# Software demonstration: DrugUtilisation, an R Package to implement Patient-level Drug Utilisation Studies analysis using the OMOP common data model

**Yuchen Guo[1], Mike Du[1], Kim Lopez-Guell[1], Xintong Li[1], Ger Inberg[2], Therese Buckhard[1], Annika M. Jödicke[1], Artem Gorbachev[3], Daniel Prieto-Alhambra[1,2], Edward Burn[1], Martí Català[1]**

**[1] Pharmaco- and Device Epidemiology, Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDROMS), University of Oxford, UK**
**[2] Department of Medical Informatics, Erasmus University Medical Center**
**[3] Odysseus Data Services**

## Background

The efficient management of medication use is essential for patient safety and quality care. Patient-level drug utilisation studies can provide crucial information to enable rational drug usage. These studies involve analysing data from various sources to gain insights into the patterns and determinants of medication use in real-world settings.

DARWIN-EU® aims to produce standardised, reliable, transparent, reproducible, and fast real-world evidence across Europe. In this context, robust analytical pipelines are needed to maximise scalability.

We aimed to build an R package that standardises the way Drug Utilisation Studies are conducted within DARWIN-EU®, in line with an EMA-approved Catalogue of Standard Analyses

## Methods

The *DrugUtilisation* package was developed using the R programming language and followed a test-driven development approach based on DARWIN EU®'s approved Quality Assurance of Software Development procedures. Although all lines of code are written in R, SQL is executed on the database to maximise efficiency. This is possible thanks to the environment created by *CDMConnector* and *tidyverse* functions.

The package includes various functions designed to analyse drug utilization data. To ensure the accuracy and reliability of the package, tests were developed to encode the software's requirements for each function. These tests included checks on the output format of results, logical checks for temporal properties of start and end dates, and edge case handling. Expected errors were also tested, and informative error messages were returned to the user in such circumstances.

The package was documented using the *Roxygen2* package to provide detailed information about each function's usage and inputs. *PaRe* report was used to ensure that the code follow DARWIN-EU code standards. Finally, the package has been submitted to the Comprehensive R Archive Network (CRAN) for distribution to the R community.

## Results

*DrugUtilisation* was built using R version 4.2.1. All lines of code were covered by tests, ensuring that the code can run and produce the desired results. The tests were run in DuckDB, PostgreSQL, SQL Server and Redshift. We aim also to run tests in snowflake and databricks so the package is tested across all database management systems supported within DARWIN-EU initiative.

The main functionalities of the package are:

- generateDrugUtilisationCohortSet: obtain new and prevalent drug user cohorts based on concepts ids with some constrains like washout or prior history.

- addIndication: flag if a certain indication, unknown or none is present within a certain window before drug initiation.
- addDrugUse: obtain the information about the dosing within a certain cohort of a certain ingredient.

The package contains functions to summarise these results like summariseIndication or summariseDrugUse.

It also contains functions to summarise the drug users' characteristics:

- summariseTableOne: standardised baseline characteristics table.
- summariseLargeScaleCharacterization: concept-level characterization within prespecified time windows.
- summariseCharacteristicsFromCodelist: characterization based on bespoke concept sets within prespecified windows. Used to characterize at ATC or ICD10 class level/s.

Output format for summarise functions can be exported into a csv file. All outputs have a common format to facilitate the postprocessing of results and integration within the DARWIN-EU® reporting system.

Drug dosing was calculated following standardized formula and patterns (see accompanying abstract: 'Burkard et al, Paving the way to estimate dose in OMOP CDM').

The package is publicly available on CRAN: https://cran.r-project.org/package=DrugUtilisation.

**Conclusion**

DrugUtilisation can characterise the utilisation of predefined drugs at the patient level across databases mapped to the OMOP common data model. This package included different functionality like getting indications for the drug utilisation cohort and getting dose information for the *drug utilisation cohort. These functionalities* are based on the DARWIN EU® catalogue of Standard Analytics and will be used in upcoming regulatory studies across various clinical settings.