

Software demonstration for PatientProfiles: an R package for patient characterisation based on pre-defined phenotypes and cohorts



An R package to add characteristics to your cohort CDM environment.

Background: Several R packages have been developed to facilitate standardised analytics for OMOP CDM as part of the Darwin EU project, including CDMConnector (1) and IncidencePrevalence (2). CDMConnector provides a standardised approach to read, write, and connect to the OMOP CDM. IncidencePrevalence supports the analysis of incidence and prevalence for data mapped to the OMOP CDM in a reproducible manner. Although there are existing tools that provide functionality for defining and characterising cohorts of patients (e.g. the FeatureExtraction and Characterization R packages), no existing package provides a simple and flexible user interface for describing the demographics of patients in the OMOP CDM and working with the intersections between multiple patient cohorts. This functionality is often needed in epidemiologic research for creating a cohort-based “table one” as well as estimating outcomes for cohorts of interest.

PatientProfiles aims to

- 1) provide functionality for identifying the demographics of patients in the OMOP CDM (e.g., age, sex, days of prior history)
- 2) indicate the intersection between patient cohorts (e.g., identify days from entering one cohort to the entering another).

Methods: PatientProfiles is written in R (version 4.2.1) and was built using the most common R package tools like roxygen or devtools. This R package was built using a test-driven development approach. Tests were conducted on mock OMOP CDM data using the mockPatientProfiles function (currently >97%). The tests included checks on input/output formats, logical checks, expected true tests, and edge cases. User acceptance tests were also conducted to ensure informative function/argument naming.

Results: PatientProfiles is freely available under the Apache License (Version 2.0) and can be obtained from CRAN (v0.2.0), where full details and instructions for setup and use are provided (3). Detailed vignettes on the package’s functionality are also freely available online from our github repository. (See QR code below)

Conclusion: The R package PatientProfiles provides a standardised process for working with patient characteristics captured in the Observational Medical Outcomes Partnership (OMOP) standard data model. It simplifies writing analytic code, and increases the reliability of analysis, making it a helpful tool for researchers utilising OMOP CDM.

Figure 1. Example code snippets for PatientProfiles

```
Example code snippet: adding patient demographics

cdm$cohort1 <- cdm$cohort1 %>%
  addAge(
    cdm = cdm,
    ageGroup = list(c(0, 18), c(19, 65), c(66, 100))
  ) %>%
  addSex(cdm = cdm) %>%
  addPriorHistory(cdm = cdm)

Example code snippet: cohort intersections

cdm$cohort1 <- cdm$cohort1 %>%
  addCohortIntersectDate(
    cdm = cdm,
    targetCohortTable = "cohort2",
    targetCohortId = 1,
    order = "first",
    window = c(-Inf, Inf)
  )
```

Key functions:

Add individual patient characteristics

Add patient characteristics to a table in the OMOP Common Data Model

addAge()

Compute the age of the individuals at a certain date

addDateOfBirth()

Add a column with the individual birth date

addFutureObservation()

Compute the number of days till the end of the observation period at a certain date

addInObservation()

Indicate if a certain record is within the observation period

addPriorHistory()

Compute the number of days of prior history in the current observation period at a certain date

addSex()

Compute the sex of the individuals

Add a value from a cohort intersection

Add a variable indicating the intersection between a table in the OMOP Common Data Model and a cohort table.

addCohortIntersectCount()

It creates columns to indicate the number of occurrences of intersections with a cohort

addCohortIntersectDate()

Date of cohorts that are present in a certain window

addCohortIntersectDays()

It creates columns to indicate the number of days between the current table and a target cohort

addCohortIntersectFlag()

It creates columns to indicate the presence of cohorts



Click here for more information



Mike Du¹, Yuchen Guo¹, Kim López-Güell¹, Xintong Li¹, Nuria Mercade Besora¹, Daniel Prieto-Alhambra^{1,2}, Edward Burn¹, Marti Catala¹

¹Pharmaco- and Device Epidemiology, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford, UK

²Erasmus University Medical Centre, Rotterdam, The Netherlands