## Clinical Data Warehouse of the Bordeaux University Hospital

First published: 01/02/2024

Last updated: 26/05/2025



## Administrative details

## Administrative details

#### Data source ID

1111112

#### Data source acronym

CDWBordeaux

#### Data holder

Bordeaux University Hospital (CHU de Bordeaux)

#### Data source type

Other

#### Data source type, other

Electronic health records

#### Main financial support

Funding by own institution National, regional, or municipal public funding

#### **Care setting**

Hospital inpatient care Hospital outpatient care

#### Data source qualification

If the data source has successfully undergone a formal qualification process (e.g., from the EMA, ISO or other certifications), this should be described.

No

#### Data source website

https://www.chu-bordeaux.fr/

## Contact details

## Romain Griffier romain.griffier@chu-bordeaux.fr

Main

romain.griffier@chu-bordeaux.fr

## Vianney Jouhet vianney.jouhet@chu-bordeaux.fr

Alternate

vianney.jouhet@chu-bordeaux.fr

## Data source regions and languages

#### **Data source countries**

France

#### **Data source languages**

French

## Data source establishment

Data source established 08/04/2022

#### Data source time span

**First collection:** 15/01/2010 The date when data started to be collected or extracted.

## **Publications**

## Data source publications

Linkage of Hospital Records and Death Certificates by a Search Engine and Machine Learning

The benefit of augmenting open data with clinical data-warehouse EHR for forecasting SARS-CoV-2 hospitalizations in Bordeaux area, France

Appropriateness of psychotropic drug prescriptions in the elderly: structuring tools based on data extracted from the hospital information system to understand physician practices.

## Studies

# List of studies that have been conducted using the data source

DARWIN EU® Characterization of patients with chronic hepatitis B and C

DARWIN EU® Multiple myeloma: patient characterisation, treatments and survival in the period 2012-2022

DARWIN EU® Drug utilization study of prescription opioids

DARWIN EU® - Co-prescribing of endothelin receptor antagonists (ERAs) and phosphodiesterate-5 inhibitors (PDE-5is) in pulmonary arterial hypertension (PAH)

DARWIN EU® Treatment patterns of drugs used in adult and paediatric population with systemic lupus erythematosus

DARWIN EU® Drug utilisation study of medicines with prokinetic properties in children and adults diagnosed with gastroparesis

DARWIN EU® Natural history of dermatomyositis (DM) and polymyositis (PM) in adults and paediatric populations

DARWIN EU® Age specific incidence rates of RSV related disease in Europe

DARWIN EU® Monitoring prescription of essential medicines administered in ICU

DARWIN EU® Overall survival in patients with locally advanced or metastatic non-small cell lung cancer treated with selected immunotherapies as first line of treatment

DARWIN EU® - Chondrosarcoma: patient demographics, treatments, and survival in the period 2010-2023

DARWIN EU® - Characterising interstitial lung disease in Europe

DARWIN EU® – Paracetamol prescribing and paracetamol overdose in Europe: a descriptive analysis of trends and patient characteristics

DARWIN EU® - Prescription trends of ketamine and esketamine

DARWIN EU® – Monitoring prescription of medicines for public health emergencies at risk of shortages

DARWIN EU® Drug Utilisation Study of prescription opioids

DARWIN EU® - Monitoring prescription of essential medicines administered in ICU

## Data elements collected

# The data source contains the following information

#### **Disease information**

Does the data source collect information with a focus on a specific disease? This might be a patient registry or other similar initiatives.

No

#### **Rare diseases**

Are rare diseases captured? In the European Union a rare disease is one that affects no more than 5 people in 10,000.

Yes

#### **Pregnancy and/or neonates**

Does the data source collect information on pregnant women and/or neonatal subpopulation (under 28 days of age)?

Yes

#### Hospital admission and/or discharge

Yes

#### **ICU** admission

Is information on intensive care unit admission available?

Yes

#### **Cause of death**

Captured

#### Cause of death vocabulary

Not coded (Free text)

#### **Prescriptions of medicines**

Captured

#### **Prescriptions vocabulary**

ATC

other

#### Prescriptions vocabulary, other

UCD (french inhospital drug classification) ; Free-text prescription after in-

hospital

#### **Dispensing of medicines**

Captured

#### **Dispensing vocabulary**

ATC

other

#### Dispensing vocabulary, other

UCD (french inhospital drug classification)

#### Advanced therapy medicinal products (ATMP)

Is information on advanced therapy medicinal products included? A medicinal product for human use that is either a gene therapy medicinal product, a somatic cell therapy product or a tissue engineered products as defined in Regulation (EC) No 1394/2007 [Reg (EC) No 1394/2007 Art 1(1)].

No

#### Contraception

Is information on the use of any type of contraception (oral, injectable, devices etc.) available?

Yes

#### Indication for use

Does the data source capture information on the therapeutic indication for the use of medicinal products?

Captured

#### Indication vocabulary

Not coded (Free text)

#### **Medical devices**

Is information on medicinal devices (e.g., pens, syringes, inhalers) available?

Yes

#### Administration of vaccines

Yes

#### Procedures

Does the data source capture information on procedures (e.g., diagnostic tests, therapeutic, surgical interventions)?

Captured

#### **Procedures vocabulary**

Other

#### Procedures vocabulary, other

CCAM (french terminology)

#### Healthcare provider

Is information on the person providing healthcare (e.g., physician, pharmacist, specialist) available? The healthcare provider refers to individual health professionals or a health facility organisation licensed to provide health care diagnosis and treatment services including medication, surgery and medical devices.

Yes

#### **Clinical measurements**

Is information on clinical measurements (e.g., BMI, blood pressure, height) available?

Yes

#### **Genetic data**

Are data related to genotyping, genome sequencing available?

Captured

#### Genetic data vocabulary

Other

#### Genetic data vocabulary, other

Free-text

#### **Biomarker data**

Does the data source capture biomarker information? The term "biomarker" refers to a broad subcategory of medical signs ( objective indications of medical state observed from outside the patient), which can be measured accurately and reproducibly. For example, haematological assays, infectious disease markers or metabolomic biomarkers.

#### Captured

#### **Biomarker data vocabulary**

Other

#### Biomarker vocabulary, other

Free-text

#### **Patient-reported outcomes**

Is information on patient-reported outcomes (e.g., quality of life) available?

No

#### **Patient-generated data**

Is patient-generated information (e.g., from wearable devices) available?

No

#### Units of healthcare utilisation

Are units of healthcare utilisation (e.g., number of visits to GP per year, number of hospital days) available or can they be derived? Units of healthcare utilisation refer to the quantification of the use of services for the purpose of preventing or curing health problems.

No

#### **Unique identifier for persons**

Are patients uniquely identified in the data source?

Yes

#### **Diagnostic codes**

Captured

#### Diagnosis / medical event vocabulary

ICD-10

#### Medicinal product information

Captured

#### Medicinal product information collected

Active ingredient(s) Batch number

Brand name

Dose

Route of administration

#### Medicinal product vocabulary

ATC

Other

#### If 'other,' what vocabulary is used?

UCD (french inhospital drug classification)

#### **Quality of life measurements**

Captured

#### **Quality of life measurements vocabulary**

Not coded (Free text)

#### Lifestyle factors

Captured

#### Lifestyle factors

Alcohol use

Tobacco use

#### Sociodemographic information

Captured

#### Sociodemographic information collected

Age

Gender

Other

## Quantitative descriptors

## Population Qualitative Data

#### Population age groups

Paediatric Population (< 18 years) Neonate Preterm newborn infants (0 – 27 days) Term newborn infants (0 – 27 days) Infants and toddlers (28 days – 23 months) Children (2 to < 12 years) Adolescents (12 to < 18 years) Adult and elderly population ( $\geq$ 18 years) Adults (18 to < 65 years) Adults (18 to < 46 years) Adults (46 to < 65 years) Elderly ( $\geq$  65 years) Adults (65 to < 75 years) Adults (75 to < 85 years) Adults (85 years and over)

## Estimated percentage of the population covered by the data source in the catchment area

Unknown : It's difficult to assess a healthcare facility's coverage level. However, the Bordeaux University Hospital is one of France's largest hospitals, and the largest in the Nouvelle-Aquitaine region, with a focus on medicine / surgery and obstetrics care.

\*It was not possible to differenciate preterm / term newborn

### Description of the population covered by the data source in the catchment area whose data are not collected (e.g., people who are registered only for private care)

Data from the Bordeaux University Hospital Information System are integrated into an initial EDS in i2b2 format. From this first EDS, data are integrated into a second EDS in OMOP format, including: patient data, visit data, prescription and drug administration data, diagnostic data, procedure data, biology data and free-text documents.

## Family linkage

#### Family linkage available in the data source permanently or can be created on an ad hoc basis

Ad hoc

## Population

#### **Population size**

2389909

#### Active population size

431441

## Population by age group

Age group	Population size	Active population size
Paediatric Population (< 18 years)	481757	82907
Term newborn infants (0 – 27 days)	46187	5051
Infants and toddlers (28 days – 23 months)	82742	14098
Children (2 to < 12 years)	229707	40752
Adolescents (12 to < 18 years)	123121	23006
Adults (18 to < 46 years)	878966	144048
Adults (46 to < 65 years)	488367	93344
Elderly (≥ 65 years)	540819	111142

Age group	Population size	Active population size
Adults (65 to < 75 years)	246823	55119
Adults (75 to < 85 years)	184083	40987
Adults (85 years and over)	109913	15036

## Median observation time

Median time (years) between first and last available records for unique individuals captured in the data source

0.04

Median time (years) between first and last available records for unique active individuals (alive and currently registered) capt 0.84

## Data flows and management

## Access and validation

#### **Biospecimen access**

Are biospecimens available in the data source (e.g., tissue samples)?

No

#### Access to subject details

No

#### **Description of data collection**

Data are integrated daily from the hospital information system into an CDW in i2b2 format.

Data integrated in i2b2 are loaded every 4 months into an CDW in OMOP-CDM format. 80% of data is mapped to standard terminologies.

## Event triggering registration

#### Event triggering registration of a person in the data source

Other

Event triggering registration of a person in the data source, other

Hospital visit (in and outhospital)

#### Event triggering de-registration of a person in the data source

Other

#### Event triggering de-registration of a person in the data source, other

Patient opposition to secondary use of data

#### Event triggering creation of a record in the data source

Hospital encounter (in-hospital and out-hospital)

## Data source linkage

#### Linkage

Is the data source described created by the linkage of other data sources (prelinked data source) and/or can the data source be linked to other data source on an ad-hoc basis?

Yes

#### Linkage description, possible linkage

Probabilistic matching based on identity traits (https://doi.org/10.1093/jamiaopen/ooab005)

## Linked data sources

#### Pre linked

Is the data source described created by the linkage of other data sources?

No

#### Data source, other

French death registry (open data)

#### Linkage strategy

Probabilistic

#### Linkage variable

Last names, First Name, Birth date, Gender, Birth location, Birth country, Date of death, Last visit date

#### Linkage completeness

The recall and precision of our linkage strategy is 97.5% and 99.97% 2/3 of known deaths are derived from this external database.

# Data management specifications that apply for the data source

#### Data source refresh

Quarterly

#### Informed consent for use of data for research

Other

#### Possibility of data validation

Can validity of the data in the data source be verified (e.g., access to original medical charts)?

Yes

#### **Data source preservation**

Are records preserved in the data source indefinitely?

No

#### Data source preservation length (years)

15 years

#### **Approval for publication**

Is an approval needed for publishing the results of a study using the data source?

Yes

#### Informed consent, other

General non-opposition to secondary use of health data

Data source last refresh

24/10/2024

## Common Data Model (CDM) mapping

#### **CDM** mapping

Has the data source been converted (ETL-ed) to a common data model?

Yes

#### **CDM Mappings**

#### **CDM** name

i2b2

#### **CDM** website

https://community.i2b2.org/wiki/display/BUN/i2b2+Common+Data+Model+Documentation

#### Data source ETL frequency

0,03 months

#### Data source ETL status

Completed

#### Data source ETL specifications (link)

https://gitub.u-bordeaux.fr/scossi910e/ehden-bordeaux/-/wikis/home

#### **CDM** name

OMOP

#### **CDM** website

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

#### Data source ETL CDM version

5.4

#### Data source ETL frequency

3,00 months

#### Data source ETL status

In progress

#### Data source ETL specifications (link)

https://gitub.u-bordeaux.fr/scossi910e/ehden-bordeaux-etl-omop/-/wikis/home