

# Package ‘CohortDiagnostics’

April 21, 2021

**Type** Package

**Title** Diagnostics for OHDSI Cohorts

**Version** 2.1.0

**Date** 2021-04-21

**Maintainer** Gowtham Rao <rao@ohdsi.org>

**Description** Diagnostics for cohorts that use the OMOP Common Data Model and the OHDSI tools.

**Depends** DatabaseConnector (>= 4.0.0),  
FeatureExtraction (>= 3.1.1),  
R (>= 4.0.0)

**Imports** Andromeda,  
checkmate,  
CirceR,  
digest,  
dplyr (>= 1.0.0),  
ParallelLogger (>= 2.0.0),  
pool,  
readr (>= 1.4.0),  
rlang,  
RJSONIO,  
ROhdsiWebApi (>= 1.2.0),  
SqlRender (>= 1.7.0),  
stringr,  
tidyr (>= 1.0.0),  
methods

**Suggests** purrr,  
shiny,  
shinydashboard,  
shinyWidgets,  
ggplot2,  
ggiraph,  
DT,  
plotly,  
htmltools,  
RColorBrewer,  
scales,  
knitr,  
rmarkdown,  
Eunomia,

testthat,  
RSQLite ( $\geq 2.2.1$ ),  
withr,  
zip

**Remotes** ohdsi/Eunomia,  
ohdsi/FeatureExtraction,  
ohdsi/ROhdsiWebApi,  
ohdsi/CirceR

**License** Apache License

**VignetteBuilder** knitr

**URL** <https://ohdsi.github.io/CohortDiagnostics>, <https://github.com/OHDSI/CohortDiagnostics>

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

**RoxygenNote** 7.1.1

**Encoding** UTF-8

**Language** en-US

## R topics documented:

checkInputFileEncoding . . . . .	2
createResultsDataModel . . . . .	3
getCohortCounts . . . . .	3
getResultsDataModelSpecifications . . . . .	4
instantiateCohortSet . . . . .	4
launchCohortExplorer . . . . .	7
launchDiagnosticsExplorer . . . . .	8
preMergeDiagnosticsFiles . . . . .	9
runCohortDiagnostics . . . . .	9
uploadResults . . . . .	13

<b>Index</b>	<b>15</b>
--------------	-----------

---

checkInputFileEncoding

*Check character encoding of input file*

---

### Description

For its input files, CohortDiagnostics only accepts UTF-8 or ASCII character encoding. This function can be used to check whether a file meets these criteria.

### Usage

```
checkInputFileEncoding(fileName)
```

### Arguments

fileName            The path to the file to check

**Value**

Throws an error if the input file does not have the correct encoding.

---

```
createResultsDataModel
```

*Create the results data model tables on a database server.*

---

**Description**

Create the results data model tables on a database server.

**Usage**

```
createResultsDataModel(connection = NULL, connectionDetails = NULL, schema)
```

**Arguments**

connection	An object of type connection as created using the <a href="#">connect</a> function in the DatabaseConnector package. Can be left NULL if connectionDetails is provided, in which case a new connection will be opened at the start of the function, and closed when the function finishes.
connectionDetails	An object of type connectionDetails as created using the <a href="#">createConnectionDetails</a> function in the DatabaseConnector package. Can be left NULL if connection is provided.
schema	The schema on the postgres server where the tables will be created.

**Details**

Only PostgreSQL servers are supported.

---

```
getCohortCounts
```

*Count the cohort(s)*

---

**Description**

Computes the subject and entry count per cohort

**Usage**

```
getCohortCounts(
  connectionDetails = NULL,
  connection = NULL,
  cohortDatabaseSchema,
  cohortTable = "cohort",
  cohortIds = c()
)
```

**Arguments**

connectionDetails	An object of type connectionDetails as created using the <a href="#">createConnectionDetails</a> function in the DatabaseConnector package. Can be left NULL if connection is provided.
connection	An object of type connection as created using the <a href="#">connect</a> function in the DatabaseConnector package. Can be left NULL if connectionDetails is provided, in which case a new connection will be opened at the start of the function, and closed when the function finishes.
cohortDatabaseSchema	Schema name where your cohort table resides. Note that for SQL Server, this should include both the database and schema name, for example 'scratch.dbo'.
cohortTable	Name of the cohort table.
cohortIds	The cohort Id(s) used to reference the cohort in the cohort table. If left empty, all cohorts in the table will be included.

**Value**

A tibble with cohort counts

---

getResultsDataModelSpecifications

*Get specifications for Cohort Diagnostics results data model*

---

**Description**

Get specifications for Cohort Diagnostics results data model

**Usage**

```
getResultsDataModelSpecifications()
```

**Value**

A tibble data frame object with specifications

---

instantiateCohortSet    *Instantiate a set of cohort*

---

**Description**

This function instantiates a set of cohort in the cohort table, using definitions that are fetched from a WebApi interface. Optionally, the inclusion rule statistics are computed and stored in the inclusionStatisticsFolder.

**Usage**

```

instantiateCohortSet(
  connectionDetails = NULL,
  connection = NULL,
  cdmDatabaseSchema,
  vocabularyDatabaseSchema = cdmDatabaseSchema,
  tempEmulationSchema = NULL,
  oracleTempSchema = NULL,
  cohortDatabaseSchema = cdmDatabaseSchema,
  cohortTable = "cohort",
  cohortIds = NULL,
  packageName = NULL,
  cohortToFile = "settings/CohortsToCreate.csv",
  baseUrl = NULL,
  cohortSetReference = NULL,
  generateInclusionStats = FALSE,
  inclusionStatisticsFolder = NULL,
  createCohortTable = TRUE,
  incremental = FALSE,
  incrementalFolder = NULL
)

```

**Arguments**

- |                          |  |
|--------------------------|--|
| connectionDetails        | An object of type connectionDetails as created using the <a href="#">createConnectionDetails</a> function in the DatabaseConnector package. Can be left NULL if connection is provided.  |
| connection               | An object of type connection as created using the <a href="#">connect</a> function in the DatabaseConnector package. Can be left NULL if connectionDetails is provided, in which case a new connection will be opened at the start of the function, and closed when the function finishes. |
| 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'.   |
| vocabularyDatabaseSchema | Schema name where your OMOP vocabulary data resides. This is commonly the same as cdmDatabaseSchema. Note that for SQL Server, this should include both the database and schema name, for example 'vocabulary.dbo'.  |
| tempEmulationSchema      | Some database platforms like Oracle and Impala do not truly support temp tables. To emulate temp tables, provide a schema with write privileges where temp tables can be created.  |
| oracleTempSchema         | DEPRECATED by DatabaseConnector: use tempEmulationSchema instead.  |
| cohortDatabaseSchema     | Schema name where your cohort table resides. Note that for SQL Server, this should include both the database and schema name, for example 'scratch.dbo'.   |
| cohortTable              | Name of the cohort table.  |
| cohortIds                | Optionally, provide a subset of cohort IDs to restrict the construction to.  |

<code>packageName</code>	The name of the package containing the cohort definitions. Can be left NULL if <code>baseUrl</code> and <code>cohortSetReference</code> have been specified.
<code>cohortToCreateFile</code>	The location of the cohortToCreate file within the package. Is ignored if <code>baseUrl</code> and <code>cohortSetReference</code> have been specified. The cohortToCreateFile must be .csv file that is expected to be read into a dataframe object identical to requirements for <code>cohortSetReference</code> argument. This csv file is expected to be encoded in either ASCII or UTF-8, if not, an error message will be displayed and process stopped.
<code>baseUrl</code>	The base URL for the WebApi instance, for example: "http://server.org:80/WebAPI". Can be left NULL if <code>packageName</code> and <code>cohortToCreateFile</code> have been specified.
<code>cohortSetReference</code>	A data frame with four columns, as described in the details. Can be left NULL if <code>packageName</code> and <code>cohortToCreateFile</code> have been specified.
<code>generateInclusionStats</code>	Compute and store inclusion rule statistics?
<code>inclusionStatisticsFolder</code>	The folder where the inclusion rule statistics are stored. Can be left NULL if <code>generateInclusionStats = FALSE</code> .
<code>createCohortTable</code>	Create the cohort table? If <code>incremental = TRUE</code> and the table already exists this will be skipped.
<code>incremental</code>	Create only cohorts that haven't been created before?
<code>incrementalFolder</code>	If <code>incremental = TRUE</code> , specify a folder where records are kept of which definition has been executed.

## Details

Currently two ways of executing this function are supported, either (1) [Package Mode] embedded in a study package, assuming the cohort definitions are stored in that package using the `ROhdsiWebApi::insertCohortDefinition` or (2) [WebApi Mode] By using a WebApi interface to retrieve the cohort definitions.

When using this function in Package Mode: Use the `packageName` and `cohortToCreateFile` to specify the name of the study package, and the name of the cohortToCreate file within that package, respectively

When using this function in WebApi Mode: use the `baseUrl` and `cohortSetReference` to specify how to connect to the WebApi, and which cohorts to fetch, respectively.

Note: if the parameters for both Package Mode and WebApi Mode are provided, then Package mode is preferred.

The `cohortSetReference` argument must be a data frame with the following columns:

**referentConceptId** A standard omop concept id that serves as the referant phenotype definition for the cohort Id.

**cohortId** The cohort Id is the id used to identify a cohort definition. This is required to be unique. It will be used to create file names. It is recommended to be  $(\text{referentConceptId} * 1000) + \text{a number between 3 to 999}$

**webApiCohortId** Cohort Id in the webApi/atlas instance. It is a required field to run Cohort Diagnostics in WebApi mode. It is discarded in package mode.

**cohortName** The full name of the cohort. This will be shown in the Shiny app.

**logicDescription** A human understandable brief description of the cohort definition. This logic does not have to a fully specified description of the cohort definition, but should provide enough context to help user understand the meaning of the cohort definition

## Value

A data frame with cohort counts

---

launchCohortExplorer    *Launch the CohortExplorer Shiny app*

---

## Description

Launch the CohortExplorer Shiny app

## Usage

```
launchCohortExplorer(
  connectionDetails,
  cdmDatabaseSchema,
  cohortDatabaseSchema,
  cohortTable,
  cohortId,
  sampleSize = 100,
  subjectIds = 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'.
cohortDatabaseSchema	Schema name where your cohort table resides. Note that for SQL Server, this should include both the database and schema name, for example 'scratch.dbo'.
cohortTable	Name of the cohort table.
cohortId	The ID of the cohort.
sampleSize	Number of subjects to sample from the cohort. Ignored if subjectIds is specified.
subjectIds	A vector of subject IDs to view.

## Details

Launches a Shiny app that allows the user to explore a cohort of interest.

---

launchDiagnosticsExplorer

*Launch the Diagnostics Explorer Shiny app*


---

## Description

Launch the Diagnostics Explorer Shiny app

## Usage

```
launchDiagnosticsExplorer(
  dataFolder = "data",
  dataFile = "PreMerged.RData",
  connectionDetails = NULL,
  resultsDatabaseSchema = NULL,
  vocabularyDatabaseSchema = NULL,
  vocabularyDatabaseSchemas = resultsDatabaseSchema,
  aboutText = NULL,
  runOverNetwork = FALSE,
  port = 80,
  launch.browser = FALSE
)
```

## Arguments

dataFolder	A folder where the premerged file is stored. Use the <a href="#">preMergeDiagnosticsFiles</a> function to generate this file.
dataFile	(Optional) The name of the .RData file with results. It is commonly known as the Premerged file.
connectionDetails	An object of type connectionDetails as created using the <a href="#">createConnectionDetails</a> function in the DatabaseConnector package, specifying how to connect to the server where the CohortDiagnostics results have been uploaded using the <a href="#">uploadResults</a> function.
resultsDatabaseSchema	The schema on the database server where the CohortDiagnostics results have been uploaded.
vocabularyDatabaseSchema	(Deprecated) Please use vocabularyDatabaseSchemas.
vocabularyDatabaseSchemas	(optional) A list of one or more schemas on the database server where the vocabulary tables are located. The default value is the value of the resultsDatabaseSchema. We can provide a list of vocabulary schema that might represent different versions of the OMOP vocabulary tables. It allows us to compare the impact of vocabulary changes on Diagnostics.
aboutText	Text (using HTML markup) that will be displayed in an About tab in the Shiny app. If not provided, no About tab will be shown.
runOverNetwork	(optional) Do you want the app to run over your network?
port	(optional) Only used if runOverNetwork = TRUE.
launch.browser	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 diagnostics

---

```
preMergeDiagnosticsFiles
```

*Premerge Shiny diagnostics files*

---

**Description**

This function combines diagnostics results from one or more databases into a single file. The result is a single file that can be used as input for the Diagnostics Explorer Shiny app.

It also checks whether the results conform to the results data model specifications.

**Usage**

```
preMergeDiagnosticsFiles(dataFolder, tempFolder = tempdir())
```

**Arguments**

dataFolder	folder where the exported zip files for the diagnostics are stored. Use the <a href="#">runCohortDiagnostics</a> function to generate these zip files. Zip files containing results from multiple databases may be placed in the same folder.
tempFolder	A folder on the local file system where the zip files are extracted to. Will be cleaned up when the function is finished. Can be used to specify a temp folder on a drive that has sufficient space if the default system temp space is too limited.

---

```
runCohortDiagnostics
```

*Run cohort diagnostics*

---

**Description**

Runs the cohort diagnostics on all (or a subset of) the cohorts instantiated using the `ROhdsiWebApi::insertCohortDefinition` function. Assumes the cohorts have already been instantiated.

Characterization: If `runTemporalCohortCharacterization` argument is `TRUE`, then the following default covariateSettings object will be created using `RFeatureExtraction::createTemporalCovariateSettings`. Alternatively, a covariate setting object may be created using the above as an example.

**Usage**

```
runCohortDiagnostics(
  packageName = NULL,
  cohortToCreateFile = "settings/CohortsToCreate.csv",
  baseUrl = NULL,
  cohortSetReference = NULL,
  phenotypeDescriptionFile = NULL,
  connectionDetails = NULL,
  connection = NULL,
  cdmDatabaseSchema,
```

```

oracleTempSchema = NULL,
tempEmulationSchema = NULL,
cohortDatabaseSchema,
vocabularyDatabaseSchema = cdmDatabaseSchema,
cohortTable = "cohort",
cohortIds = NULL,
inclusionStatisticsFolder = file.path(exportFolder, "inclusionStatistics"),
exportFolder,
databaseId,
databaseName = databaseId,
databaseDescription = databaseId,
cdmVersion = 5,
runInclusionStatistics = TRUE,
runIncludedSourceConcepts = TRUE,
runOrphanConcepts = TRUE,
runTimeDistributions = TRUE,
runVisitContext = TRUE,
runBreakdownIndexEvents = TRUE,
runIncidenceRate = TRUE,
runCohortOverlap = TRUE,
runCohortCharacterization = TRUE,
covariateSettings = createDefaultCovariateSettings(),
runTemporalCohortCharacterization = TRUE,
temporalCovariateSettings = createTemporalCovariateSettings(useConditionOccurrence =
  TRUE, useDrugEraStart = TRUE, useProcedureOccurrence = TRUE, useMeasurement = TRUE,
  temporalStartDays = c(-365, -30, 0, 1, 31), temporalEndDays = c(-31, -1, 0, 30, 365)),
minCellCount = 5,
incremental = FALSE,
incrementalFolder = file.path(exportFolder, "incremental")
)

```

## Arguments

packageName	The name of the package containing the cohort definitions. Can be left NULL if baseUrl and cohortSetReference have been specified.
cohortToCreateFile	The location of the cohortToCreate file within the package. Is ignored if baseUrl and cohortSetReference have been specified. The cohortToCreateFile must be .csv file that is expected to be read into a dataframe object identical to requirements for cohortSetReference argument. This csv file is expected to be encoded in either ASCII or UTF-8, if not, an error message will be displayed and process stopped.
baseUrl	The base URL for the WebApi instance, for example: "http://server.org:80/WebAPI". Can be left NULL if packageName and cohortToCreateFile have been specified.
cohortSetReference	A data frame with four columns, as described in the details. Can be left NULL if packageName and cohortToCreateFile have been specified.
phenotypeDescriptionFile	(Optional) The location of the phenotype description file within the package. The file must be .csv file and have the following columns that may be read into

following data types: phenotypeId (double), phenotypeName (character), referentConceptId (double), clinicalDescription (character), literatureReview (character), phenotypeNotes (character). Note: the field names are in snake\_case. Also, character fields should not have 'NA' in it. 'NA's are commonly added by R when R functions exports data from dataframe into CSV. Instead please use "" (empty string) to represent absence of data. The literature\_review field is expected to be a html link to page that contains resources related to literature review for the phenotype, and will be used to create link-out object. This csv file is expected to be encoded in either ASCII or UTF-8, if not, an error message will be displayed and process stopped.

#### connectionDetails

An object of type connectionDetails as created using the [createConnectionDetails](#) function in the DatabaseConnector package. Can be left NULL if connection is provided.

#### connection

An object of type connection as created using the [connect](#) function in the DatabaseConnector package. Can be left NULL if connectionDetails is provided, in which case a new connection will be opened at the start of the function, and closed when the function finishes.

#### 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

DEPRECATED by DatabaseConnector: use tempEmulationSchema instead.

#### tempEmulationSchema

Some database platforms like Oracle and Impala do not truly support temp tables. To emulate temp tables, provide a schema with write privileges where temp tables can be created.

#### cohortDatabaseSchema

Schema name where your cohort table resides. Note that for SQL Server, this should include both the database and schema name, for example 'scratch.dbo'.

#### vocabularyDatabaseSchema

Schema name where your OMOP vocabulary data resides. This is commonly the same as cdmDatabaseSchema. Note that for SQL Server, this should include both the database and schema name, for example 'vocabulary.dbo'.

#### cohortTable

Name of the cohort table.

#### cohortIds

Optionally, provide a subset of cohort IDs to restrict the diagnostics to.

#### inclusionStatisticsFolder

The folder where the inclusion rule statistics are stored. Can be left NULL if runInclusionStatistics = FALSE.

#### exportFolder

The folder where the output will be exported to. If this folder does not exist it will be created.

#### databaseId

A short string for identifying the database (e.g. 'Synpuf').

#### databaseName

The full name of the database. If NULL, defaults to databaseId.

#### databaseDescription

A short description (several sentences) of the database. If NULL, defaults to databaseId.

#### cdmVersion

The version of the OMOP CDM. Default 5. (Note: only 5 is supported.)

<code>runInclusionStatistics</code>	Generate and export statistic on the cohort inclusion rules?
<code>runIncludedSourceConcepts</code>	Generate and export the source concepts included in the cohorts?
<code>runOrphanConcepts</code>	Generate and export potential orphan concepts?
<code>runTimeDistributions</code>	Generate and export cohort time distributions?
<code>runVisitContext</code>	Generate and export index-date visit context?
<code>runBreakdownIndexEvents</code>	Generate and export the breakdown of index events?
<code>runIncidenceRate</code>	Generate and export the cohort incidence rates?
<code>runCohortOverlap</code>	Generate and export the cohort overlap? Overlaps are checked within cohortIds that have the same phenotype ID sourced from the CohortSetReference or cohortToCreateFile.
<code>runCohortCharacterization</code>	Generate and export the cohort characterization? Only records with values greater than 0.0001 are returned.
<code>covariateSettings</code>	Either an object of type covariateSettings as created using one of the createCovariateSettings function in the FeatureExtraction package, or a list of such objects.
<code>runTemporalCohortCharacterization</code>	Generate and export the temporal cohort characterization? Only records with values greater than 0.001 are returned.
<code>temporalCovariateSettings</code>	Either an object of type covariateSettings as created using one of the createTemporalCovariateSettings function in the FeatureExtraction package, or a list of such objects.
<code>minCellCount</code>	The minimum cell count for fields contains person counts or fractions.
<code>incremental</code>	Create only cohort diagnostics that haven't been created before?
<code>incrementalFolder</code>	If incremental = TRUE, specify a folder where records are kept of which cohort diagnostics has been executed.

## Details

Currently two ways of executing this function are supported, either (1) [Package Mode] embedded in a study package, assuming the cohort definitions are stored in that package using the `ROhdsiWebApi::insertCohortDe` or (2) [WebApi Mode] By using a WebApi interface to retrieve the cohort definitions.

When using this function in Package Mode: Use the `packageName` and `cohortToCreateFile` to specify the name of the study package, and the name of the cohortToCreate file within that package, respectively

When using this function in WebApi Mode: use the `baseUrl` and `cohortSetReference` to specify how to connect to the WebApi, and which cohorts to fetch, respectively.

Note: if the parameters for both Package Mode and WebApi Mode are provided, then Package mode is preferred.

The cohortSetReference argument must be a data frame with the following columns:

**referentConceptId** A standard omop concept id that serves as the referant phenotype definition for the cohort Id.

**cohortId** The cohort Id is the id used to identify a cohort definition. This is required to be unique. It will be used to create file names. It is recommended to be (referentConceptId \* 1000) + a number between 3 to 999

**webApiCohortId** Cohort Id in the webApi/atlas instance. It is a required field to run Cohort Diagnostics in WebApi mode. It is discarded in package mode.

**cohortName** The full name of the cohort. This will be shown in the Shiny app.

**logicDescription** A human understandable brief description of the cohort definition. This logic does not have to a fully specified description of the cohort definition, but should provide enough context to help user understand the meaning of the cohort definition

---

uploadResults	<i>Upload results to the database server.</i>
---------------	---

---

## Description

Requires the results data model tables have been created using the [createResultsDataModel](#) function.

Set the POSTGRES\_PATH environmental variable to the path to the folder containing the psql executable to enable bulk upload (recommended).

## Usage

```
uploadResults(
  connectionDetails = NULL,
  schema,
  zipFileName,
  forceOverWriteOfSpecifications = FALSE,
  purgeSiteDataBeforeUploading = TRUE,
  tempFolder = tempdir()
)
```

## Arguments

connectionDetails	An object of type connectionDetails as created using the <a href="#">createConnectionDetails</a> function in the DatabaseConnector package.
schema	The schema on the postgres server where the tables have been created.
zipFileName	The name of the zip file.
forceOverWriteOfSpecifications	If TRUE, specifications of the phenotypes, cohort definitions, and analysis will be overwritten if they already exist on the database. Only use this if these specifications have changed since the last upload.
purgeSiteDataBeforeUploading	If TRUE, before inserting data for a specific databaseId all the data for that site will be dropped. This assumes the input zip file contains the full data for that data site.

`tempFolder` A folder on the local file system where the zip files are extracted to. Will be cleaned up when the function is finished. Can be used to specify a temp folder on a drive that has sufficient space if the default system temp space is too limited.

# Index

checkInputFileEncoding, [2](#)  
connect, [3–5](#), [11](#)  
createConnectionDetails, [3–5](#), [7](#), [8](#), [11](#), [13](#)  
createResultsDataModel, [3](#), [13](#)  
  
getCohortCounts, [3](#)  
getResultsDataModelSpecifications, [4](#)  
  
instantiateCohortSet, [4](#)  
  
launchCohortExplorer, [7](#)  
launchDiagnosticsExplorer, [8](#)  
  
preMergeDiagnosticsFiles, [8](#), [9](#)  
  
runCohortDiagnostics, [9](#), [9](#)  
  
uploadResults, [8](#), [13](#)