Clinical Practice Research Datalink (CPRD) GOLD

First published: 01/02/2024

Last updated: 28/02/2025

Data source | Human | Primary care medical records

Administrative details

Administrative details

PURI

https://redirect.ema.europa.eu/resource/1111113

Data source ID

1111113

Data source acronym

CPRD GOLD

Data holder

Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences (NDORMS), University of Oxford

Data source type

Primary care medical records

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.cprd.com/primary-care-data-public-health-research

Contact details

Antonella Delmestri

Main

antonella.delmestri@ndorms.ox.ac.uk

Data source regions and languages

Data source countries

United Kingdom

Data source languages

English

Data source establishment

Data source established

09/09/1987

Data source time span

First collection: 09/09/1987

The date when data started to be collected or extracted.

Last collection: 01/07/2024

If data collection in the data source has ceased, the date new records last entered the data source.

Publications

Data source publications

The General Practice Research Database: Now and the Future (2007)

The accuracy of date of death recording in the Clinical Practice Research Datalink GOLD database in England compared with the Office for National Statistics death registrations (2019)

Data Resource Profile: Clinical Practice Research Datalink (CPRD) (2015)

Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource (2012)

The General Practice Research Database: Role in Pharmacovigilance (2004)

Studies

List of studies that have been conducted using the data source

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

Risk of Skin Cancer and Lymphoma in Users of Topical Tacrolimus, Pimecrolimus, and Corticosteroids: Protopic JOint European Longitudinal Lymphoma and skin cancer Evaluation (JOELLE) Study

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® EHDS Use Case: Natural history of coagulopathy in COVID-19 patients and persons vaccinated against SARS-CoV-2 in the context of the OMICRON variant

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® Rates of occurrence of treatment-related intercurrent events in patients with major depressive disorder

DARWIN EU® Effectiveness of COVID-19 vaccines on severe COVID-19 and post acute outcomes of SARS-CoV-2 infection

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

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

DARWIN EU® Effectiveness of Human Papillomavirus Vaccines (HPV) to prevent cervical cancer

DARWIN EU® Comparing direct and indirect methods to estimate prevalence of chronic diseases using real-world data

DARWIN EU® – Frailty and polypharmacy among adults with selected cancers at the time of diagnosis

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

DARWIN EU® - Characterising interstitial lung disease in Europe

SAFETY-VAC: Network of Data Sources for Vaccine Safety Evaluation

ADEPT: The utilisation of antiseizure medications in pregnant women, other women of childbearing potential, and men: a multi-database study from 7 European countries

DARWIN EU® – Trends in utilisation of Attention-Deficit Hyperactivity Disorder (ADHD) Medications

DARWIN EU® - Drug Utilisation Study on GLP-1 Receptor Agonists

SAFETY-VAC: Background incidence estimation of flares of pre-existing chronic diseases using pan-European electronic healthcare data sources. (SAFETY VAC)

ADEPT: feasibility of estimating the risk of adverse pregnancy, neonatal and child outcomes following either in utero ASM exposure through the mother, or peri-conceptional ASM exposure through the father

DARWIN EU® - Background incidence rates of selected vaccine adverse events of special interest (AESIs) in Europe

DARWIN EU® - Suicidality following exposure to doxycycline

SAFETY-VAC: Phenotype proposal and rates of immunocompromised populations in real-world data sources.

DARWIN EU® - Incidence of suicidality in patients with specific chronic skin conditions

DARWIN EU® - Azathioprine - user characteristics

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

DARWIN EU® - DUS Characterising STOPP criteria medication use in people with recurrent falls

DARWIN EU® - Incidence of myoclonus in heart failure: a descriptive analysis in patients treated with sacubitril/valsartan and other treatments

A post-authorisation safety study of ABRYSVO in immunocompromised, or renally or hepatically impaired adults aged 60 years and older in a real world setting in Europe and UK

DARWIN EU® -Drug utilisation of salbutamol products for inhalation and therapeutic alternative inhalation products

DARWIN EU® - Suicidality incidence rates in adult male patients and in patients treated with finasteride and dutasteride

DARWIN EU® - Characterisation of exposure to acitretin and purpura and related conditions

DARWIN EU® - Prevalence of hypertrophic cardiomyopathy (HCM) and obstructive hypertrophic cardiomyopathy (oHCM) in six European countries

DARWIN EU® - Prescription trends of ketamine and esketamine

DARWIN EU® – Incidence rates of venous thromboembolic events in patients with selected cancers

DARWIN EU® - Association of venous thromboembolism with non-steroidal anti-inflammatory drug use in women 15-49 years using hormonal contraceptives

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

Not Captured

Prescriptions of medicines

Captured

Dispensing of medicines

Not Captured

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

Read

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

Read

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?

Not Captured

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.

Not Captured

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

Read

Medicinal product information

Captured

Medicinal product information collected

Active ingredient(s)

Brand name

Formulation

Route of administration

Strength

Medicinal product vocabulary

dm+d

Gemscript

Quality of life measurements

Not Captured

Lifestyle factors

Captured

Lifestyle factors

Alcohol use

Diet

Tobacco use

Sociodemographic information

Captured

Sociodemographic information collected

Age

Ethnicity

Gender

Marital status

Quantitative descriptors

Population Qualitative Data

Population age groups

Paediatric Population (< 18 years)

Adolescents (12 to < 18 years)

Adults (18 to < 46 years)

Adults (46 to < 65 years)

Elderly (≥ 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

In CPRD GOLD July 2024 there were 2,894,922 current acceptable patients (i.e. registered at currently contributing practices that use Vision software, excluding transferred out, deceased patients and those flagged by CPRD as not acceptable for clinical research for data quality issues) equal to 4.32% based on the UK population estimates of 67,026,300 from the Office of National Statistics (July 2024). Our Oxford OMOP CDM data include 2,894,714 after excluding 208 patients as described for the full population below: the percentage remains the same.

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)

CPRD GOLD July 2024 included 24,380,228 patients of which 2,896,079 were flagged as not acceptable for clinical research. Our CDM data include 17,521,504 by excluding those and also patients with gender not equal to Male or Female (1,143), year of birth before 1875 (785), practice up-to-standard after the patient died (3) or left the practice or after the last collection date or patients with no observations (227,221). The numbers below refer to this CDM population.

Population

Population size

17521504

Active population size

2894714

Population by age group

Age group	Population size	Active population size
Paediatric Population (< 18 years)	3147229	519902
Children (2 to < 12 years)	1856658	287819
Adolescents (12 to < 18 years)	963131	200949
Adults (18 to < 46 years)	7516645	1061418
Adults (46 to < 65 years)	3377115	725924
Elderly (≥ 65 years)	3480514	587470
Adults (65 to < 75 years)	1377791	303212
Adults (75 to < 85 years)	1187660	205960
Adults (85 years and over)	915064	78298

Median observation time

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

5.87

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

13.36

Data flows and management

Access and validation

Governance details

Documents or webpages that describe the overall governance of the data source and processes and procedures for data capture and management, data quality check and validation results (governing data access or utilisation for research purposes).

https://cprd.com/

Biospecimen access

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

No

Access to subject details

Can individual patients/practitioners/practices included in the data source be contacted?

No

Description of data collection

https://cprd.com/primary-care-data-public-health-research

https://ohdsi.github.io/ETL-LambdaBuilder/docs/CPRD

Event triggering registration

Event triggering registration of a person in the data source

Practice registration

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

Death

Practice deregistration

Event triggering creation of a record in the data source

Patient has contact with a GP practice

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, pre-linked

CPRD GOLD can be linked to several sources, however our Oxford OMOP CDM data are not linked to other data sources

---

Linkage description, possible linkage

https://cprd.com/cprd-linked-data#Data%20from%20NHS%20Digital

Linked data sources

Pre linked

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

No

Data source, other

Hospital Episode Statistics (HES): HES Accepted Patient Care (APC), HES Accident & Emergency (A&E), HES Outpatient (OP). Office of National Statistics (ONS) mortality records. Index of Multiple Deprivation (IMD), Townsend index, Carstairs index, Rural Urban classification

Linkage variable

Patient identifier

Data management specifications that apply for the data source

Data source refresh

Every 6 months

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)?

No

Data source preservation

Are records preserved in the data source indefinitely?

Yes

Approval for publication

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

Yes

Informed consent, other

Patient are assumed to consent, but they can withdraw consent at any time

Data source last refresh

01/07/2024

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

OMOP

CDM website

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

Data source ETL CDM version

5.4.1

Data source ETL frequency

6,00 months

Data source ETL status

Completed

Data source ETL specifications (link)

https://ohdsi.github.io/ETL-LambdaBuilder/docs/CPRD