

# UK Biobank

**First published:** 01/02/2024

**Last updated:** 05/05/2026

Data source

Human

Biobank

Death registry

Other

## Administrative details

### Administrative details

**Data source ID**

1111233

---

**Data source acronym**

UKB

---

**Data holder**

[Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences \(NDORMS\), University of Oxford](#)

---

**Data source type**

Biobank

Death registry

Other

---

## Data source type, other

Electronic health records

---

## Main financial support

Funding from public-private partnership  
National, regional, or municipal public funding

---

## Care setting

Hospital inpatient care  
Primary care - GP, community pharmacist level

---

## 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.ukbiobank.ac.uk/>

## Contact details

Antonella Delmestri [antonella.delmestri@ndorms.ox.ac.uk](mailto:antonella.delmestri@ndorms.ox.ac.uk)

Main

[antonella.delmestri@ndorms.ox.ac.uk](mailto:antonella.delmestri@ndorms.ox.ac.uk)

## Data source regions and languages

### Data source countries

United Kingdom

---

## Data source languages

English

---

## Data source regions

England

Scotland

Wales [Cymru GB-CYM]

## Data source establishment

### Data source established

15/06/2007

---

### Data source time span

**First collection:** 01/01/1996

The date when data started to be collected or extracted.

## Publications

### Data source publications

[Cohort profile of the UK Biobank: diagnosis and characteristics of cerebrovascular disease](#)

[The UK Biobank resource with deep phenotyping and genomic data](#)

[The sequences of 150,119 genomes in the UK Biobank](#)

[Plasma proteomic associations with genetics and health in the UK Biobank](#)

[Transforming and evaluating the UK Biobank to the OMOP Common Data Model for COVID-19 research and beyond](#)

## Studies

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

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

DARWIN EU® - Association between genetic polymorphisms of interest and risk of myopathy among statin users

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

DARWIN EU® - Time to onset of thromboembolic events in adults with selected types of cancer

## 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.

Yes

---

#### **Disease details (other)**

Cancer

---

#### **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)?

No

---

### **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**

ICD-10

---

### **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**

OPCS

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?

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.

Captured

---

### **Patient-reported outcomes**

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

Yes

---

### **Patient-generated data**

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

Yes

---

### **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

Read

---

**Medicinal product information**

Captured

---

**Medicinal product information collected**

Active ingredient(s)

Brand name

Formulation

Route of administration

Strength

---

**Medicinal product vocabulary**

dm+d

Other

---

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

BNF

---

**Quality of life measurements**

Not Captured

---

## **Lifestyle factors**

Captured

---

## **Lifestyle factors**

Alcohol use

Diet

Frequency of exercise

Other

Tobacco use

---

## **Sociodemographic information**

Captured

---

## **Sociodemographic information collected**

Age

Deprivation index

Education level

Ethnicity

Gender

Marital status

Socioeconomic status

## Quantitative descriptors

## Population Qualitative Data

### **Population age groups**

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**

We are reporting numbers of the OMOP Biobank data.

In April 2024 there were 457,441 current active participants (i.e. excluding died or withdrawn from the cohort) equal to  $\sim$ 2% of target population of adults aged older than 45 years based on the estimates from the Office of National Statistics.

---

### **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)**

Secondary care, mortality and cancer data are available for all UK Biobank participants but primary care data currently only linkable for half of them, and only covering up to 2017.

## Population

### **Population size**

501939

---

### **Active population size**

457441

## Population by age group

Age group	Population size	Active population size
Adults (18 to < 46 years)	110	0
Adults (46 to < 65 years)	110300	102154
Elderly ( $\geq$ 65 years)	391529	355287
Adults (65 to < 75 years)	172656	152758
Adults (75 to < 85 years)	205285	188968
Adults (85 years and over)	13588	13561

## Median observation time

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

25.42

---

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

25.62

## 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://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/governance>

## **Biospecimen access**

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

Yes

---

## **Biospecimen access conditions**

Request to be made in the research application for approval.

---

## **Access to subject details**

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

No

---

## **Description of data collection**

<https://www.ukbiobank.ac.uk/enable-your-research/about-our-data>

# 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**

Patient was recruited into BioBank cohort and gave written consent

---

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

Death

Loss to follow up

Practice deregistration

---

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

Patient had contact with a GP practice and/or had a hospital admission and/or gave self reported information

# 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**

[https://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Data\\_providers\\_and\\_dates](https://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Data_providers_and_dates)

# Linked data sources

## **Pre linked**

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

Yes

---

## **Data source, other**

Death Data (England, Wales, Scotland)

---

**Linkage variable**

eid

---

**Linkage completeness**

[https://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Data\\_providers\\_and\\_dates](https://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Data_providers_and_dates)

**Pre linked**

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

Yes

---

**Data source, other**

Hospital Inpatient Data (England, Wales, Scotland)

---

**Linkage variable**

eid

---

**Linkage completeness**

[https://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Data\\_providers\\_and\\_dates](https://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Data_providers_and_dates)

**Pre linked**

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

Yes

---

**Data source, other**

Primary Care (England, Wales, Scotland)

---

**Linkage variable**

eid

---

**Linkage completeness**

[https://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Data\\_providers\\_and\\_dates](https://biobank.ctsu.ox.ac.uk/crystal/exinfo.cgi?src=Data_providers_and_dates)

## Data management specifications that apply for the data source

**Data source refresh**

Yearly

---

**Informed consent for use of data for research**

Waiver

---

**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

---

**Data source last refresh**

01/04/2024

## 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

---

### **Data source ETL frequency**

12,00 months

---

### **Data source ETL status**

Completed

---

### **Data source ETL specifications (link)**

[https://github.com/oxford-pharmacoepi/etl\\_ndorms](https://github.com/oxford-pharmacoepi/etl_ndorms)