ClinicalCharacteristics: A table-shell approach to Characterization in OMOP

Martin Lavallee, Ajit Londhe, Katy Sadowski, Ron Herrera
Boehringer Ingelheim

Background

When conducting real-world evidence (RWE) studies, analysts often need to populate table-shells defined by clinical colleagues. In principle, these table-shells should be simple to generate, as they consist primarily of counts and percentages for categorical variables and summary statistics for continuous variables. Surprisingly, this task proves to be non-trivial when using the available OHDSI tools, whether that is because the requested table shells include summaries on OMOP Common Data Model tables lacking full analytic support (e.g. Cost, Episode, Location) or because the significant time needed wrangling cohort features via tools like `FeatureExtraction`[1].

Though not OHDSI's preferred approach, table-shells are a reality of RWE studies. OHDSI would benefit from tools that simplify this task. We have developed a new OHDSI R package called `ClinicalCharacteristics`[2] to characterize cohorts of interest via a table-shell approach. Our main design principle is explicitness; instead of pulling all concepts from a domain, we limit characterization to the specified concept set(s). In this presentation, we will showcase how to use `ClinicalCharacteristics` in four scenarios: first, describing demographics; second, describing comorbidities and co-medications; third, describing health care resource utilization; and fourth, characterizing lab measurements.

Methods

We developed `ClinicalCharacteristics` for two reasons: (1) to adhere to explicitly authored table-shell designs and (2) to minimize the need for custom code. The typical OHDSI method for characterization is `FeatureExtraction`, which is designed to extract as many clinical features as possible to use as covariates for high-dimensional modelling. While helpful for large-scale analyses, `FeatureExtraction`'s fit for populating table-shells is awkward because it pulls all concepts from a domain and performs characterization at the individual concept level. One could also use ATLAS[3] to characterize, however, it lacks coverage for the cost table and geographies, while also being a graphical user interface that runs counter to our preference of programmatically pipelining many analyses together,

`ClinicalCharacteristics` adheres to `circe-be`[4] logic, meaning all SQL is OHDSI-SQL compliant. Specifically, `ClinicalCharacteristics` leverages `circe-be` to build concept set logic to enumerate occurrences within an OMOP domain. A characterization requires the following set of components:

- **Target Cohort**: the study population we wish to characterize, given it has been generated to a cohort table (using methods such as `CohortGenerator`[5]).
- **Domain**: specification of the OMOP table we want to link to for a summary
- **Concept Set**: "logical expressions that can be used to identify a set of concepts to be used throughout" standardize OHDSI analyses[6]
- **Window**: the period of time during which we wish to observe the event
- **Summary type**: either a binary measure of presence or a numeric measure of time, cost or count
- **Attribute**: (optional) allows for counting the occurrence of an event based on other logic; for example, only count conditions found in the inpatient setting or drugs with a specific days supply
- **Transformation**: (optional) provides ability to dichotomize a continuous value into categories (e.g. age groups or lab value ranges). Other times we wish to apply a weight to a binary to generate a continuous score (e.g. Charlson comorbidity index)

`ClinicalCharacteristics` aims to capture repeated patterns based on these criteria to provide a large set of possible summary tables.

Results

We will demonstrate four use cases of `ClinicalCharacteristics`: demographics, comorbidities, HCRU and labs using the following workflow:

1) Build target cohorts
2) Identify concept sets to summarize
3) Build the `ClinicalCharacteristics` object, which populates the intended table-shell
4) Run the characterization query
5) Review the results

A code snippet of the general workflow is provided in the appendix.

Conclusion

With `ClinicalCharacteristics`, we aim to provide the OHDSI community with a valuable new approach to characterizing cohorts via a table-shell approach. `ClinicalCharacteristics` provides commonly needed analyses as well as characterization of previously unsupported domains and features.

Citations

1. Schuemie, Martijn, Marc Suchard, Patrick Ryan, Jenna Reps, Anthony Sena, and Ger Inberg. 2024. *FeatureExtraction: Generating Features for a Cohort*. https://ohdsi.github.io/FeatureExtraction.

2. Lavallee, Martin, Katy Sadowski, and Ajit Londhe. 2024. ClinicalCharacteristics: Table Shell Approach to OMOP Characterization. https://github.com/OHDSI/ClinicalCharacteristics.
3. Knoll, Christopher, Anthony Sena, and OHDSI et al. 2024. *Atlas: An Open Source Software Tool for Researchers to Conduct Scientific Analyses on Standardized Observational Data*. https://github.com/OHDSI/Atlas.
4. Knoll, Chris, Anthony Sena. 2024. *Circe-Be: CIRCE is a cohort definition and syntax compiler tool for OMOP CDMv5.* https://github.com/OHDSI/circe-be.
5. Sena A, Gilbert J, Rao G, Schuemie M (2024). *CohortGenerator: An R Package for Cohort Generation Against the OMOP CDM*. https://ohdsi.github.io/CohortGenerator/, https://github.com/OHDSI/CohortGenerator.
6. OHDSI. *The Book of OHDSI: Atlas.* https://ohdsi.github.io/TheBookOfOhdsi/OhdsiAnalyticsTools.html#atlas.

Appendix

```
library(ClinicalCharacteristics)
library(Capr)

# make connection details
connectionDetails <- DatabaseConnector::createConnectionDetails(
  dbms = '<dbms>',
  user = "<user>",
  password = "<password>",
  connectionString = "<jdbcString>"
)

executionSettings <- list(
  databaseName = "my_omop_data",
  cdmDatabaseSchema = "cdm",
  workDatabaseSchema = "scratch",
  cohortTable = "cohorts"
)

# define output folder
outputFolder <- here::here("my_output")

# establish connection
connection <- DatabaseConnector::connect(connectionDetails)

clinChar <- makeClinChar(
  targetCohortIds = 1,
  targetCohortNames = "Target",
  dbms = connection@dbms,
  database = executionSettings$databaseName,
  cdmDatabaseSchema = executionSettings$cdmDatabaseSchema,
```

```
  workDatabaseSchema = executionSettings$workDatabaseSchema,
  cohortTable = executionSettings$cohortTable
) |>
  addAgeChar(categorize = age10yrGrp()) |> # age summary reported in 10yr gro
ups
  addGenderChar() |> # gender summary
  addRaceChar() |> # race summary
  addYearChar(categorize = year5yrGrp()) |> # year summary reported in 5 yr g
roups
  addConditionPresence( # presence of charlson comorbidities
    conceptSets = charlsonConcepts(),
    timeWindows = makeTimeTable(time_a = -365, time_b = -1),
    limit = "first",
    score = charlsonIndexScore(ageId = 1) # score binary
  ) |>
  addDrugPresence( # presence of atc2 class
    conceptSets = atcConcepts(),
    timeWindows = makeTimeTable(
      time_a = c(-365, 0, 0, 0),
      time_b = c(-1, 183, 365, 730)
      ),
    limit = "first"
  ) |>
  addVisitCount( # count number of visits for IP, OP, ER
    conceptSets = standardVisitConcepts(),
    timeWindows = makeTimeTable(
      time_a = c(-365, 0, 0, 0),
      time_b = c(-1, 183, 365, 730)
    )
  )


# run the clinical characteristics
dt <- runClinicalCharacteristics(connection = connection,
                                       clinChar = clinChar,
                                       saveName = "internals_test",
                                       savePath = outputFolder,
                                       dropDat = FALSE)

# preview the categorical results
previewClincalCharacteristics(dt, type = "categorical")
# preview the continuous results
previewClincalCharacteristics(dt, type = "continuous")
#build a report
createReport(clinChar, outputFolder = outputFolder)
```