Use cases for development, optimisation and implementation of artificial intelligence methods for real world data analyses in regulatory decision-making and health technology assessment along the product lifecycle (Real4Reg)

First published: 29/06/2023

Last updated: 20/11/2025

Study Ongoing

Administrative details

EU PAS number

EUPAS105544

Study ID

105545

DARWIN EU® study

No

Study countries

Denmark

Finland

Germany

Portugal

Study description

The use of real world data (RWD) is established in regulatory processes such as safety monitoring, but evidentiary value for further use cases, especially in the pre-authorisation and evaluation phase of medicinal products, is rudimentary. Also, the use of RWD in post-authorisation steps is constrained by data variability and by challenges in analysing data from different settings and sources. Moreover, there are emerging opportunities in the use of artificial intelligence (AI), but there is a lack of knowledge on its appropriate application to heterogeneous RWD sources to increase evidentiary value in the regulatory decision-making and health technology assessment (HTA) context. Thus, the development of new and optimised AI-supported methodologies for RWD analyses is essential. In addition, the four use cases to be investigated in this study contain phenotype-specific open questions of high regulatory interest.

Study status

Ongoing

Research institutions and networks

Institutions

Pharmacoepidemiology Research, Federal Institute for Drugs and Medical Devices (BfArM)

Germany

First published: 02/06/2023

Last updated: 23/04/2024

Institution | Regulatory Authority | ENCePP partner

Real-World Evidence Team, University of Eastern Finland (RWE team)

Finland

First published: 20/12/2017

Last updated: 27/08/2024

Institution | Educational Institution | ENCePP partner

Aarhus University & Aarhus University Hospital DEPARTMENT OF CLINICAL EPIDEMIOLOGY

Denmark

First published: 20/07/2021

Last updated: 08/05/2026

Institution | Educational Institution

Data Analytic Center (DAC), Danish Medicine Agency

Denmark

First published: 17/04/2023

Last updated: 17/04/2023

Institution | EU Institution/Body/Agency | ENCePP partner

INFARMED – National Authority of Medicines and Health Products, I.P.

Portugal

First published: 19/11/2025

Last updated: 08/04/2026

Institution | EEA National Competent authority | Healthcare payer | Regulatory Authority | ENCePP partner

Deutsches Zentrum für Neurodegenerative Erkrankungen e. V. (DZNE)

Germany

First published: 20/11/2025

Last updated: 20/11/2025

Institution | Hospital/Clinic/Other health care facility

Fraunhofer Society, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Research Group AI and Data Science Sankt Augustin, Germany;
IT Centre for Science (CSC) Espoo, Finland;
European Organization for Professionals and Patients with ALS (EUpALS) Leuven, Belgium;
European Institute of Women's Health (EIWH) Dublin, Ireland.

Networks

Real4Reg

- Belgium
- Denmark
- Finland
- Germany
- Ireland
- Portugal

First published: 01/02/2024

Last updated: 08/05/2024

Network

Contact details

Study institution contact

Joana Fernandes real4reg@infarmed.pt

Study contact

real4reg@infarmed.pt

Primary lead investigator

Britta Haenisch 0000-0002-4828-6058

Primary lead investigator

ORCID number:

0000-0002-4828-6058

Study timelines

Date when funding contract was signed

Planned: 05/11/2022

Actual: 11/11/2022

Study start date

Planned: 01/09/2023

Actual: 11/09/2023

Date of final study report

Planned: 31/12/2026

Sources of funding

- Other

More details on funding

HORIZON Europe (European Union, European Comission)

Study protocol

Study_protocols_Real4Reg_2023_June_29.pdf (1.35 MB)

Real4Reg_protocols_usecases_1-4.pdf (2.74 MB)

Study_protocols_Real4Reg_2024_April_11.pdf (1.48 MB)

Regulatory

Was the study required by a regulatory body?

No

Is the study required by a Risk Management Plan (RMP)?

Not applicable

Methodological aspects

Study type

Study type list

Study topic:

Disease /health condition

Human medicinal product

Study type:

Non-interventional study

Scope of the study:

Assessment of risk minimisation measure implementation or effectiveness

Disease epidemiology

Drug utilisation

Effectiveness study (incl. comparative)

Method development or testing

Safety study (incl. comparative)

Data collection methods:

Secondary use of data

Study design:

See study protocols.

Main study objective:

The study develops tools & technologies for the effective analyses of real-world data (RWD) in regulatory decision-making & HTA based on four highly relevant uses cases along the pre-& post-authorisation steps of the product life cycle. Portugal, Finland, Denmark, & Germany provide routine care RWD.

Study Design

Non-interventional study design

Cohort

Other

Non-interventional study design, other

External control arm study

Study drug and medical condition

Anatomical Therapeutic Chemical (ATC) code

(A10BH) Dipeptidyl peptidase 4 (DPP-4) inhibitors

Dipeptidyl peptidase 4 (DPP-4) inhibitors

(A10BK) Sodium-glucose co-transporter 2 (SGLT2) inhibitors

Sodium-glucose co-transporter 2 (SGLT2) inhibitors

(J01) ANTIBACTERIALS FOR SYSTEMIC USE

ANTIBACTERIALS FOR SYSTEMIC USE

(J01MA) Fluoroquinolones

Fluoroquinolones

Medical condition to be studied

Type 2 diabetes mellitus

Amyotrophic lateral sclerosis

Breast neoplasm

Additional medical condition(s)

Cardiac arrythmias, Peripheral neuropathies, Drug-induced liver injury, Cardiac failure, Sudden cardiac death

Population studied

Short description of the study population

The source population come from health registers for the entire adult population (≥18 years) in Denmark, Finland, and Portugal and from claims records of public health insurance providers in Germany (~90% of population). Populations studied vary by use case, see settings below.

Age groups

- Adults (18 to < 46 years)
- Adults (46 to < 65 years)
- Adults (65 to < 75 years)
- Adults (75 to < 85 years)
- Adults (85 years and over)

Estimated number of subjects

88000000

Study design details

Setting

Adult population (≥18 years unless otherwise specified).
Inpatient and outpatient data is used.

Setting varies by use case, in short:
1&2: a) persons diagnosed with incident primary breast cancer and external controls,
b) persons diagnosed with incident ALS and external controls,
3. persons with prescriptions of oral antibiotics,
4. persons with prescriptions of noninsulinic antidiabetes drugs (≥40 years).

Study period:

for use cases 1&2: 2000/2005/2008*-2021;

for use case 3: 2010-2021*;

for use case 4: 2012-2021*.

*depending on data availability

Comparators

Use case 3: fluoroquinolones compared to active comparators

Use case 4: SGLT-2 inhibitors compared to active comparators

Outcomes

For details, see protocols.

Use case 1&2: incidence, prevalence, survival time, mortality; further outcomes see protocol;

Use case 3: number of prescriptions, trends in user characteristics, sudden cardiac death or cardiac arrhythmia, sudden cardiac death, aortic aneurysm and dissection, acute toxic liver disease, polyneuropathy;

Use case 4: proportion of SGLT-2 inhibitor prescription/non-insulin antidiabetic prescriptions, change in user characteristics, hospital admission, heart failure diagnosis, mortality.

Data analysis plan

Use Case (UC) 1

* Define a simplified common data model based on OMOP to harmonise metadata

* Determine the population as well as rates of occurrence and outcome using

the STROBE standards (Strengthening the Reporting of Observational studies in Epidemiology).
* Examine heterogeneity within the dataset, including coding practices in the four partners, representativeness etc.
* Develop a workflow to subset and display RWD by user definable inclusion/exclusion criteria; workflow allows to display summary statistics and disease trajectories of selected patients.
* Generate synthetic data based on RWD of BC patients using a generative AI technique such as HALO [https://www.nature.com/articles/s41467-023-41093-0]
UC 2
* Construct synthetic and external control arms for RWD ALS patients (based on data from the ProACT database: https://ncri1.partners.org/ProACT) by implementing different propensity score matching algorithms from the literature
* Generate synthetic data based on ProACT by training genreative AI approaches such as Multimodal Neural ODEs [https://pubmed.ncbi.nlm.nih.gov/17910531/]
UCs 3 and 4
* Drug utilisation study: description of outcomes:
UC 3: proportion of fluorquinolone (FQ) and comparison antibiotics prescriptions out of all oral antibiotic prescriptions, change in user characteristics;
UC 4: proportion of SGLT-2 inhibitors users and DPP-4 inhibitors users out of all non-insulin antidiabetic drug users, change in user characteristics
* Adverse drug reaction study concerning FQ use (UC 3) and effectiveness analysis concerning SGLT-2 inhibitor use (UC4):
· Use active comparator new user (ACNU) design,
· Derive propensity scores for ACNU design from pre-specified covariates using different approaches
· Derive average treatment effect using Target Maximum Likelihood Estimation
· Estimate conditional average treatment effect using different approaches

See study protocols for more details.

Documents

Study, other information

WP leaders R4R_2025.pdf (407.74 KB)

Checklists protocols UC1-4_signed.pdf (9.2 MB)

Declarations_of_Interest_Real4Reg_2025.pdf (2.35 MB)

Study publications

The EU project Real4Reg: unlocking real-world data with AI Background The u…

Supplementing Single-Arm Trials with External Control Arms—Evaluation of German…

Data management

ENCePP Seal

> The use of the ENCePP Seal has been discontinued since February 2025. The ENCePP Seal fields are retained in the display mode for transparency but are no longer maintained.

This study has been awarded the ENCePP seal

Conflicts of interest of investigators

Declarations_of_Interest_Real4Reg_2024.pdf (2.27 MB)

Composition of steering group and observers

WP leaders R4R_2024.pdf (77.94 KB)

Signed code of conduct

ENCePPCoCAnnex3_DeclarationofcompliancewiththeENCePPCodeofConduct_Real4Reg.pdf (411.29 KB)

Signed code of conduct checklist

Checklist_Code_of_Conduct_Real4Reg.pdf (2.64 MB)

Signed checklist for study protocols

Checklist_study_protocols_Real4Reg_spring_2024.pdf (1.52 MB)

Data sources

Data source(s)

Danish registries (access/analysis)

Data source(s), other

- Registo Oncológico Nacional (RON) - Portugal;
- Individual-level data collected from the Finnish national healthcare, mortality and population registers and censuses - Finland;
- Health Data Lab - Germany.

Data sources (types)

Administrative healthcare records (e.g., claims)

Disease registry

Electronic healthcare records (EHR)

Population registry

Use of a Common Data Model (CDM)

CDM mapping

Yes

CDM Mappings

CDM name

OMOP

CDM website

https://www.ohdsi.org/Data-standardization/

Data quality specifications

Check conformance

Unknown

Check completeness

Unknown

Check stability

Unknown

Check logical consistency

Unknown

Data characterisation

Data characterisation conducted

Unknown