Identification of type 2 diabetes cases in a set of databases participating to the EMIF project

First published: 01/10/2015

Last updated: 02/07/2024

Study Finalised

Administrative details

EU PAS number

EUPAS11157

Study ID

21959

DARWIN EU® study

No

Study countries

- Denmark
- Estonia
- Italy
- Netherlands

Spain

United Kingdom

Study description

The European Medical Information Framework (EMIF) project has the main objective of building an infrastructure for the efficient re-use of existing health care data for epidemiological research. Within the project, the EMIF-Platform represents a federation of heterogeneous sources of health data (e.g. administrative, hospital or primary care databases, disease registries, biobanks). One of the major challenges for the EMIF project is to deal with the different characteristics of the participating data sources in order to facilitate the execution of large multi-data base observational studies and generate high quality. For this purpose, a template data derivation process was specifically developed and the identification of Type 2 diabetes mellitus (T2DM) was used as a test case.The objectives of this study are: a) to establish a set of standard algorithms, each based on a single data domain, that will be used for the identification of patients with type 2 diabetes (T2DM) across heterogeneous sources of health data, b) to describe the data source-tailored combinations of standard algorithms recommended by the relevant local data base experts, c) to assess the impact of individual standard algorithms on the population of cases identified across different data sources.

Study status

Finalised

Research institutions and networks

Institutions

Agenzia regionale di sanità della Toscana (ARS Toscana)

Italy

First published: 01/02/2024

Last updated: 23/03/2026

Institution | EU Institution/Body/Agency | ENCePP partner

Epidemiology (Rotterdam Study), Erasmus Medical Center (ErasmusMC)

Netherlands

First published: 13/01/2015

Last updated: 29/10/2015

Institution | Outdated | Educational Institution | ENCePP partner

The PHARMO Institute for Drug Outcomes Research (PHARMO Institute)

Netherlands

First published: 07/01/2022

Last updated: 19/12/2025

Institution | Non-Pharmaceutical company | ENCePP partner

IMIM-Hospital del Mar Medical Research Institute and Universitat Pompeu Fabra Barcelona, Spain, The Health Improvement Network, Cegedim Strategic Data Medical Research Ltd London, United Kingdom, Arsenàl.IT Consortium, Veneto's Research Centre for eHealth Innovation Treviso, Italy, Health Search, Italian College of General Practitioners and Primary Care Florence, Italy, Department of Clinical Epidemiology, Aarhus University Hosptial Aarhus, Denmark, Quretec, Software Technology and Applications Competence Center, University of Tartu Tartu, Estonia

Networks

European Medical Information Framework (EMIF)

European Union

First published: 01/02/2024

Last updated: 12/03/2024

Network

Contact details

Study institution contact

Roberto Giuseppe giuseppe.roberto@ars.toscana.it

Study contact

giuseppe.roberto@ars.toscana.it

Primary lead investigator

Roberto Giuseppe

Primary lead investigator

Study timelines

Date when funding contract was signed

Actual: 01/01/2013

Study start date

Actual: 30/04/2014

Date of final study report

Planned: 31/01/2016

Actual: 01/02/2016

Sources of funding

- Other

More details on funding

EU/EFPIA

Study protocol

Protocol T2DM_V 1.5.pdf (197.81 KB)

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

Other

Study topic, other:

Disease/Epidemiology study

Study type:

Non-interventional study

Scope of the study:

Other

If ‘other’, further details on the scope of the study

Proof of concept study to test a data derivation workflow designed to identify any event of interest from data sources with heterogeneous characteristics

Data collection methods:

Secondary use of data

Main study objective:

- to establish a set of standard algorithms useful to identify of patients with type 2 diabetes (T2DM) across heterogeneous sources of health data- to describe the data source-tailored combinations of standard algorithms recommended by the relevant local data base experts- to assess the impact of individual standard algorithms on the population of cases identified across different data sources

Study Design

Non-interventional study design

Cross-sectional

Study drug and medical condition

Medical condition to be studied

Type 2 diabetes mellitus

Population studied

Short description of the study population

All active type 2 diabetes mellitus subjects in the participating databases at two distinct index dates: 1st January 2009 and 1st January 2012

Age groups

- Term newborn infants (0 – 27 days)
- Infants and toddlers (28 days – 23 months)
- Children (2 to < 12 years)
- Adolescents (12 to < 18 years)
- 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)

Special population of interest

Other

Special population of interest, other

Diabetes mellitus patients

Estimated number of subjects

12.000.000

Study design details

Data analysis plan

This is a descriptive, cross-sectional, retrospective multi-database study. The study population will include all active subjects in the participating databases at two distinct index dates: 1st January 2009 and 1st January 2012, respectively. Results from case-identification algorithms will be computed and presented as age band distribution of the percentage of the data base population identified. A validation of the records extracted in the databases will not be performed, since this is out of the scope of this study. Results will not be intended as disease frequency estimates.

Documents

Study results

Poster_EMIF_ICPE_Boston_2016.pdf (1.08 MB)

Study publications

Roberto G, Leal I, Sattar N, Loomis AK, Avillach P, Egger P, Van Wijngaarden R,...

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.

Data sources

Data source(s)

THIN® (The Health Improvement Network®)

Health Search/IQVIA Health Longitudinal Patient Database

Integrated Primary Care Information (IPCI)

PHARMO Data Network

ARS Toscana

Data source(s), other

IMASIS Spain, EGCUT Estonia, Pedianet Italy, AUH Denmark

Data sources (types)

Administrative healthcare records (e.g., claims)

Drug dispensing/prescription data

Electronic healthcare records (EHR)

Use of a Common Data Model (CDM)

CDM mapping

No

Data quality specifications

Check conformance

Unknown

Check completeness

Unknown

Check stability

Unknown

Check logical consistency

Unknown

Data characterisation

Data characterisation conducted

Unknown