

# Package ‘FeatureExtraction’

March 19, 2018

**Type** Package

**Title** Generating Features for a Cohort

**Version** 2.1.1

**Date** 2018-03-19

**Author** Martijn J. Schuemie [aut, cre],  
Marc A. Suchard [aut],  
Patrick B. Ryan [aut],  
Jenna Reys [aut]

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

**Description** An R package for generating features (covariates) for a cohort using data in the Common Data Model.

**License** Apache License 2.0

**VignetteBuilder** knitr

**Depends** R (>= 3.2.2),  
DatabaseConnector (>= 2.0.2),

**Imports** methods,  
bit,  
ff,  
ffbase (>= 0.12.1),  
plyr,  
Rcpp (>= 0.11.2),  
rJava,  
jsonlite,  
SqlRender (>= 1.4.5),

**Suggests** testthat,  
knitr,  
rmarkdown

**LinkingTo** Rcpp

**NeedsCompilation** yes

**RoxygenNote** 6.0.1

## R topics documented:

aggregateCovariates . . . . .	2
byMaxFf . . . . .	3

bySumFf . . . . .	3
computeStandardizedDifference . . . . .	4
convertPrespecSettingsToDetailedSettings . . . . .	4
createAnalysisDetails . . . . .	5
createCohortAttrCovariateSettings . . . . .	6
createCovariateSettings . . . . .	7
createDefaultCovariateSettings . . . . .	17
createDefaultTemporalCovariateSettings . . . . .	18
createDetailedCovariateSettings . . . . .	18
createDetailedTemporalCovariateSettings . . . . .	19
createHdpsCovariateSettings . . . . .	20
createTable1 . . . . .	23
createTable1CovariateSettings . . . . .	23
createTemporalCovariateSettings . . . . .	24
FeatureExtraction . . . . .	28
filterByRowId . . . . .	28
getDbCohortAttrCovariatesData . . . . .	29
getDbCovariateData . . . . .	30
getDbDefaultCovariateData . . . . .	31
getDbHdpsCovariateData . . . . .	33
getDefaultTable1Specifications . . . . .	34
loadCovariateData . . . . .	34
saveCovariateData . . . . .	35
tidyCovariateData . . . . .	36

<b>Index</b>	<b>37</b>
--------------	-----------

---

aggregateCovariates	<i>Aggregate covariate data</i>
---------------------	---------------------------------

---

## Description

Aggregate covariate data

## Usage

```
aggregateCovariates(covariateData)
```

## Arguments

**covariateData** An object of type covariateData as generated using getDbCovariateData.

## Value

An object of class covariateData.

---

`byMaxFf`*Compute max of values binned by a second variable*

---

**Description**

Compute max of values binned by a second variable

**Usage**

```
byMaxFf(values, bins)
```

**Arguments**

<code>values</code>	An ff object containing the numeric values to take the max of.
<code>bins</code>	An ff object containing the numeric values to bin by.

**Examples**

```
values <- ff::as.ff(c(1, 1, 2, 2, 1))
bins <- ff::as.ff(c(1, 1, 1, 2, 2))
byMaxFf(values, bins)
```

---

`bySumFf`*Compute sum of values binned by a second variable*

---

**Description**

Compute sum of values binned by a second variable

**Usage**

```
bySumFf(values, bins)
```

**Arguments**

<code>values</code>	An ff object containing the numeric values to take the sum of.
<code>bins</code>	An ff object containing the numeric values to bin by.

**Examples**

```
values <- ff::as.ff(c(1, 1, 2, 2, 1))
bins <- ff::as.ff(c(1, 1, 1, 2, 2))
bySumFf(values, bins)
```

---

`computeStandardizedDifference`*Compute standardized difference of mean for all covariates.*

---

**Description**

Computes the standardized difference for all covariates between two cohorts. The standardized difference is defined as the difference between the mean divided by the overall standard deviation.

**Usage**

```
computeStandardizedDifference(covariateData1, covariateData2)
```

**Arguments**

`covariateData1` The covariate data of the first cohort. Needs to be in aggregated format.

`covariateData2` The covariate data of the second cohort. Needs to be in aggregated format.

**Value**

A data frame with means and standard deviations per cohort as well as the standardized difference of mean.

---

`convertPrespecSettingsToDetailedSettings`*Convert prespecified covariate settings into detailed covariate settings*

---

**Description**

Convert prespecified covariate settings into detailed covariate settings

**Usage**

```
convertPrespecSettingsToDetailedSettings(covariateSettings)
```

**Arguments**

`covariateSettings`

An object of type `covariateSettings` as created for example by the [createCovariateSettings](#) function.

**Details**

For advanced users only.

**Value**

An object of type `covariateSettings`, to be used in other functions.

---

createAnalysisDetails *Create detailed covariate settings*

---

## Description

Create detailed covariate settings

## Usage

```
createAnalysisDetails(analysisId, sqlFileName, parameters,
  includedCovariateConceptIds = c(), addDescendantsToInclude = FALSE,
  excludedCovariateConceptIds = c(), addDescendantsToExclude = FALSE,
  includedCovariateIds = c())
```

## Arguments

analysisId	An integer between 0 and 999 that uniquely identifies this analysis.
sqlFileName	The name of the parameterized SQL file embedded in the featureExtraction package.
parameters	The list of parameter values used to render the template SQL.
includedCovariateConceptIds	A list of concept IDs that should be used to construct covariates.
addDescendantsToInclude	Should descendant concept IDs be added to the list of concepts to include?
excludedCovariateConceptIds	A list of concept IDs that should NOT be used to construct covariates.
addDescendantsToExclude	Should descendant concept IDs be added to the list of concepts to exclude?
includedCovariateIds	A list of covariate IDs that should be restricted to.

## Details

creates an object specifying in detail how covariates should be constructed from data in the CDM model. Warning: this function is for advanced users only.

## Value

An object of type analysisDetail, to be used in [createDetailedCovariateSettings](#) or [createDetailedTemporalCovariateSettings](#)

## Examples

```
analysisDetails <- createAnalysisDetails(analysisId = 1,
  sqlFileName = "DemographicsGender.sql",
  parameters = list(analysisId = 1,
    analysisName = "Gender",
    domainId = "Demographics"),
  includedCovariateConceptIds = c(),
  addDescendantsToInclude = FALSE,
  excludedCovariateConceptIds = c(),
  addDescendantsToExclude = FALSE,
```

```
includedCovariateIds = c())
```

---

```
createCohortAttrCovariateSettings
```

```
Create cohort attribute covariate settings
```

---

## Description

Create cohort attribute covariate settings

## Usage

```
createCohortAttrCovariateSettings(attrDatabaseSchema,
  attrDefinitionTable = "attribute_definition",
  cohortAttrTable = "cohort_attribute", includeAttrIds = c())
```

## Arguments

`attrDatabaseSchema`

The database schema where the attribute definition and cohort attribute table can be found.

`attrDefinitionTable`

The name of the attribute definition table.

`cohortAttrTable`

The name of the cohort attribute table.

`includeAttrIds` (optional) A list of attribute definition IDs to restrict to.

## Details

Creates an object specifying where the cohort attributes can be found to construct covariates. The attributes should be defined in a table with the same structure as the `attribute_definition` table in the Common Data Model. It should at least have these columns:

**attribute\_definition\_id** A unique identifier of type integer.

**attribute\_name** A short description of the attribute.

The cohort attributes themselves should be stored in a table with the same format as the `cohort_attribute` table in the Common Data Model. It should at least have these columns:

**cohort\_definition\_id** A key to link to the cohort table.

**subject\_id** A key to link to the cohort table.

**cohort\_start\_date** A key to link to the cohort table.

**attribute\_definition\_id** An foreign key linking to the attribute definition table.

**value\_as\_number** A real number.

## Value

An object of type `covariateSettings`, to be used in other functions.

---

createCovariateSettings

*Create covariate settings*


---

## Description

Create covariate settings

## Usage

```
createCovariateSettings(useDemographicsGender = FALSE,
  useDemographicsAge = FALSE, useDemographicsAgeGroup = FALSE,
  useDemographicsRace = FALSE, useDemographicsEthnicity = FALSE,
  useDemographicsIndexYear = FALSE, useDemographicsIndexMonth = FALSE,
  useDemographicsPriorObservationTime = FALSE,
  useDemographicsPostObservationTime = FALSE,
  useDemographicsTimeInCohort = FALSE,
  useDemographicsIndexYearMonth = FALSE,
  useConditionOccurrenceAnyTimePrior = FALSE,
  useConditionOccurrenceLongTerm = FALSE,
  useConditionOccurrenceMediumTerm = FALSE,
  useConditionOccurrenceShortTerm = FALSE,
  useConditionOccurrenceInpatientAnyTimePrior = FALSE,
  useConditionOccurrenceInpatientLongTerm = FALSE,
  useConditionOccurrenceInpatientMediumTerm = FALSE,
  useConditionOccurrenceInpatientShortTerm = FALSE,
  useConditionEraAnyTimePrior = FALSE, useConditionEraLongTerm = FALSE,
  useConditionEraMediumTerm = FALSE, useConditionEraShortTerm = FALSE,
  useConditionEraOverlapping = FALSE, useConditionEraStartLongTerm = FALSE,
  useConditionEraStartMediumTerm = FALSE,
  useConditionEraStartShortTerm = FALSE,
  useConditionGroupEraAnyTimePrior = FALSE,
  useConditionGroupEraLongTerm = FALSE,
  useConditionGroupEraMediumTerm = FALSE,
  useConditionGroupEraShortTerm = FALSE,
  useConditionGroupEraOverlapping = FALSE,
  useConditionGroupEraStartLongTerm = FALSE,
  useConditionGroupEraStartMediumTerm = FALSE,
  useConditionGroupEraStartShortTerm = FALSE,
  useDrugExposureAnyTimePrior = FALSE, useDrugExposureLongTerm = FALSE,
  useDrugExposureMediumTerm = FALSE, useDrugExposureShortTerm = FALSE,
  useDrugEraAnyTimePrior = FALSE, useDrugEraLongTerm = FALSE,
  useDrugEraMediumTerm = FALSE, useDrugEraShortTerm = FALSE,
  useDrugEraOverlapping = FALSE, useDrugEraStartLongTerm = FALSE,
  useDrugEraStartMediumTerm = FALSE, useDrugEraStartShortTerm = FALSE,
  useDrugGroupEraAnyTimePrior = FALSE, useDrugGroupEraLongTerm = FALSE,
  useDrugGroupEraMediumTerm = FALSE, useDrugGroupEraShortTerm = FALSE,
  useDrugGroupEraOverlapping = FALSE, useDrugGroupEraStartLongTerm = FALSE,
  useDrugGroupEraStartMediumTerm = FALSE,
  useDrugGroupEraStartShortTerm = FALSE,
  useProcedureOccurrenceAnyTimePrior = FALSE,
```

```

useProcedureOccurrenceLongTerm = FALSE,
useProcedureOccurrenceMediumTerm = FALSE,
useProcedureOccurrenceShortTerm = FALSE,
useDeviceExposureAnyTimePrior = FALSE, useDeviceExposureLongTerm = FALSE,
useDeviceExposureMediumTerm = FALSE, useDeviceExposureShortTerm = FALSE,
useMeasurementAnyTimePrior = FALSE, useMeasurementLongTerm = FALSE,
useMeasurementMediumTerm = FALSE, useMeasurementShortTerm = FALSE,
useMeasurementValueAnyTimePrior = FALSE,
useMeasurementValueLongTerm = FALSE,
useMeasurementValueMediumTerm = FALSE,
useMeasurementValueShortTerm = FALSE,
useMeasurementRangeGroupAnyTimePrior = FALSE,
useMeasurementRangeGroupLongTerm = FALSE,
useMeasurementRangeGroupMediumTerm = FALSE,
useMeasurementRangeGroupShortTerm = FALSE,
useObservationAnyTimePrior = FALSE, useObservationLongTerm = FALSE,
useObservationMediumTerm = FALSE, useObservationShortTerm = FALSE,
useCharlsonIndex = FALSE, useDcsi = FALSE, useChads2 = FALSE,
useChads2Vasc = FALSE, useDistinctConditionCountLongTerm = FALSE,
useDistinctConditionCountMediumTerm = FALSE,
useDistinctConditionCountShortTerm = FALSE,
useDistinctIngredientCountLongTerm = FALSE,
useDistinctIngredientCountMediumTerm = FALSE,
useDistinctIngredientCountShortTerm = FALSE,
useDistinctProcedureCountLongTerm = FALSE,
useDistinctProcedureCountMediumTerm = FALSE,
useDistinctProcedureCountShortTerm = FALSE,
useDistinctMeasurementCountLongTerm = FALSE,
useDistinctMeasurementCountMediumTerm = FALSE,
useDistinctMeasurementCountShortTerm = FALSE,
useDistinctObservationCountLongTerm = FALSE,
useDistinctObservationCountMediumTerm = FALSE,
useDistinctObservationCountShortTerm = FALSE,
useVisitCountLongTerm = FALSE, useVisitCountMediumTerm = FALSE,
useVisitCountShortTerm = FALSE, useVisitConceptCountLongTerm = FALSE,
useVisitConceptCountMediumTerm = FALSE,
useVisitConceptCountShortTerm = FALSE, longTermStartDays = -365,
mediumTermStartDays = -180, shortTermStartDays = -30, endDays = 0,
includedCovariateConceptIds = c(), addDescendantsToInclude = FALSE,
excludedCovariateConceptIds = c(), addDescendantsToExclude = FALSE,
includedCovariateIds = c())

```

## Arguments

useDemographicsGender

Gender of the subject. (analysis ID 1)

useDemographicsAge

Age of the subject on the index date (in years). (analysis ID 2)

useDemographicsAgeGroup

Age of the subject on the index date (in 5 year age groups) (analysis ID 3)

useDemographicsRace

Race of the subject. (analysis ID 4)



useDemographicsEthnicity  
Ethnicity of the subject. (analysis ID 5)

useDemographicsIndexYear  
Year of the index date. (analysis ID 6)

useDemographicsIndexMonth  
Month of the index date. (analysis ID 7)

useDemographicsPriorObservationTime  
Number of continuous days of observation time preceding the index date. (analysis ID 8)

useDemographicsPostObservationTime  
Number of continuous days of observation time preceding the index date. (analysis ID 9)

useDemographicsTimeInCohort  
Number of days of observation time during cohort period. (analysis ID 10)

useDemographicsIndexYearMonth  
Both calendar year and month of the index date in a single variable. (analysis ID 11)

useConditionOccurrenceAnyTimePrior  
One covariate per condition in the condition\_occurrence table starting any time prior to index. (analysis ID 101)

useConditionOccurrenceLongTerm  
One covariate per condition in the condition\_occurrence table starting in the long term window. (analysis ID 102)

useConditionOccurrenceMediumTerm  
One covariate per condition in the condition\_occurrence table starting in the medium term window. (analysis ID 103)

useConditionOccurrenceShortTerm  
One covariate per condition in the condition\_occurrence table starting in the short term window. (analysis ID 104)

useConditionOccurrenceInpatientAnyTimePrior  
One covariate per condition observed in an inpatient setting in the condition\_occurrence table starting any time prior to index. (analysis ID 105)

useConditionOccurrenceInpatientLongTerm  
One covariate per condition observed in an inpatient setting in the condition\_occurrence table starting in the long term window. (analysis ID 106)

useConditionOccurrenceInpatientMediumTerm  
One covariate per condition observed in an inpatient setting in the condition\_occurrence table starting in the medium term window. (analysis ID 107)

useConditionOccurrenceInpatientShortTerm  
One covariate per condition observed in an inpatient setting in the condition\_occurrence table starting in the short term window. (analysis ID 108)

useConditionEraAnyTimePrior  
One covariate per condition in the condition\_era table overlapping with any time prior to index. (analysis ID 201)

useConditionEraLongTerm  
One covariate per condition in the condition\_era table overlapping with any part of the long term window. (analysis ID 202)

useConditionEraMediumTerm  
One covariate per condition in the condition\_era table overlapping with any part of the medium term window. (analysis ID 203)

useConditionEraShortTerm  
One covariate per condition in the condition\_era table overlapping with any part of the short term window. (analysis ID 204)

useConditionEraOverlapping  
One covariate per condition in the condition\_era table overlapping with the end of the risk window. (analysis ID 205)

useConditionEraStartLongTerm  
One covariate per condition in the condition\_era table starting in the long term window. (analysis ID 206)

useConditionEraStartMediumTerm  
One covariate per condition in the condition\_era table starting in the medium term window. (analysis ID 207)

useConditionEraStartShortTerm  
One covariate per condition in the condition\_era table starting in the short term window. (analysis ID 208)

useConditionGroupEraAnyTimePrior  
One covariate per condition era rolled up to groups in the condition\_era table overlapping with any time prior to index. (analysis ID 209)

useConditionGroupEraLongTerm  
One covariate per condition era rolled up to groups in the condition\_era table overlapping with any part of the long term window. (analysis ID 210)

useConditionGroupEraMediumTerm  
One covariate per condition era rolled up to groups in the condition\_era table overlapping with any part of the medium term window. (analysis ID 211)

useConditionGroupEraShortTerm  
One covariate per condition era rolled up to groups in the condition\_era table overlapping with any part of the short term window. (analysis ID 212)

useConditionGroupEraOverlapping  
One covariate per condition era rolled up to groups in the condition\_era table overlapping with the end of the risk window. (analysis ID 213)

useConditionGroupEraStartLongTerm  
One covariate per condition era rolled up to groups in the condition\_era table starting in the long term window. (analysis ID 214)

useConditionGroupEraStartMediumTerm  
One covariate per condition era rolled up to groups in the condition\_era table starting in the medium term window. (analysis ID 215)

useConditionGroupEraStartShortTerm  
One covariate per condition era rolled up to groups in the condition\_era table starting in the short term window. (analysis ID 216)

useDrugExposureAnyTimePrior  
One covariate per drug in the drug\_exposure table starting any time prior to index. (analysis ID 301)

useDrugExposureLongTerm  
One covariate per drug in the drug\_exposure table starting in the long term window. (analysis ID 302)

useDrugExposureMediumTerm  
One covariate per drug in the drug\_exposure table starting in the medium term window. (analysis ID 303)

useDrugExposureShortTerm  
One covariate per drug in the drug\_exposure table starting in the short term window. (analysis ID 304)

useDrugEraAnyTimePrior	One covariate per drug in the drug_era table overlapping with any time prior to index. (analysis ID 401)
useDrugEraLongTerm	One covariate per drug in the drug_era table overlapping with any part of the long term window. (analysis ID 402)
useDrugEraMediumTerm	One covariate per drug in the drug_era table overlapping with any part of the medium term window. (analysis ID 403)
useDrugEraShortTerm	One covariate per drug in the drug_era table overlapping with any part of the short window. (analysis ID 404)
useDrugEraOverlapping	One covariate per drug in the drug_era table overlapping with the end of the risk window. (analysis ID 405)
useDrugEraStartLongTerm	One covariate per drug in the drug_era table starting in the long term window. (analysis ID 406)
useDrugEraStartMediumTerm	One covariate per drug in the drug_era table starting in the medium term window. (analysis ID 407)
useDrugEraStartShortTerm	One covariate per drug in the drug_era table starting in the long short window. (analysis ID 408)
useDrugGroupEraAnyTimePrior	One covariate per drug rolled up to ATC groups in the drug_era table overlapping with any time prior to index. (analysis ID 409)
useDrugGroupEraLongTerm	One covariate per drug rolled up to ATC groups in the drug_era table overlapping with any part of the long term window. (analysis ID 410)
useDrugGroupEraMediumTerm	One covariate per drug rolled up to ATC groups in the drug_era table overlapping with any part of the medium term window. (analysis ID 411)
useDrugGroupEraShortTerm	One covariate per drug rolled up to ATC groups in the drug_era table overlapping with any part of the short term window. (analysis ID 412)
useDrugGroupEraOverlapping	One covariate per drug rolled up to ATC groups in the drug_era table overlapping with the end of the risk window. (analysis ID 413)
useDrugGroupEraStartLongTerm	One covariate per drug rolled up to ATC groups in the drug_era table starting in the long term window. (analysis ID 414)
useDrugGroupEraStartMediumTerm	One covariate per drug rolled up to ATC groups in the drug_era table starting in the medium term window. (analysis ID 415)
useDrugGroupEraStartShortTerm	One covariate per drug rolled up to ATC groups in the drug_era table starting in the short term window. (analysis ID 416)
useProcedureOccurrenceAnyTimePrior	One covariate per procedure in the procedure_occurrence table any time prior to index. (analysis ID 501)

useProcedureOccurrenceLongTerm  
     One covariate per procedure in the procedure\_occurrence table in the long term window. (analysis ID 502)

useProcedureOccurrenceMediumTerm  
     One covariate per procedure in the procedure\_occurrence table in the medium term window. (analysis ID 503)

useProcedureOccurrenceShortTerm  
     One covariate per procedure in the procedure\_occurrence table in the short term window. (analysis ID 504)

useDeviceExposureAnyTimePrior  
     One covariate per device in the device exposure table starting any time prior to index. (analysis ID 601)

useDeviceExposureLongTerm  
     One covariate per device in the device exposure table starting in the long term window. (analysis ID 602)

useDeviceExposureMediumTerm  
     One covariate per device in the device exposure table starting in the medium term window. (analysis ID 603)

useDeviceExposureShortTerm  
     One covariate per device in the device exposure table starting in the short term window. (analysis ID 604)

useMeasurementAnyTimePrior  
     One covariate per measurement in the measurement table any time prior to index. (analysis ID 701)

useMeasurementLongTerm  
     One covariate per measurement in the measurement table in the long term window. (analysis ID 702)

useMeasurementMediumTerm  
     One covariate per measurement in the measurement table in the medium term window. (analysis ID 703)

useMeasurementShortTerm  
     One covariate per measurement in the measurement table in the short term window. (analysis ID 704)

useMeasurementValueAnyTimePrior  
     One covariate containing the value per measurement-unit combination any time prior to index. (analysis ID 705)

useMeasurementValueLongTerm  
     One covariate containing the value per measurement-unit combination in the long term window. (analysis ID 706)

useMeasurementValueMediumTerm  
     One covariate containing the value per measurement-unit combination in the medium term window. (analysis ID 707)

useMeasurementValueShortTerm  
     One covariate containing the value per measurement-unit combination in the short term window. (analysis ID 708)

useMeasurementRangeGroupAnyTimePrior  
     Covariates indicating whether measurements are below, within, or above normal range any time prior to index. (analysis ID 709)

useMeasurementRangeGroupLongTerm  
     Covariates indicating whether measurements are below, within, or above normal range in the long term window. (analysis ID 710)

useMeasurementRangeGroupMediumTerm	Covariates indicating whether measurements are below, within, or above normal range in the medium term window. (analysis ID 711)
useMeasurementRangeGroupShortTerm	Covariates indicating whether measurements are below, within, or above normal range in the short term window. (analysis ID 712)
useObservationAnyTimePrior	One covariate per observation in the observation table any time prior to index. (analysis ID 801)
useObservationLongTerm	One covariate per observation in the observation table in the long term window. (analysis ID 802)
useObservationMediumTerm	One covariate per observation in the observation table in the medium term window. (analysis ID 803)
useObservationShortTerm	One covariate per observation in the observation table in the short term window. (analysis ID 804)
useCharlsonIndex	The Charlson comorbidity index (Romano adaptation) using all conditions prior to the window end. (analysis ID 901)
useDcsi	The Diabetes Comorbidity Severity Index (DCSI) using all conditions prior to the window end. (analysis ID 902)
useChads2	The CHADS2 score using all conditions prior to the window end. (analysis ID 903)
useChads2Vasc	The CHADS2VAsc score using all conditions prior to the window end. (analysis ID 904)
useDistinctConditionCountLongTerm	The number of distinct condition concepts observed in the long term window. (analysis ID 905)
useDistinctConditionCountMediumTerm	The number of distinct condition concepts observed in the medium term window. (analysis ID 906)
useDistinctConditionCountShortTerm	The number of distinct condition concepts observed in the short term window. (analysis ID 907)
useDistinctIngredientCountLongTerm	The number of distinct ingredients observed in the long term window. (analysis ID 908)
useDistinctIngredientCountMediumTerm	The number of distinct ingredients observed in the medium term window. (analysis ID 909)
useDistinctIngredientCountShortTerm	The number of distinct ingredients observed in the short term window. (analysis ID 910)
useDistinctProcedureCountLongTerm	The number of distinct procedures observed in the long term window. (analysis ID 911)

useDistinctProcedureCountMediumTerm  
 The number of distinct procedures observed in the medium term window. (analysis ID 912)

useDistinctProcedureCountShortTerm  
 The number of distinct procedures observed in the short term window. (analysis ID 913)

useDistinctMeasurementCountLongTerm  
 The number of distinct measurements observed in the long term window. (analysis ID 914)

useDistinctMeasurementCountMediumTerm  
 The number of distinct measurements observed in the medium term window. (analysis ID 915)

useDistinctMeasurementCountShortTerm  
 The number of distinct measurements observed in the short term window. (analysis ID 916)

useDistinctObservationCountLongTerm  
 The number of distinct observations observed in the long term window. (analysis ID 917)

useDistinctObservationCountMediumTerm  
 The number of distinct observations observed in the medium term window. (analysis ID 918)

useDistinctObservationCountShortTerm  
 The number of distinct observations observed in the short term window. (analysis ID 919)

useVisitCountLongTerm  
 The number of visits observed in the long term window. (analysis ID 920)

useVisitCountMediumTerm  
 The number of visits observed in the medium term window. (analysis ID 921)

useVisitCountShortTerm  
 The number of visits observed in the short term window. (analysis ID 922)

useVisitConceptCountLongTerm  
 The number of visits observed in the long term window, stratified by visit concept ID. (analysis ID 923)

useVisitConceptCountMediumTerm  
 The number of visits observed in the medium term window, stratified by visit concept ID. (analysis ID 924)

useVisitConceptCountShortTerm  
 The number of visits observed in the short term window, stratified by visit concept ID. (analysis ID 925)

longTermStartDays  
 What is the start day (relative to the index date) of the long-term window?

mediumTermStartDays  
 What is the start day (relative to the index date) of the medium-term window?

shortTermStartDays  
 What is the start day (relative to the index date) of the short-term window?

endDays  
 What is the end day (relative to the index date) of the window?

includedCovariateConceptIds  
 A list of concept IDs that should be used to construct covariates.

addDescendantsToInclude	Should descendant concept IDs be added to the list of concepts to include?
excludedCovariateConceptIds	A list of concept IDs that should NOT be used to construct covariates.
addDescendantsToExclude	Should descendant concept IDs be added to the list of concepts to exclude?
includedCovariateIds	A list of covariate IDs that should be restricted to.

```

useDrugExposureMediumTerm = FALSE,
useDrugExposureShortTerm = FALSE,
useDrugEraAnyTimePrior = FALSE,
useDrugEraLongTerm = FALSE,
useDrugEraMediumTerm = FALSE,
useDrugEraShortTerm = FALSE,
useDrugEraOverlapping = FALSE,
useDrugEraStartLongTerm = FALSE,
useDrugEraStartMediumTerm = FALSE,
useDrugEraStartShortTerm = FALSE,
useDrugGroupEraAnyTimePrior = FALSE,
useDrugGroupEraLongTerm = TRUE,
useDrugGroupEraMediumTerm = FALSE,
useDrugGroupEraShortTerm = TRUE,
useDrugGroupEraOverlapping = TRUE,
useDrugGroupEraStartLongTerm = FALSE,
useDrugGroupEraStartMediumTerm = FALSE,
useDrugGroupEraStartShortTerm = FALSE,
useProcedureOccurrenceAnyTimePrior = FALSE,
useProcedureOccurrenceLongTerm = TRUE,
useProcedureOccurrenceMediumTerm = FALSE,
useProcedureOccurrenceShortTerm = TRUE,
useDeviceExposureAnyTimePrior = FALSE,
useDeviceExposureLongTerm = TRUE,
useDeviceExposureMediumTerm = FALSE,
useDeviceExposureShortTerm = TRUE,
useMeasurementAnyTimePrior = FALSE,
useMeasurementLongTerm = TRUE,
useMeasurementMediumTerm = FALSE,
useMeasurementShortTerm = TRUE,
useMeasurementValueAnyTimePrior = FALSE,
useMeasurementValueLongTerm = FALSE,
useMeasurementValueMediumTerm = FALSE,
useMeasurementValueShortTerm = FALSE,
useMeasurementRangeGroupAnyTimePrior = FALSE,
useMeasurementRangeGroupLongTerm = TRUE,
useMeasurementRangeGroupMediumTerm = FALSE,
useMeasurementRangeGroupShortTerm = FALSE,
useObservationAnyTimePrior = FALSE,
useObservationLongTerm = TRUE,
useObservationMediumTerm = FALSE,
useObservationShortTerm = TRUE,
useCharlsonIndex = TRUE,
useDcsi = TRUE,
useChads2 = TRUE,
useChads2Vasc = TRUE,
useDistinctConditionCountLongTerm = FALSE,
useDistinctConditionCountMediumTerm = FALSE,
useDistinctConditionCountShortTerm = FALSE,
useDistinctIngredientCountLongTerm = FALSE,
useDistinctIngredientCountMediumTerm = FALSE,
useDistinctIngredientCountShortTerm = FALSE,
useDistinctProcedureCountLongTerm = FALSE,
useDistinctProcedureCountMediumTerm = FALSE,
useDistinctProcedureCountShortTerm = FALSE,
useDistinctMeasurementCountLongTerm = FALSE,
useDistinctMeasurementCountMediumTerm = FALSE,

```



```

useDistinctMeasurementCountShortTerm = FALSE,
useDistinctObservationCountLongTerm = FALSE,
useDistinctObservationCountMediumTerm = FALSE,
useDistinctObservationCountShortTerm = FALSE,
useVisitCountLongTerm = FALSE,
useVisitCountMediumTerm = FALSE,
useVisitCountShortTerm = FALSE,
useVisitConceptCountLongTerm = FALSE,
useVisitConceptCountMediumTerm = FALSE,
useVisitConceptCountShortTerm = FALSE,
longTermStartDays = -365,
mediumTermStartDays = -180,
shortTermStartDays = -30,
endDays = 0,
includedCovariateConceptIds = c(),
addDescendantsToInclude = FALSE,
excludedCovariateConceptIds = c(),
addDescendantsToExclude = FALSE,
includedCovariateIds = c())

```

---

createDefaultCovariateSettings

*Create default covariate settings*

---

## Description

Create default covariate settings

## Usage

```

createDefaultCovariateSettings(includedCovariateConceptIds = c(),
  addDescendantsToInclude = FALSE, excludedCovariateConceptIds = c(),
  addDescendantsToExclude = FALSE, includedCovariateIds = c())

```

## Arguments

includedCovariateConceptIds

A list of concept IDs that should be used to construct covariates.

addDescendantsToInclude

Should descendant concept IDs be added to the list of concepts to include?

excludedCovariateConceptIds

A list of concept IDs that should NOT be used to construct covariates.

addDescendantsToExclude

Should descendant concept IDs be added to the list of concepts to exclude?

includedCovariateIds

A list of covariate IDs that should be restricted to.

## Value

An object of type covariateSettings, to be used in other functions.

---

```
createDefaultTemporalCovariateSettings
```

*Create default covariate settings*

---

### Description

Create default covariate settings

### Usage

```
createDefaultTemporalCovariateSettings(includedCovariateConceptIds = c(),
  addDescendantsToInclude = FALSE, excludedCovariateConceptIds = c(),
  addDescendantsToExclude = FALSE, includedCovariateIds = c())
```

### Arguments

`includedCovariateConceptIds`  
A list of concept IDs that should be used to construct covariates.

`addDescendantsToInclude`  
Should descendant concept IDs be added to the list of concepts to include?

`excludedCovariateConceptIds`  
A list of concept IDs that should NOT be used to construct covariates.

`addDescendantsToExclude`  
Should descendant concept IDs be added to the list of concepts to exclude?

`includedCovariateIds`  
A list of covariate IDs that should be restricted to.

### Value

An object of type `covariateSettings`, to be used in other functions.

---

```
createDetailedCovariateSettings
```

*Create detailed covariate settings*

---

### Description

Create detailed covariate settings

### Usage

```
createDetailedCovariateSettings(analyses = list())
```

### Arguments

`analyses` A list of `analysisDetail` objects as created using [createAnalysisDetails](#).

**Details**

creates an object specifying in detail how covariates should be constructed from data in the CDM model. Warning: this function is for advanced users only.

**Value**

An object of type covariateSettings, to be used in other functions.

---

```
createDetailedTemporalCovariateSettings
```

*Create detailed temporal covariate settings*

---

**Description**

Create detailed temporal covariate settings

**Usage**

```
createDetailedTemporalCovariateSettings(analyses = list(),  
    temporalStartDays = -365:-1, temporalEndDays = -365:-1)
```

**Arguments**

analyses	A list of analysis detail objects as created using <a href="#">createAnalysisDetails</a> .
temporalStartDays	A list of integers representing the start of a time period, relative to the index date. 0 indicates the index date, -1 indicates the day before the index date, etc. The start day is included in the time period.
temporalEndDays	A list of integers representing the end of a time period, relative to the index date. 0 indicates the index date, -1 indicates the day before the index date, etc. The end day is included in the time period.

**Details**

creates an object specifying in detail how temporal covariates should be constructed from data in the CDM model. Warning: this function is for advanced users only.

**Value**

An object of type covariateSettings, to be used in other functions.

---

```
createHdpsCovariateSettings
```

*Create HDPS covariate settings*

---

## Description

Create HDPS covariate settings

## Usage

```
createHdpsCovariateSettings(useCovariateCohortIdIs1 = FALSE,
  useCovariateDemographics = TRUE, useCovariateDemographicsGender = TRUE,
  useCovariateDemographicsRace = TRUE,
  useCovariateDemographicsEthnicity = TRUE,
  useCovariateDemographicsAge = TRUE, useCovariateDemographicsYear = TRUE,
  useCovariateDemographicsMonth = TRUE,
  useCovariateConditionOccurrence = TRUE,
  useCovariate3DigitIcd9Inpatient180d = FALSE,
  useCovariate3DigitIcd9Inpatient180dMedF = FALSE,
  useCovariate3DigitIcd9Inpatient180d75F = FALSE,
  useCovariate3DigitIcd9Ambulatory180d = FALSE,
  useCovariate3DigitIcd9Ambulatory180dMedF = FALSE,
  useCovariate3DigitIcd9Ambulatory180d75F = FALSE,
  useCovariateDrugExposure = FALSE,
  useCovariateIngredientExposure180d = FALSE,
  useCovariateIngredientExposure180dMedF = FALSE,
  useCovariateIngredientExposure180d75F = FALSE,
  useCovariateProcedureOccurrence = FALSE,
  useCovariateProcedureOccurrenceInpatient180d = FALSE,
  useCovariateProcedureOccurrenceInpatient180dMedF = FALSE,
  useCovariateProcedureOccurrenceInpatient180d75F = FALSE,
  useCovariateProcedureOccurrenceAmbulatory180d = FALSE,
  useCovariateProcedureOccurrenceAmbulatory180dMedF = FALSE,
  useCovariateProcedureOccurrenceAmbulatory180d75F = FALSE,
  excludedCovariateConceptIds = c(), includedCovariateConceptIds = c(),
  deleteCovariatesSmallCount = 100)
```

## Arguments

`useCovariateCohortIdIs1`

A boolean value (TRUE/FALSE) to determine if a covariate should be constructed for whether the cohort ID is 1 (currently primarily used in Cohort-Method).

`useCovariateDemographics`

A boolean value (TRUE/FALSE) to determine if demographic covariates (age in 5-yr increments, gender, race, ethnicity, year of index date, month of index date) will be created and included in future models.

`useCovariateDemographicsGender`

A boolean value (TRUE/FALSE) to determine if gender should be included in the model.

useCovariateDemographicsRace	A boolean value (TRUE/FALSE) to determine if race should be included in the model.
useCovariateDemographicsEthnicity	A boolean value (TRUE/FALSE) to determine if ethnicity should be included in the model.
useCovariateDemographicsAge	A boolean value (TRUE/FALSE) to determine if age (in 5 year increments) should be included in the model.
useCovariateDemographicsYear	A boolean value (TRUE/FALSE) to determine if calendar year should be included in the model.
useCovariateDemographicsMonth	A boolean value (TRUE/FALSE) to determine if calendar month should be included in the model.
useCovariateConditionOccurrence	A boolean value (TRUE/FALSE) to determine if covariates derived from CONDITION_OCCURRENCE table will be created and included in future models.
useCovariate3DigitIcd9Inpatient180d	A boolean value (TRUE/FALSE) to determine if covariates will be created and used in models that look for presence/absence of condition within inpatient setting in 180d window prior to or on cohort index date. Conditions are aggregated at the ICD-9 3-digit level. Only applicable if useCovariateConditionOccurrence = TRUE.
useCovariate3DigitIcd9Inpatient180dMedF	Similar to useCovariate3DigitIcd9Inpatient180d, but now only if the frequency of the ICD-9 code is higher than the median.
useCovariate3DigitIcd9Inpatient180d75F	Similar to useCovariate3DigitIcd9Inpatient180d, but now only if the frequency of the ICD-9 code is higher than the 75th percentile.
useCovariate3DigitIcd9Ambulatory180d	A boolean value (TRUE/FALSE) to determine if covariates will be created and used in models that look for presence/absence of condition within ambulatory setting in 180d window prior to or on cohort index date. Conditions are aggregated at the ICD-9 3-digit level. Only applicable if useCovariateConditionOccurrence = TRUE.
useCovariate3DigitIcd9Ambulatory180dMedF	Similar to useCovariate3DigitIcd9Ambulatory180d, but now only if the frequency of the ICD-9 code is higher than the median.
useCovariate3DigitIcd9Ambulatory180d75F	Similar to useCovariate3DigitIcd9Ambulatory180d, but now only if the frequency of the ICD-9 code is higher than the 75th percentile.
useCovariateDrugExposure	A boolean value (TRUE/FALSE) to determine if covariates derived from DRUG_EXPOSURE table will be created and included in future models.
useCovariateIngredientExposure180d	A boolean value (TRUE/FALSE) to determine if covariates will be created and used in models that look for presence/absence of drug ingredients within inpatient setting in 180d window prior to or on cohort index date. Only applicable if useCovariateDrugExposure = TRUE.

useCovariateIngredientExposure180dMedF	Similar to useCovariateIngredientExposure180d, but now only if the frequency of the ingredient is higher than the median.
useCovariateIngredientExposure180d75F	Similar to useCovariateIngredientExposure180d, but now only if the frequency of the ingredient is higher than the 75th percentile.
useCovariateProcedureOccurrence	A boolean value (TRUE/FALSE) to determine if covariates derived from PROCEDURE_OCCURRENCE table will be created and included in future models.
useCovariateProcedureOccurrenceInpatient180d	A boolean value (TRUE/FALSE) to determine if covariates will be created and used in models that look for presence/absence of procedures within inpatient setting in 180d window prior to or on cohort index date. Only applicable if useCovariateProcedureOccurrence = TRUE.
useCovariateProcedureOccurrenceInpatient180dMedF	Similar to useCovariateProcedureOccurrenceInpatient180d, but now only if the frequency of the procedure code is higher than the median.
useCovariateProcedureOccurrenceInpatient180d75F	Similar to useCovariateProcedureOccurrenceInpatient180d, but now only if the frequency of the procedure code is higher than the 75th percentile.
useCovariateProcedureOccurrenceAmbulatory180d	A boolean value (TRUE/FALSE) to determine if covariates will be created and used in models that look for presence/absence of procedures within ambulatory setting in 180d window prior to or on cohort index date. Only applicable if useCovariateProcedureOccurrence = TRUE.
useCovariateProcedureOccurrenceAmbulatory180dMedF	Similar to useCovariateProcedureOccurrenceAmbulatory180d, but now only if the frequency of the procedure code is higher than the median.
useCovariateProcedureOccurrenceAmbulatory180d75F	Similar to useCovariateProcedureOccurrenceAmbulatory180d, but now only if the frequency of the procedure code is higher than the 75th percentile.
excludedCovariateConceptIds	A list of concept IDs that should NOT be used to construct covariates.
includedCovariateConceptIds	A list of concept IDs that should be used to construct covariates.
deleteCovariatesSmallCount	A numeric value used to remove covariates that occur in both cohorts fewer than deleteCovariateSmallCounts time.

## Details

creates an object specifying how covariates should be constructed from data in the CDM model.

## Value

An object of type HdpsCovariateSettings, to be used in other functions.



**Arguments**

**specifications** A specifications object for generating a table using the [createTable1](#) function.

**covariateSettings** The covariate settings object to use as the basis for the filtered covariate settings.

**includedCovariateConceptIds** A list of concept IDs that should be used to construct covariates.

**addDescendantsToInclude** Should descendant concept IDs be added to the list of concepts to include?

**excludedCovariateConceptIds** A list of concept IDs that should NOT be used to construct covariates.

**addDescendantsToExclude** Should descendant concept IDs be added to the list of concepts to exclude?

**includedCovariateIds** A list of covariate IDs that should be restricted to.

**Value**

A covariate settings object, for example to be used when calling the [getDbCovariateData](#) function.

---

```
createTemporalCovariateSettings
      Create covariate settings
```

---

**Description**

Create covariate settings

**Usage**

```
createTemporalCovariateSettings(useDemographicsGender = FALSE,
  useDemographicsAge = FALSE, useDemographicsAgeGroup = FALSE,
  useDemographicsRace = FALSE, useDemographicsEthnicity = FALSE,
  useDemographicsIndexYear = FALSE, useDemographicsIndexMonth = FALSE,
  useDemographicsPriorObservationTime = FALSE,
  useDemographicsPostObservationTime = FALSE,
  useDemographicsTimeInCohort = FALSE,
  useDemographicsIndexYearMonth = FALSE, useConditionOccurrence = FALSE,
  useConditionOccurrenceInpatient = FALSE, useConditionEraStart = FALSE,
  useConditionEraOverlap = FALSE, useConditionEraGroupStart = FALSE,
  useConditionEraGroupOverlap = FALSE, useDrugExposure = FALSE,
  useDrugEraStart = FALSE, useDrugEraOverlap = FALSE,
  useDrugEraGroupStart = FALSE, useDrugEraGroupOverlap = FALSE,
  useProcedureOccurrence = FALSE, useDeviceExposure = FALSE,
  useMeasurement = FALSE, useMeasurementValue = FALSE,
  useMeasurementRangeGroup = FALSE, useObservation = FALSE,
  useCharlsonIndex = FALSE, useDcsi = FALSE, useChads2 = FALSE,
  useChads2Vasc = FALSE, useDistinctConditionCount = FALSE,
  useDistinctIngredientCount = FALSE, useDistinctProcedureCount = FALSE,
  useDistinctMeasurementCount = FALSE, useDistinctObservationCount = FALSE,
```



```

useVisitCount = FALSE, useVisitConceptCount = FALSE,
temporalStartDays = -365:-1, temporalEndDays = -365:-1,
includedCovariateConceptIds = c(), addDescendantsToInclude = FALSE,
excludedCovariateConceptIds = c(), addDescendantsToExclude = FALSE,
includedCovariateIds = c())

```

## Arguments

```

useDemographicsGender
  Gender of the subject. (analysis ID 1)
useDemographicsAge
  Age of the subject on the index date (in years). (analysis ID 2)
useDemographicsAgeGroup
  Age of the subject on the index date (in 5 year age groups) (analysis ID 3)
useDemographicsRace
  Race of the subject. (analysis ID 4)
useDemographicsEthnicity
  Ethnicity of the subject. (analysis ID 5)
useDemographicsIndexYear
  Year of the index date. (analysis ID 6)
useDemographicsIndexMonth
  Month of the index date. (analysis ID 7)
useDemographicsPriorObservationTime
  Number of days of observation time preceding the index date. (analysis ID 8)
useDemographicsPostObservationTime
  Number of days of observation time preceding the index date. (analysis ID 9)
useDemographicsTimeInCohort
  Number of days of observation time preceding the index date. (analysis ID 10)
useDemographicsIndexYearMonth
  Calendar month of the index date. (analysis ID 11)
useConditionOccurrence
  One covariate per condition in the condition_occurrence table starting in the
  time window. (analysis ID 101)
useConditionOccurrenceInpatient
  One covariate per condition observed in an inpatient setting in the condition_occurrence
  table starting in the time window. (analysis ID 102)
useConditionEraStart
  One covariate per condition in the condition_era table starting in the time win-
  dow. (analysis ID 201)
useConditionEraOverlap
  One covariate per condition in the condition_era table overlapping with any part
  of the time window. (analysis ID 202)
useConditionEraGroupStart
  One covariate per condition era rolled up to SNOMED groups in the condi-
  tion_era table starting in the time window. (analysis ID 203)
useConditionEraGroupOverlap
  One covariate per condition era rolled up to SNOMED groups in the condi-
  tion_era table overlapping with any part of the time window. (analysis ID 203)

```

<code>useDrugExposure</code>	One covariate per drug in the <code>drug_exposure</code> table starting in the time window. (analysis ID 301)
<code>useDrugEraStart</code>	One covariate per drug in the <code>drug_era</code> table starting in the time window. (analysis ID 401)
<code>useDrugEraOverlap</code>	One covariate per drug in the <code>drug_era</code> table overlapping with any part of the time window. (analysis ID 402)
<code>useDrugEraGroupStart</code>	One covariate per drug rolled up to ATC groups in the <code>drug_era</code> table starting in the time window. (analysis ID 403)
<code>useDrugEraGroupOverlap</code>	One covariate per drug rolled up to ATC groups in the <code>drug_era</code> table overlapping with any part of the time window. (analysis ID 403)
<code>useProcedureOccurrence</code>	One covariate per procedure in the <code>procedure_occurrence</code> table in the time window. (analysis ID 501)
<code>useDeviceExposure</code>	One covariate per device in the device exposure table starting in the time window. (analysis ID 601)
<code>useMeasurement</code>	One covariate per measurement in the measurement table in the time window. (analysis ID 701)
<code>useMeasurementValue</code>	One covariate containing the value per measurement-unit combination in the time window. If multiple values are found, the last is taken. (analysis ID 702)
<code>useMeasurementRangeGroup</code>	Covariates indicating whether measurements are below, within, or above normal range within the time period. (analysis ID 703)
<code>useObservation</code>	One covariate per observation in the observation table in the time window. (analysis ID 801)
<code>useCharlsonIndex</code>	The Charlson comorbidity index (Romano adaptation) using all conditions prior to the window end. (analysis ID 901)
<code>useDcsi</code>	The Diabetes Comorbidity Severity Index (DCSI) using all conditions prior to the window end. (analysis ID 902)
<code>useChads2</code>	The CHADS2 score using all conditions prior to the window end. (analysis ID 903)
<code>useChads2Vasc</code>	The CHADS2VASC score using all conditions prior to the window end. (analysis ID 904)
<code>useDistinctConditionCount</code>	The number of distinct condition concepts observed in the time window. (analysis ID 905)
<code>useDistinctIngredientCount</code>	The number of distinct ingredients observed in the time window. (analysis ID 906)
<code>useDistinctProcedureCount</code>	The number of distinct procedures observed in the time window. (analysis ID 907)



```
useConditionEraOverlap = FALSE,
useConditionEraGroupStart = FALSE,
useConditionEraGroupOverlap = TRUE,
useDrugExposure = FALSE,
useDrugEraStart = FALSE,
useDrugEraOverlap = FALSE,
useDrugEraGroupStart = FALSE,
useDrugEraGroupOverlap = TRUE,
useProcedureOccurrence = TRUE,
useDeviceExposure = TRUE,
useMeasurement = TRUE,
useMeasurementValue = FALSE,
useMeasurementRangeGroup = TRUE,
useObservation = TRUE,
useCharlsonIndex = TRUE,
useDcsi = TRUE,
useChads2 = TRUE,
useChads2Vasc = TRUE,
useDistinctConditionCount = FALSE,
useDistinctIngredientCount = FALSE,
useDistinctProcedureCount = FALSE,
useDistinctMeasurementCount = FALSE,
useDistinctObservationCount = FALSE,
useVisitCount = FALSE,
useVisitConceptCount = FALSE,
temporalStartDays = -365:-1,
temporalEndDays = -365:-1,
includedCovariateConceptIds = c(),
addDescendantsToInclude = FALSE,
excludedCovariateConceptIds = c(),
addDescendantsToExclude = FALSE,
includedCovariateIds = c())
```

FeatureExtraction	<i>FeatureExtraction</i>
<b>Description</b>	
FeatureExtraction	
filterByRowId	<i>Filter covariates by row ID</i>

**Description**

Filter covariates by row ID

**Usage**

```
filterByRowId(object, rowIds)
```

**Arguments**

object	Either an object of type covariateData, or an ffd object containing covariate values.
rowIds	A vector (or ff object) containing the rowIds to keep.

**Value**

Either an object of type covariateData, or an ffd object containing covariate values. (depending on the type of the object argument.

---

getDbCohortAttrCovariatesData

*Getcovariate information from the database through the cohort\_attribute table*

---

**Description**

Constructs a large default set of covariates for one or more cohorts using data in the CDM schema. Includes covariates for all drugs, drug classes, condition, condition classes, procedures, observations, etc.

**Usage**

```
getDbCohortAttrCovariatesData(connection, oracleTempSchema = NULL,
  cdmDatabaseSchema, cohortTable = "#cohort_person", cohortId = -1,
  cdmVersion = "5", rowIdField = "subject_id", covariateSettings,
  aggregated = FALSE)
```

**Arguments**

connection	A connection to the server containing the schema as created using the connect function in the DatabaseConnector package.
oracleTempSchema	A schema where temp tables can be created in Oracle.
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'.
cohortTable	Name of the table holding the cohort for which we want to construct covariates. If it is a temp table, the name should have a hash prefix, e.g. '#temp_table'. If it is a non-temp table, it should include the database schema, e.g. 'cdm_database.cohort'.
cohortId	For which cohort ID should covariates be constructed? If set to -1, covariates will be constructed for all cohorts in the specified cohort table.
cdmVersion	The version of the Common Data Model used. Currently only cdmVersion = "5" is supported.
rowIdField	The name of the field in the cohort temp table that is to be used as the row_id field in the output table. This can be especially usefull if there is more than one period per person.

**covariateSettings**

An object of type covariateSettings as created using the [createCohortAttrCovariateSettings](#) function.

**aggregated**

Should aggregate statistics be computed instead of covariates per cohort entry?

**Details**

This function uses the data in the CDM to construct a large set of covariates for the provided cohort. The cohort is assumed to be in an existing temp table with these fields: 'subject\_id', 'cohort\_definition\_id', 'cohort\_start\_date'. Optionally, an extra field can be added containing the unique identifier that will be used as rowID in the output. Typically, users don't call this function directly but rather use the [getDbCovariateData](#) function instead.

**Value**

Returns an object of type covariateData, containing information on the baseline covariates. Information about multiple outcomes can be captured at once for efficiency reasons. This object is a list with the following components:

**covariates** An ffdF object listing the baseline covariates per person in the cohorts. This is done using a sparse representation: covariates with a value of 0 are omitted to save space. The covariates object will have three columns: rowId, covariateId, and covariateValue. The rowId is usually equal to the person\_id, unless specified otherwise in the rowIdField argument.

**covariateRef** An ffdF object describing the covariates that have been extracted.

**metaData** A list of objects with information on how the covariateData object was constructed.

---

getDbCovariateData	<i>Get covariate information from the database</i>
--------------------	--

---

**Description**

Uses one or several covariate builder functions to construct covariates.

**Usage**

```
getDbCovariateData(connectionDetails = NULL, connection = NULL,
  oracleTempSchema = NULL, cdmDatabaseSchema, cdmVersion = "5",
  cohortTable = "cohort", cohortDatabaseSchema = cdmDatabaseSchema,
  cohortTableIsTemp = FALSE, cohortId = -1, rowIdField = "subject_id",
  covariateSettings, aggregated = FALSE)
```

**Arguments****connectionDetails**

An R object of type connectionDetails created using the function createConnectionDetails in the DatabaseConnector package. Either the connection or connectionDetails argument should be specified.

**connection**

A connection to the server containing the schema as created using the connect function in the DatabaseConnector package. Either the connection or connectionDetails argument should be specified.

oracleTempSchema	A schema where temp tables can be created in Oracle.
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'.
cdmVersion	Define the OMOP CDM version used: currently supported is "5".
cohortTable	Name of the (temp) table holding the cohort for which we want to construct covariates
cohortDatabaseSchema	If the cohort table is not a temp table, specify the database schema where the cohort table can be found. On SQL Server, this should specify both the database and the schema, so for example 'cdm_instance.dbo'.
cohortTableIsTemp	Is the cohort table a temp table?
cohortId	For which cohort ID should covariates be constructed? If set to -1, covariates will be constructed for all cohorts in the specified cohort table.
rowIdField	The name of the field in the cohort table that is to be used as the row_id field in the output table. This can be especially usefull if there is more than one period per person.
covariateSettings	Either an object of type covariateSettings as created using one of the create-Covariate functions, or a list of such objects.
aggregated	Should aggregate statistics be computed instead of covariates per cohort entry?

## Details

This function uses the data in the CDM to construct a large set of covariates for the provided cohort. The cohort is assumed to be in an existing table with these fields: 'subject\_id', 'cohort\_definition\_id', 'cohort\_start\_date'. Optionally, an extra field can be added containing the unique identifier that will be used as rowID in the output.

## Value

Returns an object of type covariateData, containing information on the covariates.

---

getDbDefaultCovariateData

*Get default covariate information from the database*

---

## Description

Constructs a large default set of covariates for one or more cohorts using data in the CDM schema. Includes covariates for all drugs, drug classes, condition, condition classes, procedures, observations, etc.

## Usage

```
getDbDefaultCovariateData(connection, oracleTempSchema = NULL,
  cdmDatabaseSchema, cohortTable = "#cohort_person", cohortId = -1,
  cdmVersion = "5", rowIdField = "subject_id", covariateSettings,
  targetDatabaseSchema, targetCovariateTable, targetCovariateRefTable,
  targetAnalysisRefTable, aggregated = FALSE)
```

## Arguments

connection	A connection to the server containing the schema as created using the connect function in the DatabaseConnector package.
oracleTempSchema	A schema where temp tables can be created in Oracle.
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'.
cohortTable	Name of the table holding the cohort for which we want to construct covariates. If it is a temp table, the name should have a hash prefix, e.g. '#temp_table'. If it is a non-temp table, it should include the database schema, e.g. 'cdm_database.cohort'.
cohortId	For which cohort ID should covariates be constructed? If set to -1, covariates will be constructed for all cohorts in the specified cohort table.
cdmVersion	The version of the Common Data Model used. Currently only cdmVersion = "5" is supported.
rowIdField	The name of the field in the cohort temp table that is to be used as the row_id field in the output table. This can be especially usefull if there is more than one period per person.
covariateSettings	Either an object of type covariateSettings as created using one of the create-Covariate functions, or a list of such objects.
targetDatabaseSchema	(Optional) The name of the database schema where the resulting covariates should be stored.
targetCovariateTable	(Optional) The name of the table where the resulting covariates will be stored. If not provided, results will be fetched to R. The table can be a permanent table in the targetDatabaseSchema or a temp table. If it is a temp table, do not specify targetDatabaseSchema.
targetCovariateRefTable	(Optional) The name of the table where the covariate reference will be stored.
targetAnalysisRefTable	(Optional) The name of the table where the analysis reference will be stored.
aggregated	Should aggregate statistics be computed instead of covariates per cohort entry?

## Details

This function uses the data in the CDM to construct a large set of covariates for the provided cohort. The cohort is assumed to be in an existing temp table with these fields: 'subject\_id', 'cohort\_definition\_id', 'cohort\_start\_date'. Optionally, an extra field can be added containing the unique identifier that will be used as rowID in the output. Typically, users don't call this function directly but rather use the [getDbCovariateData](#) function instead.



**Value**

Returns an object of type `covariateData`, containing information on the baseline covariates. Information about multiple outcomes can be captured at once for efficiency reasons. This object is a list with the following components:

**covariates** An `ffdf` object listing the baseline covariates per person in the cohorts. This is done using a sparse representation: covariates with a value of 0 are omitted to save space. The covariates object will have three columns: `rowId`, `covariateId`, and `covariateValue`. The `rowId` is usually equal to the `person_id`, unless specified otherwise in the `rowIdField` argument.

**covariateRef** An `ffdf` object describing the covariates that have been extracted.

**metaData** A list of objects with information on how the `covariateData` object was constructed.

---

`getDbHdpsCovariateData`

*Get HDPS covariate information from the database*

---

**Description**

Constructs the set of covariates for one or more cohorts using data in the CDM schema. This implements the covariates typically used in the HDPS algorithm.

**Usage**

```
getDbHdpsCovariateData(connection, oracleTempSchema = NULL, cdmDatabaseSchema,
  cohortTable = "cohort_person", cohortId = -1, cdmVersion = "5",
  rowIdField = "subject_id", covariateSettings, aggregated = FALSE)
```

**Arguments**

<code>connection</code>	A connection to the server containing the schema as created using the <code>connect</code> function in the <code>DatabaseConnector</code> package.
<code>oracleTempSchema</code>	A schema where temp tables can be created in Oracle.
<code>cdmDatabaseSchema</code>	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 <code>'cdm_instance.dbo'</code> .
<code>cohortTable</code>	Name of the table holding the cohort for which we want to construct covariates. If it is a temp table, the name should have a hash prefix, e.g. <code>'#temp_table'</code> . If it is a non-temp table, it should include the database schema, e.g. <code>'cdm_database.cohort'</code> .
<code>cohortId</code>	For which cohort ID should covariates be constructed? If set to -1, covariates will be constructed for all cohorts in the specified cohort table.
<code>cdmVersion</code>	The version of the Common Data Model used. Currently only <code>cdmVersion = "5"</code> is supported.
<code>rowIdField</code>	The name of the field in the cohort temp table that is to be used as the <code>row_id</code> field in the output table. This can be especially usefull if there is more than one period per person.
<code>covariateSettings</code>	An object of type <code>covariateSettings</code> as created using the <a href="#">createHdpsCovariateSettings</a> function.
<code>aggregated</code>	Should aggregate statistics be computed instead of covariates per cohort entry?

## Details

This function uses the data in the CDM to construct a large set of covariates for the provided cohort. The cohort is assumed to be in an existing temp table with these fields: 'subject\_id', 'cohort\_definition\_id', 'cohort\_start\_date'. Optionally, an extra field can be added containing the unique identifier that will be used as rowID in the output. Typically, users don't call this function directly but rather use the [getDbCovariateData](#) function instead.

## Value

Returns an object of type covariateData, containing information on the baseline covariates. Information about multiple outcomes can be captured at once for efficiency reasons. This object is a list with the following components:

**covariates** An ffdm object listing the baseline covariates per person in the cohorts. This is done using a sparse representation: covariates with a value of 0 are omitted to save space. The covariates object will have three columns: rowId, covariateId, and covariateValue. The rowId is usually equal to the person\_id, unless specified otherwise in the rowIdField argument.

**covariateRef** An ffdm object describing the covariates that have been extracted.

**metaData** A list of objects with information on how the covariateData object was constructed.

---

```
getDefaultTable1Specifications
```

*Get the default table 1 specifications*

---

## Description

Loads the default specifications for a table 1, to be used with the [createTable1](#) function.

## Usage

```
getDefaultTable1Specifications()
```

## Value

A specifications objects.

---

```
loadCovariateData
```

*Load the covariate data from a folder*

---

## Description

loadCovariateData loads an object of type covariateData from a folder in the file system.

## Usage

```
loadCovariateData(file, readOnly = FALSE)
```

**Arguments**

file	The name of the folder containing the data.
readOnly	If true, the data is opened read only.

**Details**

The data will be written to a set of files in the folder specified by the user.

**Value**

An object of class covariateData.

**Examples**

```
# todo
```

---

saveCovariateData	<i>Save the covariate data to folder</i>
-------------------	--

---

**Description**

saveCovariateData saves an object of type covariateData to folder.

**Usage**

```
saveCovariateData(covariateData, file)
```

**Arguments**

covariateData	An object of type covariateData as generated using getDbCovariateData.
file	The name of the folder where the data will be written. The folder should not yet exist.

**Details**

The data will be written to a set of files in the folder specified by the user.

**Examples**

```
# todo
```

---

tidyCovariateData	<i>Tidy covariate data</i>
-------------------	----------------------------

---

**Description**

Tidy covariate data

**Usage**

```
tidyCovariateData(covariateData, covariates, covariateRef, populationSize,
  minFraction = 0.001, normalize = TRUE, removeRedundancy = TRUE)
```

**Arguments**

covariateData	An object as generated using the <a href="#">getDbCovariateData</a> function. If provided, the covariates, covariateRef, and populationSize arguments will be ignored.
covariates	An ffdF object with the covariate values in sparse format. Will be ignored if covariateData is provided.
covariateRef	An ffdF object with the covariate definitions. Will be ignored if covariateData is provided. Only needed when removeRedundancy = TRUE.
populationSize	An integer specifying the total number of unique cohort entries (rowIds). Will be ignored if covariateData is provided. Only needed when removeRedundancy = TRUE.
minFraction	Minimum fraction of the population that should have a non-zero value for a covariate for that covariate to be kept. Set to 0 to don't filter on frequency.
normalize	Normalize the covariates? (dividing by the max)
removeRedundancy	Should redundant covariates be removed?

**Details**

Normalize covariate values by dividing by the max and/or remove redundant covariates and/or remove infrequent covariates.

# Index

aggregateCovariates, [2](#)

byMaxFf, [3](#)  
bySumFf, [3](#)

computeStandardizedDifference, [4](#)  
convertPrespecSettingsToDetailedSettings,  
    [4](#)  
createAnalysisDetails, [5](#), [18](#), [19](#)  
createCohortAttrCovariateSettings, [6](#),  
    [30](#)  
createCovariateSettings, [4](#), [7](#)  
createDefaultCovariateSettings, [17](#)  
createDefaultTemporalCovariateSettings,  
    [18](#)  
createDetailedCovariateSettings, [5](#), [18](#)  
createDetailedTemporalCovariateSettings,  
    [5](#), [19](#)  
createHdpsCovariateSettings, [20](#), [33](#)  
createTable1, [23](#), [24](#), [34](#)  
createTable1CovariateSettings, [23](#)  
createTemporalCovariateSettings, [24](#)

FeatureExtraction, [28](#)  
FeatureExtraction-package  
    (FeatureExtraction), [28](#)  
filterByRowId, [28](#)

getDbCohortAttrCovariatesData, [29](#)  
getDbCovariateData, [24](#), [30](#), [30](#), [32](#), [34](#), [36](#)  
getDbDefaultCovariateData, [31](#)  
getDbHdpsCovariateData, [33](#)  
getDefaultTable1Specifications, [34](#)

loadCovariateData, [34](#)

saveCovariateData, [35](#)

tidyCovariateData, [36](#)