Clinical Characteristics V1 demo

Contents


```
library(ClinicalCharacteristics)
```

1 Introduction

We introduce a new R package called ClinicalCharacteristics which is a table-shell approach to characterization in real-world data mapped to the OMOP CDM. The purpose of introducing this package was to deal with table-shells in studies. Often we encounter situations were our study counter-parts wish to characterize an indication based on a series of specific comorbidities identified using a set of codes. Using base OHDSI tools this request is becomes non-trivial, involving lots of data-wrangling. Our goal was to build a tool that could characterize specific comorbidities of interest and only pull results of interest omitting superflous output.

1.1 Good ole' `FeatureExtraction`

The common characterization tool `FeatureExtraction` was designed for drawing as many arbitrary features for high-dimenionsal modelling. In the modelling situation, it is to our advantage to pull any concept from a domain table to use in a model and optimize its fit. Further, a handy side-effect of `FeatureExtraction` is its ability to quantify counts and proportions for the presence of concepts. This is a fantastic tool that has been a cornerstone in OHDSI software for many years. However, it yields output that either requires additional data wrangling or does not quite answer the specified characterization. The alternative is to write custom sql to characterize. While this solution solves the immediate project problem it no longer abides by standardized analytics which is cornerstone to network studies.

1.2 ClinicalCharacteristics Input

A characterization requires input based on: one, the target cohort to use as a reference point; two, a time window to observe the occurrence of a characteristic; and three identification of a characteristic listed as follows:

- ConceptSet: a set of concepts when bundled identify a clinical event. For example, identification of Type 2 Diabetes could be done with SNOMED concept plus descendant concepts.
- Cohort: a cohort identifying a clinical event. For example identification of patients with Type 2 Diabetes who enter the cohort with a SNOMED concept and exit the cohort at the end of continuous observation
- ConceptSetGroup: bundling of multiple concept sets to identify a clinical event. For example identification of Smoking using a condition SNOMED concept and a concept in the observation table.
- Demographic: identification of a person characteristic such as age, gender, race, and ethnicity among others.

For concept sets, cohorts and concept set groups, this tool can handle the following types of aggregations:

- Presence: identification of characteristic in a specified time window
- Count: enumeration of characteristic in a specified time window
- TimeDiff: summary of time from the cohort index to the event date

The final feature implemented in `ClinicalCharacteristics` is how to report a characteristic; either categorizing a continuous variable or scoring a categorical variable. For example, we can choose to report the distribution of age in 5 year, 10 year or another custom grouping. Another example we can categorize the number of SGLT2 medications taken in a time window by converting the continuous count to a categorical break (eg. 1, 2, 3, 4+ prescriptions). Finally we can score a categorical variable by applying a weight to the presence of a comorbidity (a Charlson score for example).

1.3 ClinicalCharacteristics Design

The design of ClinicalCharacteristics follows a simple workflow implemented in SQL.

- Step 0: Implement meta table for table shell
- Step 1: Identify occurrences of a clinical characteristic
- Step 2: Synthesize occurrences to "1 row per patient"
- Step 3: Aggregate patient level data in the database
- Step 4: Clean and label aggregate table for saving and presentation

Similar to circe, `ClinicalCharacteristics` takes the input of the table shell specification and renders a sql file to build the table shell. The routing of these steps is handled in R, using R6 classes.

1.3.1 Roadmap

This R package is still under-development as of its presentation at the 2024 OHDSI symposium. An earlier version of the package (v0.3.4) is available for use but will be subject to major refactoring.

2 Example

We provide an example for running `ClinicalCharacteristics` below:

```
# create execution settings R6 class
executionSettings <- createExecutionSettings(
  connectionDetails = connectionDetails,
  cdmDatabaseSchema = "cdm_schema",
  workDatabaseSchema = "work_schema",
  tempEmulationSchema = "temp_schema",
  targetCohortTable = "cohorts",
  cdmSourceName = "my_omop_data"
)

# Using CAPR to create concept sets to implement in ClinChar

library(Capr)

## baseline conditions
cs1 <- list(
  'hf' = cs(descendants(316139), name = "hf"),
  'ckd' = cs(descendants(46271022), name = "ckd")
)

## drugs
cs2 <- list(
```

```r
  'metformin' = cs(descendants(1503297), name = "metformin"),
  'sglt2' = cs(descendants(1123627), name = "sglt2")
)

# Identification of time windows
tw <- list(
  timeInterval(lb = -365, rb = -1)
)

tw1 <- list(
  timeInterval(lb = 0, rb = 90),
  timeInterval(lb = 0, rb = 365)
)


# Specification of Table Shell

tableShell <- createTableShell(
  title = "Test",
  targetCohorts = list(
    createCohortInfo(id = 1, name = "T2D")
  ),
  lineItems = lineItems(
    createAgeLineItem()), # age as
    createGenderLineItem(),
    createConceptSetLineItemBatch(
      name = "Baseline Conditions",
      domain = "condition_occurrence",
      statistic = createPresence(),
      conceptSets = cs1,
      timeInterval = tw
    ),
    createConceptSetLineItemBatch(
      name = "Drug Usage",
      domain = "drug_exposure",
      statistic = createPresence(),
      conceptSets = cs2,
      timeIntervals = tw1
    )
  )
)
# generation of table shell
res <- generateTableShell(tableShell, executionSettings)
# review sql
reviewTableShellSql(tableShell = tableShell, executionSettings)
# explore results
previewTableShell(res, type = "continuous")
previewTableShell(res, type = "categorical")
```