Estonian Biobank

First published: 01/02/2024

Last updated: 27/01/2025

Data source | Human | Biobank

Administrative details

Administrative details

PURI

https://redirect.ema.europa.eu/resource/1111114

Data source ID

1111114

Data source acronym

EBB

Data holder

University of Tartu

Data source type

Biobank

Main financial support

Funding from industry or contract research

National, regional, or municipal public funding

Care setting

Hospital inpatient care

Hospital outpatient care

Primary care – GP, community pharmacist level

Primary care – specialist level (e.g. paediatricians)

Secondary care – specialist level (ambulatory)

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://genomics.ut.ee/en/content/estonian-biobank

Contact details

Raivo Kolde

Main

raivo.kolde@ut.ee

Data source regions and languages

Data source countries

Estonia

Data source languages

Estonian

Data source establishment

Data source established

01/06/2002

Data source time span

First collection: 01/01/2004

The date when data started to be collected or extracted.

Last collection: 31/12/2024

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

Publications

Data source publications

Reisberg S, Krebs K, Lepamets M, Kals M, Mägi R, Metsalu K, Lauschke VM, Vilo J, Milani L. Translating genotype data of 44,000 biobank participants into clinical pharmacogenetic recommendations: challenges and solutions. Genetics in Medicine. 2019 Jun;21(6):1345-54.

Singh T, Poterba T, Curtis D, Akil H, Al Eissa M, Barchas JD, Bass N, Bigdeli TB, Breen G, Bromet EJ, Buckley PF. Rare coding variants in ten genes confer substantial risk for schizophrenia. Nature. 2022 Apr 21;604(7906):509-16.

Writing group lead Andrews Shea J. 6 Kanai Masahiro 3 Cordioli Mattia 7, Manuscript analyses team lead, Manuscript analyses team members: PHEWAS Polimanti Renato 1, Manuscript analyses team members: Mendelian randomization Harerimana Nadia 8, Manuscript analyses team members: methods development Pirinen Matti 7, Manuscript analyses team members: PC

projection, gene prioritization, Project management lead Liao Rachel G. 3, Project managment support Chwialkowska Karolina 9 Trankiem Amy 3 Balaconis Mary K. 3, Nguyen H, Solomonson M. A first update on mapping the human genetic architecture of COVID-19. Nature. 2022 Aug 4;608(7921):E1-0.

Leitsalu L, Haller T, Esko T, Tammesoo ML, Alavere H, Snieder H, Perola M, Ng PC, Mägi R, Milani L, Fischer K. Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. International journal of epidemiology. 2015 Aug 1;44(4):1137-47.

Studies

List of studies that have been conducted using the data source

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

Strengthening Use of Real-World Data in Medicines Development: Metadata for Data Discoverability and Study Replicability (MINERVA)

DARWIN EU® - Background rates of serious adverse events to contextualise safety assessments in clinical trials and non-interventional studies in adolescent and adult patients with severe asthma

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

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® Drug Utilisation Study of prescription opioids

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

ICD-10

Prescriptions of medicines

Captured

Prescriptions vocabulary

ATC

other

Prescriptions vocabulary, other

Local package code

Dispensing of medicines

Captured

Dispensing vocabulary

ATC

other

Dispensing vocabulary, other

Local package code

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

ICD-10

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

SNOMED CT

Procedures vocabulary, other

Local service code, NOMESCO classification of surgical procedures, ICD-10

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

HGNC

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

Local code

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

Not coded (Free text)

SNOMED CT

Medicinal product information

Captured

Medicinal product information collected

Active ingredient(s)

Brand name

Dosage regime

Dose

Package size

Route of administration

Strength

Medicinal product vocabulary

ATC

Other

If 'other,' what vocabulary is used?

Local package code

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

Education level

Ethnicity

Gender

Living in rural area

Marital status

Quantitative descriptors

Population Qualitative Data

Population age groups

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

20% of adult people in Estonia.

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)

Joining the Estonian Biobank is voluntary. By joining, consent is given for data collection from various data sources. To join, one must be at least 18 years old. Data is collected about these individuals. Data is not collected from those individuals who have not expressed a desire to join the Estonian Biobank

Family linkage

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

Ad hoc

Population

Population size

212004

Active population size

201282

Population by age group

Age group	Population size	Active population size
Adults (18 to < 46 years)	86697	86198
Adults (46 to < 65 years)	74847	72742
Elderly (≥ 65 years)	50460	42342
Adults (65 to < 75 years)	28774	26368
Adults (75 to < 85 years)	15995	12782
Adults (85 years and over)	5691	3192

Median observation time

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

20.00

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

20.00

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://genomics.ut.ee/en/content/estonian-biobank

Biospecimen access

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

Yes

Biospecimen access conditions

A four step process. 1) Preliminary inquiry; 2) Application to Estonian Biobank Scientific committee; 3) Application to the Estonian Committee on Bioethics and Human Research 4) Application for the issuance of data or biosamples or both from the EBB

Access to subject details

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

Yes

Description of data collection

The donors fill out a survey upon joiing the biobank. Subsequently their medical data dating back to 2004 is loaded from national sources (claims, discharge reports, prescriptions and registries). The healthcare data is updated each year.

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

Application to join Estonian Biobank

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

Other

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

Application to leave Estonian Biobank

Event triggering creation of a record in the data source

Initially joining the biobank. Subsequently the health data is being kept up to date.

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

Linkage is done using unique personal identification code which is unique to all Estonian residents

Linked data sources

Pre linked

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

No

Data source, other

All the data sources that can be linked to are already included.

Pre linked

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

Yes

Data source, other

Cancer registry information

Linkage variable

Unique personal identification code

Linkage completeness

Every inhabitant of Estonia has the personal identification code that can link all their data from different data sources.

Pre linked

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

Yes

Data source, other

Death registry information

Linkage variable

Unique personal identification code

Linkage completeness

Every inhabitant of Estonia has the personal identification code that can link all their data from different data sources.

Pre linked

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

Yes

Data source, other

Digital prescription data

Linkage variable

Unique personal identification code

Linkage completeness

Every inhabitant of Estonia has the personal identification code that can link all their data from different data sources.

Pre linked

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

Yes

Data source, other

Discharge summaries

Linkage variable

Unique personal identification code

Linkage completeness

Every inhabitant of Estonia has the personal identification code that can link all their data from different data sources.

Pre linked

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

Yes

Data source, other

Insurance claims data

Linkage variable

Unique personal identification code

Linkage completeness

Every inhabitant of Estonia has the personal identification code that can link all their data from different data sources.

Data management specifications that apply for the data source

Data source refresh

Yearly

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

All donors have signed a broad informed consent upon joining the Biobank.

Data source last refresh

15/06/2024

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