Running cohort diagnostics

Gowtham Rao

2021-07-22

Contents

1 Configuring the connection to the server

We need to tell R how to connect to the server where the data are. `CohortDiagnostics` uses the `DatabaseConnector` package, which provides the `createConnectionDetails` function. Type `?createConnectionDetails` for the specific settings required for the various database management systems (DBMS). For example, one might connect to a PostgreSQL database using this code:

```r
connectionDetails <- createConnectionDetails(dbms = "postgresql",
                                             server = "localhost/ohdsi",
                                             user = "joe",
                                             password = "supersecret")


cdmDatabaseSchema <- "my_cdm_data"
tempEmulationSchema <- NULL
cohortDatabaseSchema <- "my_schema"
cohortTable <- "my_cohort_table"
```

The last four lines define the `cdmDatabaseSchema`, `tempEmulationSchema`, `cohortDatabaseSchema`, and `cohortTable` variables. We'll use the `cdmDatabaseSchema` later to tell R where the data in CDM format live. The `tempEmulationSchema` is needed only for Oracle users, since Oracle does not support temporary tables. The `cohortDatabaseSchema`, and `cohortTable` specify where we want to instantiate our cohorts. Note that for Microsoft SQL Server, databaseschemas need to specify both the database and the schema, so for example `cdmDatabaseSchema <- "my_cdm_data.dbo"`.

1.1 Creating a new cohort table

In order to run most of the cohort diagnostics, we need to instantiate the cohorts. The best way is to instantiate the cohorts in a new cohort table. We can use the `createCohortTable` to create an empty cohort table:

```r
createCohortTable(connectionDetails = connectionDetails,
                  cohortDatabaseSchema = cohortDatabaseSchema,
                  cohortTable = cohortTable)
```

Note this this function will **delete the table if it already exists** before creating it.

1.2 Inclusion rules

We have the option to also generate inclusion rule statistics while the cohorts are instantiated (recommended). If we want to do this, we need to provide a folder where the inclusion rule statistics will be stored for later use.

2 Cohort Diagnostics Output

Once diagnostics is completed, a zip file will have been created in the specified export folder. This zip file can be shared between sites, as it does not contain patient-identifiable information. When unzipped, the zip file will contain several .csv files that maybe easily audited. Note that cell counts smaller than 5 have been removed, as specified using the `minCellCount` argument, to ensure non-identifiability.

2.1 Creating a pre merged file

Assuming you completed the steps described above for one or more databases, you should now have a set of zip files, one per database. Make sure to place all zip files in a single folder, for example `c:/temp/allZipFiles`.

Optionally, we can pre-merge the zip files, so we can view results in the Shiny app:

```
preMergeDiagnosticsFiles("C:/temp/allZipFiles")
```