**A framework for cohort building in R: the CohortConstructor package for data mapped to the OMOP Common Data Model**

Cohorts are a key concept in epidemiological research, used to identify groups of people who meet specific criteria over a set period based on their clinical records. However, building and managing cohorts in medical data analysis can be complex, often resulting in code that is difficult to review and reuse.

When working with the OMOP Common Data Model (CDM), efforts have been made to define the `cohort\_table` class along with methods and object attributes to facilitate its use. This class is implemented in the `omopgenerics` R package. The cohort table contains at least four mandatory columns: cohort ID (unique cohort identifier within a cohort table), subject id (unique person identifier), and the cohort start and end dates for each individual.

Additionally, cohort objects have four key attributes: 1) Settings - links the cohort ID to its name, 2) Counts - provides the number of subjects and records in each cohort, 3) Attrition - flow chart of excluded records and individuals for each inclusion criteria, and 4) Code lists - stores clinical concept lists used to define cohort entry, exclusion, or exit. These attributes ensure transparency in cohort creation, facilitate validation, and enable easy dissemination of study results.

To streamline cohort building in R using OMOP data, we developed the R package CohortConstructor. It provides tools for cohort manipulation, including filtering by demographics, calendar time, or presence/absence in other cohorts. Additionally, the package tracks clinical codes used and population attrition for each operation.

CohortConstructor version 0.3.5 is available in CRAN at the time of abstract submission. The development version is publicly available in GitHub: <https://github.com/OHDSI/CohortConstructor/>.

The pipeline to build cohorts begins by creating base cohorts. These can be defined using clinical concepts (e.g., asthma diagnoses) or demographics (e.g., females aged >18). Once base cohorts are established, curation steps are applied to meet study-specific inclusion criteria.

The curation functions cover the most usual operations in cohort studies, as well as more complex cohort manipulations. These functions can be grouped into three categories: 1) Requirement and Filtering – demographic restrictions, event presence/absence conditions, or filtering specific records, 2) Time Manipulation – adjusting entry and exit dates to align with study periods, observation windows, or key events, and 3) Transformation and Combination – Merging, stratifying, collapsing, matching, or intersecting cohorts.

CohortConstructor enables researchers to efficiently build and refine cohorts using validated, and reusable code lists. Its user-friendly interface allows both data scientists and epidemiologists to review and apply study-specific criteria with ease. Additionally, tracking attrition throughout the process enhances cohort validation and supports research dissemination.

In our demo we will demonstrate the use of the package on synthetic data that the audience can also download and run locally. We will show how a variety of patient cohorts can be identified using the package, and how these can then be used as the foundation for subsequent data analyses. In addition, we will explain how the package works behind the scenes so that it works efficiently on big data and across different database management platforms.