSchema,
  oracleTempSchema = cdmDatabaseSchema,
  exposureDatabaseSchema = cdmDatabaseSchema,
  exposureTable = "drug_era",
  outcomeDatabaseSchema = cdmDatabaseSchema, outcomeTable = "cohort",
  outputDatabaseSchema = outcomeDatabaseSchema,
  outputTable = outcomeTable, createOutputTable = FALSE,
  exposureOutcomePairs, modelType = "poisson",
  minOutcomeCountForModel = 100, minOutcomeCountForInjection = 25,
  covariateSettings = FeatureExtraction::createCovariateSettings(useDemographicsAgeGroup
    = TRUE, useDemographicsGender = TRUE, useDemographicsIndexYear = TRUE,
    useDemographicsIndexMonth = TRUE, useConditionGroupEraLongTerm = TRUE,
```

```

useDrugGroupEraLongTerm = TRUE, useProcedureOccurrenceLongTerm = TRUE,
useMeasurementLongTerm = TRUE, useObservationLongTerm = TRUE,
useCharlsonIndex = TRUE, useDcsi = TRUE, useChads2Vasc = TRUE,
longTermStartDays = 365, endDays = 0),
prior = Cyclops::createPrior("laplace", exclude = 0, useCrossValidation
= TRUE), control = Cyclops::createControl(cvType = "auto",
startingVariance = 0.1, noiseLevel = "quiet", threads = 10),
firstExposureOnly = FALSE, washoutPeriod = 183,
riskWindowStart = 0, riskWindowEnd = 0, endAnchor = "cohort end",
addIntentToTreat = FALSE, firstOutcomeOnly = FALSE,
removePeopleWithPriorOutcomes = FALSE, maxSubjectsForModel = 1e+05,
effectSizes = c(1, 1.25, 1.5, 2, 4), precision = 0.01,
outputIdOffset = 1000, workFolder = "./SignalInjectionTemp",
cdmVersion = "5", modelThreads = 1, generationThreads = 1)

```

## Arguments

### connectionDetails

An R object of type ConnectionDetails created using the function createConnectionDetails in the DatabaseConnector package.

### cdmDatabaseSchema

A database schema containing health care data in the OMOP Common Data Model. Note that for SQL Server, both the database and schema should be specified, e.g. 'cdm\_schema.dbo'

### oracleTempSchema

Should be used in Oracle to specify a schema where the user has write privileges for storing temporary tables.

### exposureDatabaseSchema

The name of the database schema that is the location where the exposure data used to define the exposure cohorts is available. If exposureTable = DRUG\_ERA, exposureDatabaseSchema is not used and assumed to be cdmDatabaseSchema. Requires read permissions to this database.

**exposureTable** The tablename that contains the exposure cohorts. If exposureTable <> DRUG\_ERA, then expectation is exposureTable has format of COHORT table: COHORT\_DEFINITION\_ID, SUBJECT\_ID, COHORT\_START\_DATE, COHORT\_END\_DATE.

### outcomeDatabaseSchema

The database schema where the target outcome table is located. Note that for SQL Server, both the database and schema should be specified, e.g. 'cdm\_schema.dbo'

**outcomeTable** The name of the table where the outcomes will be stored.

### outputDatabaseSchema

The name of the database schema that is the location of the tables containing the new outcomes Requires write permissions to this database.

**outputTable** The name of the table names that will contain the generated outcome cohorts.

### createOutputTable

Should the output table be created prior to inserting the outcomes? If TRUE and the tables already exists, it will first be deleted. If FALSE, the table is assumed to exist and the outcomes will be inserted. Any existing outcomes with the same IDs will first be deleted.

### exposureOutcomePairs

A data frame with at least two columns:

	<ul style="list-style-type: none"> <li>• "exposureId" containing the drug_concept_ID or cohort_concept_id of the exposure variable</li> <li>• "outcomeId" containing the condition_concept_ID or cohort_concept_id of the outcome variable</li> </ul>
modelType	Can be either "poisson" or "survival"
minOutcomeCountForModel	Minimum number of outcome events required to build a model.
minOutcomeCountForInjection	Minimum number of outcome events required to inject a signal.
covariateSettings	An object of type covariateSettings as created using the createCovariateSettings function in the FeatureExtraction package.
prior	The prior used to fit the outcome model. See <a href="#">createPrior</a> for details.
control	The control object used to control the cross-validation used to determine the hyperparameters of the prior (if applicable). See <a href="#">createControl</a> for details.
firstExposureOnly	Should signals be injected only for the first exposure? (ie. assuming an acute effect)
washoutPeriod	Number of days at the start of observation for which no signals will be injected, but will be used to determine whether exposure or outcome is the first one, and for extracting covariates to build the outcome model.
riskWindowStart	The start of the risk window relative to the start of the exposure (in days). When 0, risk is assumed to start on the first day of exposure.
riskWindowEnd	The end of the risk window (in days) relative to the endAnchor.
endAnchor	The anchor point for the end of the risk window. Can be "cohort start" or "cohort end".
addIntentToTreat	If true, the signal will not only be injected in the primary time at risk, but also after the time at risk (up until the observation period end). In both time periods, the target effect size will be enforced. This allows the same positive control synthesis to be used in both on treatment and intent-to-treat analysis variants. However, this will preclude the controls to be used in self-controlled designs that consider the time after exposure. Requires firstExposureOnly = TRUE.
firstOutcomeOnly	Should only the first outcome per person be considered when modeling the outcome?
removePeopleWithPriorOutcomes	Remove people with prior outcomes?
maxSubjectsForModel	Maximum number of people used to fit an outcome model.
effectSizes	A numeric vector of effect sizes that should be inserted.
precision	The allowed ratio between target and injected signal size.
outputIdOffset	What should be the first new outcome ID that is to be created?
workFolder	Name of local folder to place intermediary results; make sure to use forward slashes (/). Do not use a folder on a network drive since this greatly impacts performance.
cdmVersion	Define the OMOP CDM version used: currently support "4" and "5".

modelThreads	Number of parallel threads to use when fitting outcome models.
generationThreads	Number of parallel threads to use when generating outcomes.
referenceSet	The name of the reference set for which positive controls need to be synthesized. Currently supported are "ohdsiMethodsBenchmark".
maxCores	How many parallel cores should be used? If more cores are made available this can speed up the analyses.
summaryFileName	The name of the CSV file where to store the summary of the final set of positive and negative controls.
connectionDetails	An R object of type ConnectionDetails created using the function createConnectionDetails in the DatabaseConnector package.
cdmDatabaseSchema	Name of database schema that contains OMOP CDM and vocabulary.
oracleTempSchema	For Oracle only: the name of the database schema where you want all temporary tables to be managed. Requires create/insert permissions to this database.
exposureDatabaseSchema	The name of the database schema that is the location where the exposure data used to define the exposure cohorts is available. If exposureTable = DRUG_ERA, exposureDatabaseSchema is not used by assumed to be cdmSchema. Requires read permissions to this database.
exposureTable	The table name that contains the exposure cohorts. If exposureTable <> DRUG_ERA, then expectation is exposureTable has format of COHORT table: cohort_concept_id, SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE.
outcomeDatabaseSchema	The name of the database schema that is the location where the data used to define the outcome cohorts is available. If exposureTable = CONDITION_ERA, exposureDatabaseSchema is not used by assumed to be cdmSchema. Requires read permissions to this database.
outcomeTable	The table name that contains the outcome cohorts. When the table name is not CONDITION_ERA This table is expected to have the same format as the COHORT table: SUBJECT_ID, COHORT_START_DATE, COHORT_END_DATE, COHORT_CONCEPT_ID (CDM v4) or COHORT_DEFINITION_ID (CDM v5 and higher).
workFolder	Path to a folder where intermediate data will be stored.

## Details

This function will synthesize positive controls for a given reference set based on the real negative controls. Data from the database will be used to fit outcome models for each negative control outcome, and these models will be used to sample additional synthetic outcomes during exposure to increase the true hazard ratio. The positive control outcome cohorts will be stored in the same database table as the negative control outcome cohorts. A summary file will be created listing all positive and negative controls. This list should then be used as input for the method under evaluation.

This function will insert additional outcomes for a given set of drug-outcome pairs. It is assumed that these drug-outcome pairs represent negative controls, so the true relative risk before inserting any outcomes should be 1. There are two models for inserting the outcomes during the specified

risk window of the drug: a Poisson model assuming multiple outcomes could occur during a single exposure, and a survival model considering only one outcome per exposure. It is possible to use bulk import to insert the generated outcomes in the database. This requires the environmental variable 'USE\_MPP\_BULK\_LOAD' to be set to 'TRUE'. See `?DatabaseConnector::insertTable` for details on how to configure the bulk upload.

### **Value**

A data.frame listing all the drug-pairs in combination with requested effect sizes and the real inserted effect size (might be different from the requested effect size because of sampling error).

### **References**

Schuemie MJ, Hripcsak G, Ryan PB, Madigan D, Suchard MA. Empirical confidence interval calibration for population-level effect estimation studies in observational healthcare data. *Proc Natl Acad Sci U S A*. 2018 Mar 13;115(11):2571-2577.

# Index

## \*Topic **datasets**

- euadrReferenceSet, [7](#)
- ohdsiNegativeControls, [12](#)
- omopReferenceSet, [13](#)

- checkCohortSourceCodes, [2](#)
- computeMdr, [3](#)
- computeMetrics, [4](#)
- computeOhdsiBenchmarkMetrics, [5](#)
- createConnectionDetails, [2](#), [8](#)
- createControl, [10](#), [19](#)
- createPrior, [10](#), [19](#)
- createReferenceSetCohorts, [6](#), [12](#)

- euadrReferenceSet, [7](#)

- findOrphanSourceCodes, [8](#)

- injectSignals, [8](#), [12](#)

- launchMethodEvaluationApp, [11](#)

- MethodEvaluation, [12](#)

- MethodEvaluation-package
  - (MethodEvaluation), [12](#)

- ohdsiNegativeControls, [12](#)

- omopReferenceSet, [13](#)

- packageOhdsiBenchmarkResults, [5](#), [11](#), [14](#)

- plotControls, [15](#)

- plotCoverageInjectedSignals, [15](#)

- plotRocsInjectedSignals, [16](#)

- synthesizePositiveControls, [11](#)

- synthesizePositiveControls, [14](#), [17](#)