Genomics England (GEL)

First published: 26/10/2022

Last updated: 17/10/2024

Data source **Human** **Biobank** **Population registry** **Other**

Administrative details

Administrative details

Data source ID

1111185

Data source acronym

GEL

Data holder

Genomics England

Data source type

Biobank

Population registry

Other

Data source type, other

Genomics registy, molecular data registry

Main financial support

National, regional, or municipal public funding

Care setting

Secondary care – specialist level (ambulatory)

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.genomicsengland.co.uk/

Contact details

General Email commercialpartnerships@genomicsengland.co.uk

Main

commercialpartnerships@genomicsengland.co.uk

Data source regions and languages

Data source countries

United Kingdom

Data source languages

English

Data source regions

England

Data source establishment

Data source established

15/06/2016

Data source time span

First collection: 15/06/2016

The date when data started to be collected or extracted.

Publications

Data source publications

https://www.genomicsengland.co.uk/research/publications

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

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

OPCS

Prescriptions of medicines

Captured

Prescriptions vocabulary

other

Dispensing of medicines

Captured

Dispensing vocabulary

other

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)].

Yes

Contraception

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

No

Indication for use

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

Captured

Medical devices

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

No

Administration of vaccines

No

Procedures

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

Captured

Procedures vocabulary

OPCS

SNOMED

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.

No

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

HGVS

Other

Genetic data vocabulary, other

ClinVar for information about genomic variation and its relationship to human health

---

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?

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.

Yes

Unique identifier for persons

Are patients uniquely identified in the data source?

Yes

Diagnostic codes

Captured

Diagnosis / medical event vocabulary

ICD-10

OPCS

Other

Diagnosis / medical event vocabulary, other

ICD-0-3

Medicinal product information

Not Captured

Quality of life measurements

Not Captured

Lifestyle factors

Not Captured

Sociodemographic information

Captured

Sociodemographic information collected

Age

Sex

Gender

Ethnicity

Deprivation index

Quantitative descriptors

Population Qualitative Data

Population age groups

Paediatric Population (< 18 years)

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)

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)

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)

Cohort of 100,000. All received whole genome sequencing and is not necessarily representative of UK population.

Population

Population size

110349

Population by age group

Age group	Population size

Children (2 to < 12 years)	9351
Adolescents (12 to < 18 years)	4786
Adults (18 to < 46 years)	36940
Adults (46 to < 65 years)	36778
Adults (65 to < 75 years)	13229
Adults (75 to < 85 years)	6272
Adults (85 years and over)	1273

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://research-help.genomicsengland.co.uk/display/GERE/9.+Data+Security+and+You

Biospecimen access

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

Yes

Biospecimen access conditions

Informed consent forms are signed by participants.

https://files.genomicsengland.co.uk/documents/Patient-Information-Research-V1.4.pdf

Access to subject details

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

No

Description of data collection

Clinical and phenotypic data: Sourced from Electronic Medical Records from NHS

Genomic data: Sequencing

Event triggering registration

Event triggering registration of a person in the data source

Disease diagnosis

Practice registration

Start of treatment

Other

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

Genomic testing

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

Other

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

Exit from resource is only possible upon change in consent status

Event triggering creation of a record in the data source

Diagnosis of disease, Hospital discharge,recording of congential or genetic abnormality, Hospital stay, Hospital procedure, Genetic sequencing,

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

Genomics Data generated upon patient enrolment is linked to provide additional clinical information for the Data source. For cancer Data, linkage to CAS (Cancer Analysis System). NCRAS and SACT are accessed within CAS. For clinical secondary care Data, linked to HES (Hospital Episode Statistics)

Linked data sources

Pre linked

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

Yes

Data source, other

Cancer Analysis System (CAS)

Linkage strategy

Deterministic

Linkage variable

“participant_id” this is the main linkage between the participants genomic Data and Other associated Data, we do also use platekey’s but they have slightly different formats “plate-key” “platekey” “germline_sample_platekey”

Linkage completeness

NCRAS:94%; SACT:44%

Pre linked

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

Yes

Data source, other

HES (Hospital Episode Statistics)

Linkage strategy

Deterministic

Linkage variable

“participant_id” this is the main linkage between the participants genomic Data and Other associated Data, we do also use platekey’s but they have slightly different formats “plate-key” “platekey” “germline_sample_platekey”

Linkage completeness

98%

Data management specifications that apply for the data source

Data source refresh

October

January

April

July

Informed consent for use of data for research

Required for all studies

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?

Yes

Approval for publication

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

Yes

Data source last refresh

13/10/2022

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 status

In progress