Cohort Diagnostics in an OHDSI Package mode

Gowtham A. Rao

2021-04-13

Contents

1 Introduction

This vignette describes how to create a self contained shareable study package that can execute cohort diagnostics on a given set of cohort definitions.

In this vignette, you will learn how to - use the Cohort Diagnostics template package - Hydrate the template package with the cohort definitions you want to diagnose. - Execute the diagnostics on one or more data sources. - Review the result set using the Diagnostics Explorer R shiny app of Cohort Diagnostics - (optional) Submit the results of your cohort diagnostics to the OHDSI Phenotype library.

1.1 Assumptions:

- You have a set of valid cohort definition sql files (in parameterized OHDSI Sql format) that you want to run diagnostics on, or you are able to export cohort definitions that you want to diagnose from Atlas by using the export tab within cohort definition module of Atlas or using ROhdsiWebApi as described here.
- You have access to person level data in OMOP CDM V5.x + format on a database, are able to execute and instantiate cohorts in cohort tables (i.e. you have access to a schema in the same database with Create, Read, Update, Delete privileges.)

1.2 identifying cohorts:

The cohorts to diagnose should have two attributes to them.

- **id**: A unique integer identifier for the cohort you are diagnosing. If you created your cohort using atlas, this is your atlas id.

- **cohortName**: The unique string name for your cohort. Usually it is the same as your Atlas cohort name, but it can be any name you wish to identify your cohort. This name will be used to display your cohort in the Diagnostics Explorer shiny app.

Example:

id	cohortName
17561	[PL 4112853001] Malignant tumor of breast referent concept incident cohort: First occurrence of referent concept + descendants with >=365d prior observation

1.3   option A: Using Hydra and ROhdsiWebApi

The skeleton cohort diagnostics study package is here. A skeleton package is a special package that is designed to be used with Hydra. The input for Hydra is study a specifications file (in json format). Example is here.

We can create this json using the shiny app 'Diagnostics Editor' available here(XXXXXX). A version is deployed at data.ohdsi.org/XXXXXXXXXXXXX. The user interface of this shiny app allows you to make selections that are then used to create the specifications json file. The application also allows you to hydrate a package using Hydra, and download a .zip that contains your study package.

Alternatively, you can take the example specification-json file and edit it in notepad. You can then use to create the study package as follows

XXXXXXXXXx

Now you may install your package and execute the study.

1.4   option B: using only ROhdsiWebApi

Start with the skeleton cohort diagnostics study package is here. See Vignette from ROhdsiWebApi here. Create the 'CohortsToCreate.csv' file by extracting the cohorts from your atlas as described. Now, open your study package and run

```
ROhdsiWebApi::insertCohortDefinitionSetInPackage(fileName = "CohortsToCreate.csv",
                                                 baseUrl = Sys.getenv("baseUrl"),
                                                 insertTableSql = TRUE,
                                                 insertCohortCreationR = TRUE,
                                                 generateStats = TRUE,
                                                 packageName = "SkeletonCohortDiagnosticsStudy")
```

This will create the relevant files in your package.

The next step is to rename the generic name of the package 'SkeletonCohortDiagnosticsStudy' into something more meaningful to identify your study such as 'CandidateHypertensionCohorts'. One way of doing this is to use Notepad ++ find and replace 'SkeletonCohortDiagnosticsStudy' with a name for your study. Also change the 'SkeletonCohortDiagnosticsStudy.Rproj' to your study name 'CandidateHypertensionCohorts.Rproj'.

Now you may install your package and execute the study.

1.5   Option C: Manual (advanced users)

Start with the skeleton cohort diagnostics study package is here. Manually populate the CohortToCreate.csv file with the atlasId, cohortId and cohortName. In most cases atlasId = cohortId and should be an integer. cohortName is a meaningful name to identify your cohorts.

Next - create a new folder called 'sql' in the 'inst' folder and put all your SQLs. The sql file should be named to match the cohortId e.g. if the cohortId is 1232 then the sql file should be 1232.sql. Optionally, you can also add the cohort json in another folder in 'inst' folder of your package called 'cohorts'. The names of the cohorts should be the same as the sql files, except they will end with .json instead of .sql.

If your cohorts have inclusion-rules, then you will have to create a file called 'InclusionRules.csv' and populate it as shown in this example:

cohortName	ruleSequence	ruleName	cohortId
14907	0	With no history of gastrointestinal hemorrhage	14907
18349	0	Gender is male	18349
18350	0	Gender is male	18350
18350	1	Metabolic disease at any time	18350
18352	0	With gastrointestinal hemorrhage at some point	18352

Now you may install your package and execute the study.