

# Package ‘SelfControlledCohort’

May 27, 2016

**Type** Package

**Title** Population-level estimation method that estimates  
incidence rate comparison of exposed/unexposed time within an exposed  
cohort

**Version** 1.1.1

**Date** 2016-4-15

**Author** Patrick Ryan, Martijn Schuemie

**Maintainer** Patrick Ryan <ryan@ohdsi.org>

**Description** This package provides a method to estimate risk by comparing time  
exposed with time unexposed among the exposed cohort.

**Depends** DatabaseConnector (>= 1.3.0)

**Imports** SqlRender (>= 1.1.1),  
RJDBC,  
OhdsiRTools (>= 1.1.1)

**Suggests** testthat

**License** Apache License 2.0

**RoxygenNote** 5.0.1

## R topics documented:

createExposureOutcome . . . . .	2
createRunSelfControlledCohortArgs . . . . .	2
createSccAnalysis . . . . .	4
loadExposureOutcomeList . . . . .	4
loadSccAnalysisList . . . . .	5
runSccAnalyses . . . . .	5
runSelfControlledCohort . . . . .	6
saveExposureOutcomeList . . . . .	9
saveSccAnalysisList . . . . .	9
SelfControlledCohort . . . . .	10
summarizeAnalyses . . . . .	10

<b>Index</b>	<b>11</b>
--------------	-----------

---

`createExposureOutcome` *Create exposure-outcome combinations.*

---

### Description

Create exposure-outcome combinations.

### Usage

```
createExposureOutcome(exposureId, outcomeId)
```

### Arguments

<code>exposureId</code>	A concept ID indentifying the drug of interest in the exposure table. If multiple strategies for picking the exposure will be tested in the analysis, a named list of numbers can be provided instead. In the analysis, the name of the number to be used can be specified using the <code>exposureType</code> parameter in the <a href="#">createSccAnalysis</a> function.
<code>outcomeId</code>	A concept ID indentifying the outcome of interest in the outcome table. If multiple strategies for picking the outcome will be tested in the analysis, a named list of numbers can be provided instead. In the analysis, the name of the number to be used can be specified using the <code>#'</code> <code>outcomeType</code> parameter in the <a href="#">createSccAnalysis</a> function.

### Details

Create a hypothesis of interest, to be used with the [runSccAnalyses](#) function.

---

`createRunSelfControlledCohortArgs`  
*Create a parameter object for the function `runSelfControlledCohort`*

---

### Description

Create a parameter object for the function `runSelfControlledCohort`

### Usage

```
createRunSelfControlledCohortArgs(firstOccurrenceDrugOnly = TRUE,
  firstOccurrenceConditionOnly = TRUE,
  outcomeConditionTypeConceptIds = c(38000247), genderConceptIds = c(8507,
  8532), minAge = "", maxAge = "", studyStartDate = "",
  studyEndDate = "", stratifyByGender = FALSE, stratifyByAge = FALSE,
  stratifyByYear = FALSE, useLengthOfExposureExposed = TRUE,
  timeAtRiskExposedStart = 1, surveillanceExposed = 30,
  useLengthOfExposureUnexposed = TRUE, timeAtRiskUnexposedStart = -1,
  surveillanceUnexposed = -30, hasFullTimeAtRisk = FALSE,
  washoutWindow = 0, followupWindow = 0, shrinkage = 1e-04)
```

**Arguments**

firstOccurrenceDrugOnly	If TRUE, only use first occurrence of each drug concept id for each person
firstOccurrenceConditionOnly	If TRUE, only use first occurrence of each condition conceptid for each person.
outcomeConditionTypeConceptIds	A list of TYPE_CONCEPT_ID values that will restrict condition occurrences. Only applicable if outcomeTable = CONDITION_OCCURRENCE.
genderConceptIds	of gender_concept_id, generally use MALE (8507) and FEMALE(8532).
minAge	Integer for minimum allowable age.
maxAge	Integer for maximum allowable age.
studyStartDate	Date for minimum allowable data for index exposure. Dateformat is 'yyyymmdd'.
studyEndDate	Date for maximum allowable data for index exposure. Dateformat is 'yyyymmdd'.
stratifyByGender	If TRUE, analysis will be calculated overall, and stratified across all gender groups.
stratifyByAge	If TRUE, analysis will be calculated overall, and stratified across all age groups (using AGE_GROUP table below).
stratifyByYear	If TRUE, analysis will be calculated overall, and stratified across all years of the index dates.
useLengthOfExposureExposed	If TRUE, use the duration from drugEraStart -> drugEraEnd as part of timeAtRisk.
timeAtRiskExposedStart	Integer of days to add to drugEraStart for start of timeAtRisk (0 to include index date, 1 to start the day after).
surveillanceExposed	Additional window to add to end of exposure period (if useLengthOfExposureExposed = TRUE, then add to exposure end date, else add to exposure start date).
useLengthOfExposureUnexposed	If TRUE, use the duration from exposure start -> exposure end as part of timeAtRisk looking back before exposure start.
timeAtRiskUnexposedStart	Integer of days to add to exposure start for start of timeAtRisk (0 to include index date, -1 to start the day before).
surveillanceUnexposed	Additional window to add to end of exposure period (if useLengthOfExposureUnexposed = TRUE, then add to exposure end date, else add to exposure start date).
hasFullTimeAtRisk	If TRUE, restrict to people who have full time-at-risk exposed and unexposed.
washoutWindow	Integer to define required time observed before exposure start.
followupWindow	Integer to define required time observed after exposure start.
shrinkage	Shrinkage used in IRR calculations, required >0 to deal with 0 case counts, but larger number means more shrinkage.

**Details**

Create an object defining the parameter values.

---

createSccAnalysis	<i>Create a SelfControlledCohort analysis specification</i>
-------------------	---

---

**Description**

Create a SelfControlledCohort analysis specification

**Usage**

```
createSccAnalysis(analysisId = 1, description = "", exposureType = NULL,
  outcomeType = NULL, runSelfControlledCohortArgs)
```

**Arguments**

analysisId	An integer that will be used later to refer to this specific set of analysis choices.
description	A short description of the analysis.
exposureType	If more than one exposure is provided for each exposureOutcome, this field should be used to select the specific exposure to use in this analysis.
outcomeType	If more than one outcome is provided for each exposureOutcome, this field should be used to select the specific outcome to use in this analysis.
runSelfControlledCohortArgs	An object representing the arguments to be used when calling the <a href="#">runSelfControlledCohort</a> function.

**Details**

Create a set of analysis choices, to be used with the [runSccAnalyses](#) function.

---

loadExposureOutcomeList	<i>Load a list of exposureOutcome from file</i>
-------------------------	---

---

**Description**

Load a list of objects of type exposureOutcome from file. The file is in JSON format.

**Usage**

```
loadExposureOutcomeList(file)
```

**Arguments**

file	The name of the file
------	----------------------

**Value**

A list of objects of type exposureOutcome.

---

loadSccAnalysisList	<i>Load a list of sccAnalysis from file</i>
---------------------	---

---

**Description**

Load a list of objects of type sccAnalysis from file. The file is in JSON format.

**Usage**

```
loadSccAnalysisList(file)
```

**Arguments**

file	The name of the file
------	----------------------

**Value**

A list of objects of type sccAnalysis.

---

runSccAnalyses	<i>Run a list of analyses</i>
----------------	-------------------------------

---

**Description**

Run a list of analyses

**Usage**

```
runSccAnalyses(connectionDetails, cdmDatabaseSchema,
  oracleTempSchema = cdmDatabaseSchema,
  exposureDatabaseSchema = cdmDatabaseSchema, exposureTable = "drug_era",
  outcomeDatabaseSchema = cdmDatabaseSchema,
  outcomeTable = "condition_occurrence", cdmVersion = 4,
  outputFolder = "./SelfControlledCohortOutput", sccAnalysisList,
  exposureOutcomeList, analysisThreads = 1)
```

**Arguments**

connectionDetails	An R object of type connectionDetails created using the function createConnectionDetails in the DatabaseConnector package.
cdmDatabaseSchema	The name of the database schema that contains the OMOP CDM instance. Requires read permissions to this database. On SQL Server, this should specify both the database and the schema, so for example 'cdm_instance.dbo'.
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 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".
outputFolder	Name of the folder where all the outputs will written to.
sccAnalysisList	A list of objects of type sccAnalysis as created using the <a href="#">createSccAnalysis</a> function.
exposureOutcomeList	A list of objects of type exposureOutcome as created using the <a href="#">createExposureOutcome</a> function.
analysisThreads	The number of parallel threads to use to execute the analyses.

## Details

Run a list of analyses for the drug-comparator-outcomes of interest. This function will run all specified analyses against all hypotheses of interest, meaning that the total number of outcome models is `length(cmAnalysisList) * length(drugComparatorOutcomesList)`.

---

runSelfControlledCohort

*Run self-controlled cohort*

---

## Description

runSelfControlledCohort generates population-level estimation by comparing exposed and un-exposed time among exposed cohort.

## Usage

```
runSelfControlledCohort(connectionDetails, cdmDatabaseSchema, cdmVersion = 5,
  oracleTempSchema, exposureIds, outcomeIds,
  exposureDatabaseSchema = cdmDatabaseSchema, exposureTable = "drug_era",
  outcomeDatabaseSchema = cdmDatabaseSchema, outcomeTable = "condition_era",
  firstOccurrenceDrugOnly = TRUE, firstOccurrenceConditionOnly = TRUE,
```

```

outcomeConditionTypeConceptIds = c(38000247), genderConceptIds = c(8507,
8532), minAge = "", maxAge = "", studyStartDate = "",
studyEndDate = "", stratifyByGender = FALSE, stratifyByAge = FALSE,
stratifyByYear = FALSE, useLengthOfExposureExposed = TRUE,
timeAtRiskExposedStart = 1, surveillanceExposed = 30,
useLengthOfExposureUnexposed = TRUE, timeAtRiskUnexposedStart = -1,
surveillanceUnexposed = -30, hasFullTimeAtRisk = FALSE,
washoutWindow = 0, followupWindow = 0, shrinkage = 1e-04)

```

## Arguments

### connectionDetails

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

### cdmDatabaseSchema

Name of database schema that contains the OMOP CDM and vocabulary.

### cdmVersion

Define the OMOP CDM version used: currently support "4" and "5".

### 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.

### exposureIds

A vector containing the drug\_concept\_ids or cohort\_definition\_ids of the exposures of interest

### outcomeIds

The condition\_concept\_ids or cohort\_definition\_ids of the outcomes of interest

### 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\_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 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.

### firstOccurrenceDrugOnly

If TRUE, only use first occurrence of each drug concept id for each person

### firstOccurrenceConditionOnly

If TRUE, only use first occurrence of each condition concept id for each person.

### outcomeConditionTypeConceptIds

A list of TYPE\_CONCEPT\_ID values that will restrict condition occurrences. Only applicable if outcomeTable = CONDITION\_OCCURRENCE.

### genderConceptIds

of gender\_concept\_id, generally use MALE (8507) and FEMALE (8532).

### minAge

Integer for minimum allowable age.

maxAge	Integer for maximum allowable age.
studyStartDate	Date for minimum allowable data for index exposure. Date format is 'yyyymmdd'.
studyEndDate	Date for maximum allowable data for index exposure. Date format is 'yyyymmdd'.
stratifyByGender	If TRUE, analysis will be calculated overall, and stratified across all gender groups.
stratifyByAge	If TRUE, analysis will be calculated overall, and stratified across all age groups (using AGE_GROUP table below).
stratifyByYear	If TRUE, analysis will be calculated overall, and stratified across all years of the index dates.
useLengthOfExposureExposed	If TRUE, use the duration from drugEraStart -> drugEraEnd as part of timeAtRisk.
timeAtRiskExposedStart	Integer of days to add to drugEraStart for start of timeAtRisk (0 to include index date, 1 to start the day after).
surveillanceExposed	Additional window to add to end of exposure period (if useLengthOfExposureExposed = TRUE, then add to exposure end date, else add to exposure start date).
useLengthOfExposureUnexposed	If TRUE, use the duration from exposure start -> exposure end as part of timeAtRisk looking back before exposure start.
timeAtRiskUnexposedStart	Integer of days to add to exposure start for start of timeAtRisk (0 to include index date, -1 to start the day before).
surveillanceUnexposed	Additional window to add to end of exposure period (if useLengthOfExposureUnexposed = TRUE, then add to exposure end date, else add to exposure start date).
hasFullTimeAtRisk	If TRUE, restrict to people who have full time-at-risk exposed and unexposed.
washoutWindow	Integer to define required time observed before exposure start.
followupWindow	Integer to define required time observed after exposure start.
shrinkage	Shrinkage used in IRR calculations, required >0 to deal with 0 case counts, but larger number means more shrinkage.

## Details

Population-level estimation method that estimates incidence rate comparison of exposed/unexposed time within an exposed cohort.

If multiple exposureIds and outcomeIds are provided, estimates will be generated for every combination of exposure and outcome.

## Value

An object of type `sccResults` containing the results of the analysis.



## References

Ryan PB, Schuemie MJ, Madigan D. Empirical performance of a self-controlled cohort method: lessons for developing a risk identification and analysis system. *Drug Safety* 36 Suppl1:S95-106, 2013

## Examples

```
## Not run:
connectionDetails <- createConnectionDetails(dbms = "sql server",
                                             server = "RNDUSRDHIT07.jnj.com")
sccResult <- runSelfControlledCohort(connectionDetails,
                                     cdmDatabaseSchema = "cdm_truven_mdcrcr.dbo",
                                     exposureIds = c(767410, 1314924, 907879),
                                     outcomeIds = 444382,
                                     outcomeTable = "condition_era")

## End(Not run)
```

---

```
saveExposureOutcomeList
```

*Save a list of exposureOutcome to file*

---

## Description

Write a list of objects of type exposureOutcome to file. The file is in JSON format.

## Usage

```
saveExposureOutcomeList(exposureOutcomeList, file)
```

## Arguments

exposureOutcomeList	The exposureOutcome list to be written to file
file	The name of the file where the results will be written

---

```
saveSccAnalysisList
```

*Save a list of sccAnalysis to file*

---

## Description

Write a list of objects of type sccAnalysis to file. The file is in JSON format.

## Usage

```
saveSccAnalysisList(sccAnalysisList, file)
```

## Arguments

sccAnalysisList	The sccAnalysis list to be written to file
file	The name of the file where the results will be written

---

SelfControlledCohort	<i>SelfControlledCohort</i>
----------------------	-----------------------------

---

**Description**  
SelfControlledCohort

---

summarizeAnalyses	<i>Create a summary report of the analyses</i>
-------------------	--

---

**Description**

Create a summary report of the analyses

**Usage**

summarizeAnalyses(resultsReference)

**Arguments**

resultsReference  
A data.frame as created by the [runSccAnalyses](#) function.

# Index

`createExposureOutcome`, [2](#), [6](#)  
`createRunSelfControlledCohortArgs`, [2](#)  
`createSccAnalysis`, [2](#), [4](#), [6](#)  
  
`loadExposureOutcomeList`, [4](#)  
`loadSccAnalysisList`, [5](#)  
  
`runSccAnalyses`, [2](#), [4](#), [5](#), [10](#)  
`runSelfControlledCohort`, [4](#), [6](#)  
  
`saveExposureOutcomeList`, [9](#)  
`saveSccAnalysisList`, [9](#)  
`SelfControlledCohort`, [10](#)  
`SelfControlledCohort-package`  
    (`SelfControlledCohort`), [10](#)  
`summarizeAnalyses`, [10](#)