THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

CREATE CHANGE

Health-Related Outcomes & PheWAS in the UKB

Tuesday 6th February 2024

Isabelle McGrath

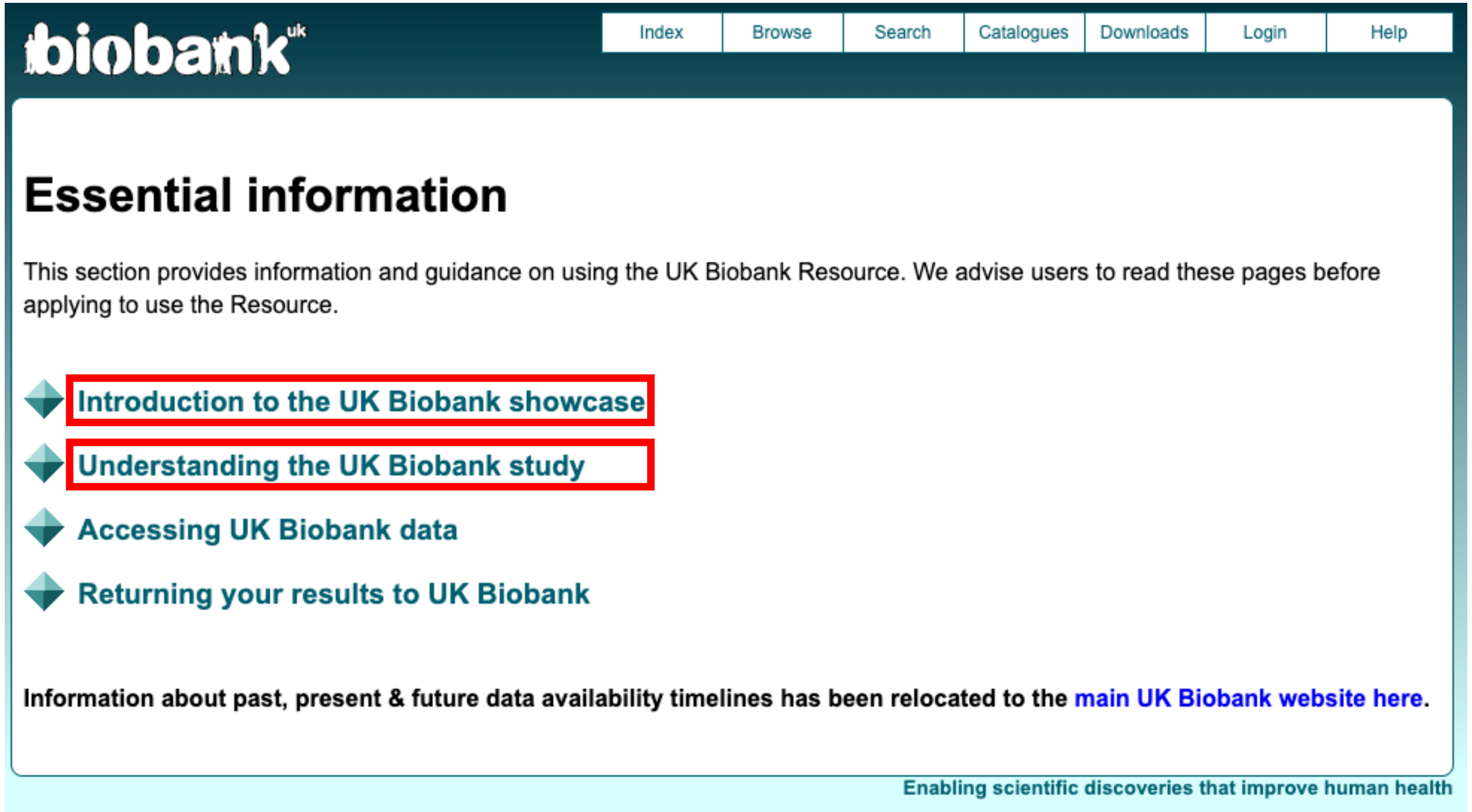
Institute for Molecular Bioscience @UQ

biobank uk

Enabling scientific discoveries that improve human health

UKB Manuals

UKB Showcase > Index > Essential Information



Essential information

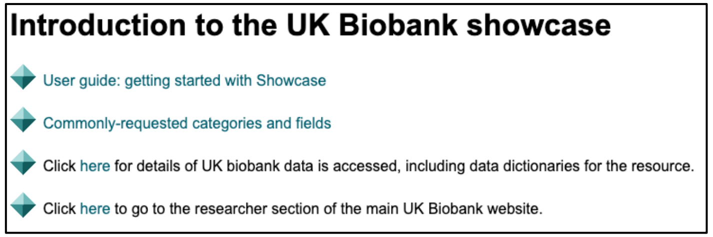
This section provides information and guidance on using the UK Biobank Resource. We advise users to read these pages before applying to use the Resource.

- Introduction to the UK Biobank showcase
- Understanding the UK Biobank study
- Accessing UK Biobank data
- Returning your results to UK Biobank

Information about past, present & future data availability timelines has been relocated to the main UK Biobank website here.

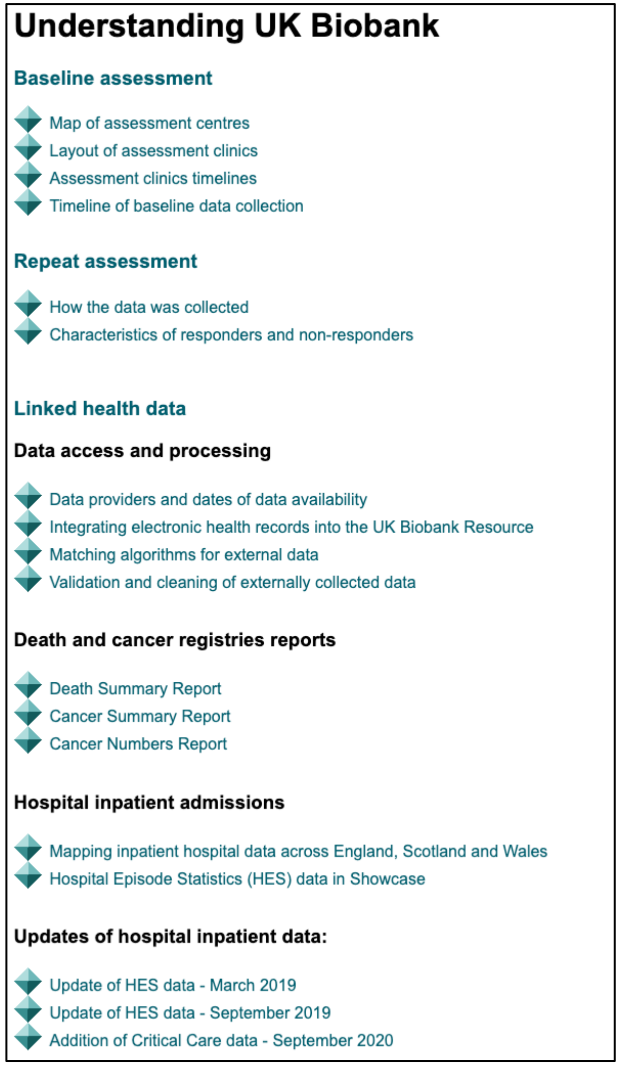
Enabling scientific discoveries that improve human health

https://biobank.ndph.ox.ac.uk/showcase/exinfo.cgi



Introduction to the UK Biobank showcase

- User guide: getting started with Showcase
- Commonly-requested categories and fields
- Click here for details of UK biobank data is accessed, including data dictionaries for the resource.
- Click here to go to the researcher section of the main UK Biobank website.



Understanding UK Biobank

Baseline assessment

- Map of assessment centres
- Layout of assessment clinics
- Assessment clinics timelines
- Timeline of baseline data collection

Repeat assessment

- How the data was collected
- Characteristics of responders and non-responders

Linked health data

Data access and processing

- Data providers and dates of data availability
- Integrating electronic health records into the UK Biobank Resource
- Matching algorithms for external data
- Validation and cleaning of externally collected data

Death and cancer registries reports

- Death Summary Report
- Cancer Summary Report
- Cancer Numbers Report

Hospital inpatient admissions

- Mapping inpatient hospital data across England, Scotland and Wales
- Hospital Episode Statistics (HES) data in Showcase

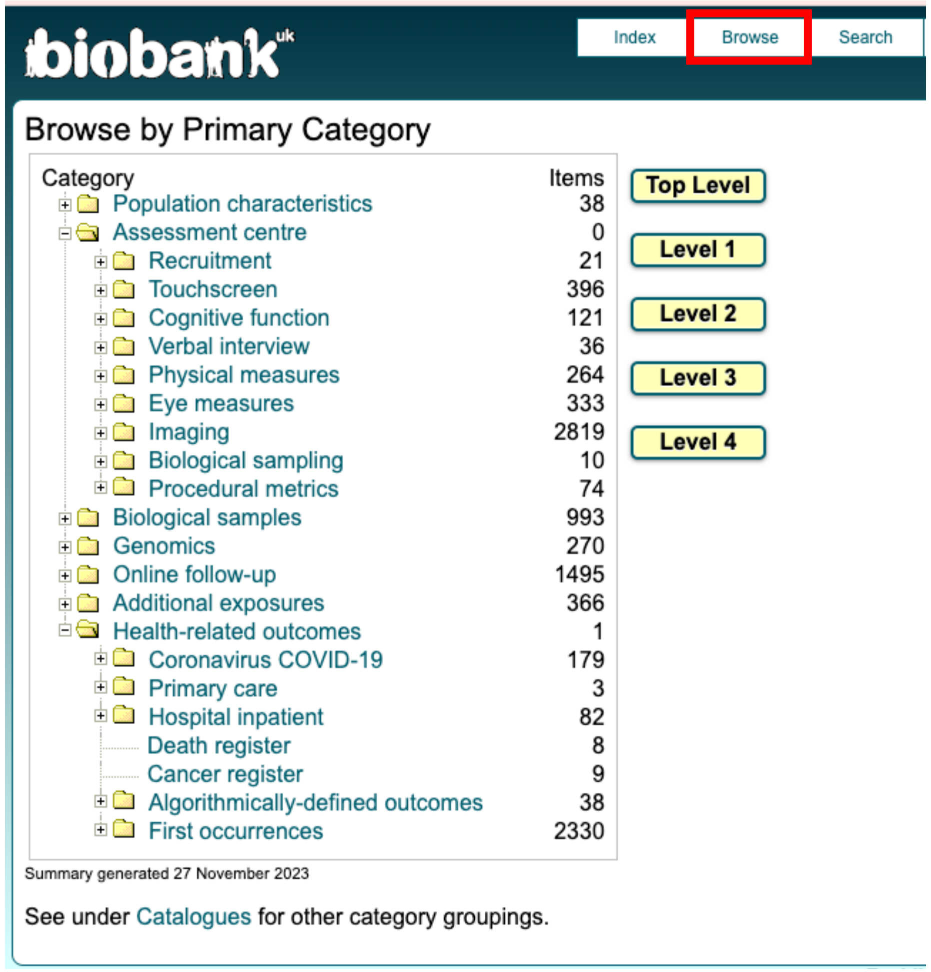
Updates of hospital inpatient data:

- Update of HES data - March 2019
- Update of HES data - September 2019
- Addition of Critical Care data - September 2020

Health-Related Outcome Categories

- Primary Care
- Hospital Inpatient
- Cancer Register
- Death Register
- Assessment centre interviews
- First Occurrences
- Algorithmically-defined outcomes
- Coronavirus COVID-19

How to browse data

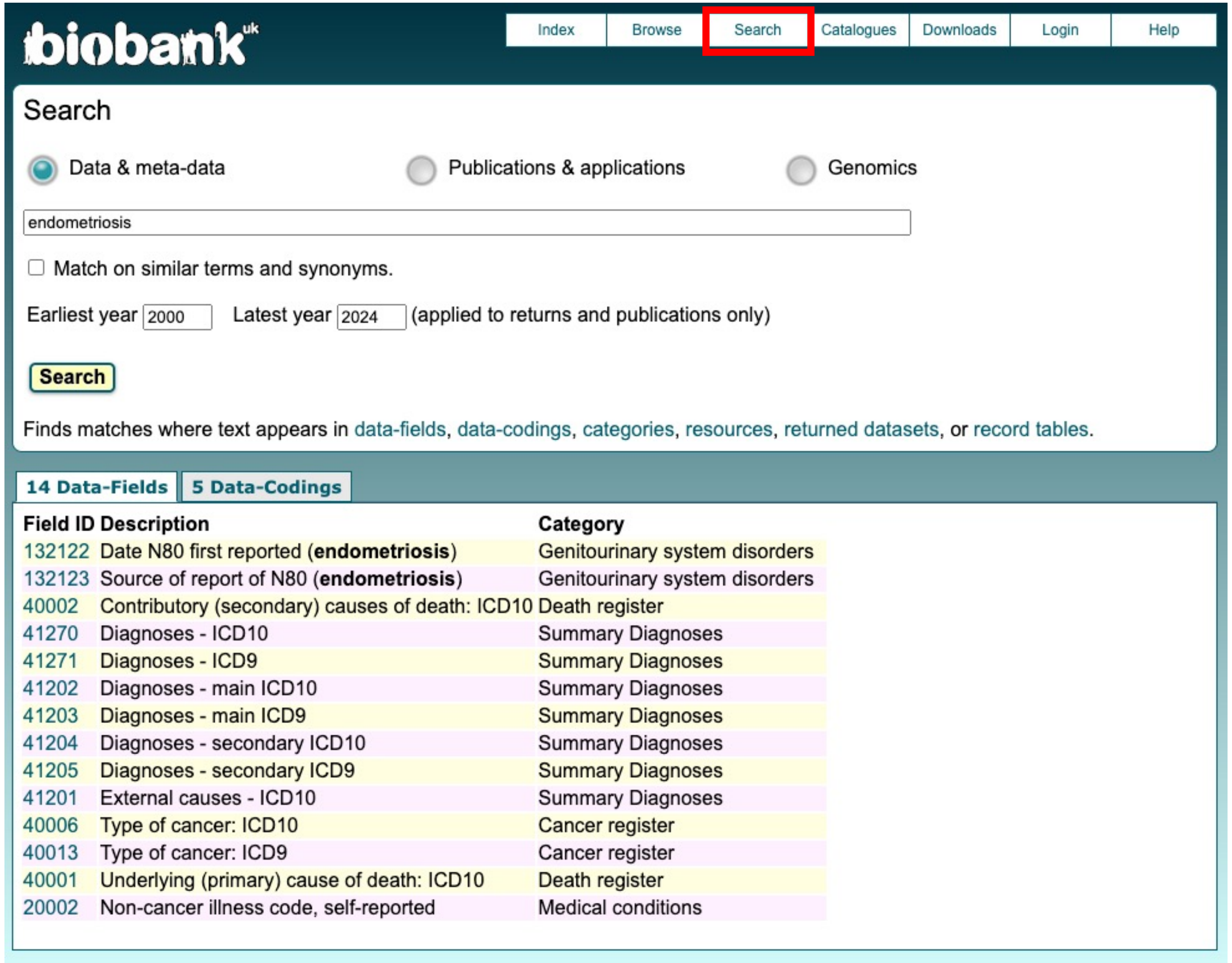


Browse by Primary Category

Category	Items
Population characteristics	38
Assessment centre	0
Recruitment	21
Touchscreen	396
Cognitive function	121
Verbal interview	36
Physical measures	264
Eye measures	333
Imaging	2819
Biological sampling	10
Procedural metrics	74
Biological samples	993
Genomics	270
Online follow-up	1495
Additional exposures	366
Health-related outcomes	1
Coronavirus COVID-19	179
Primary care	3
Hospital inpatient	82
Death register	8
Cancer register	9
Algorithmically-defined outcomes	38
First occurrences	2330

Summary generated 27 November 2023

See under Catalogues for other category groupings.



Search

● Data & meta-data ○ Publications & applications ○ Genomics

endometriosis

□ Match on similar terms and synonyms.

Earliest year 2000 Latest year 2024 (applied to returns and publications only)

Search

Finds matches where text appears in data-fields, data-codings, categories, resources, returned datasets, or record tables.

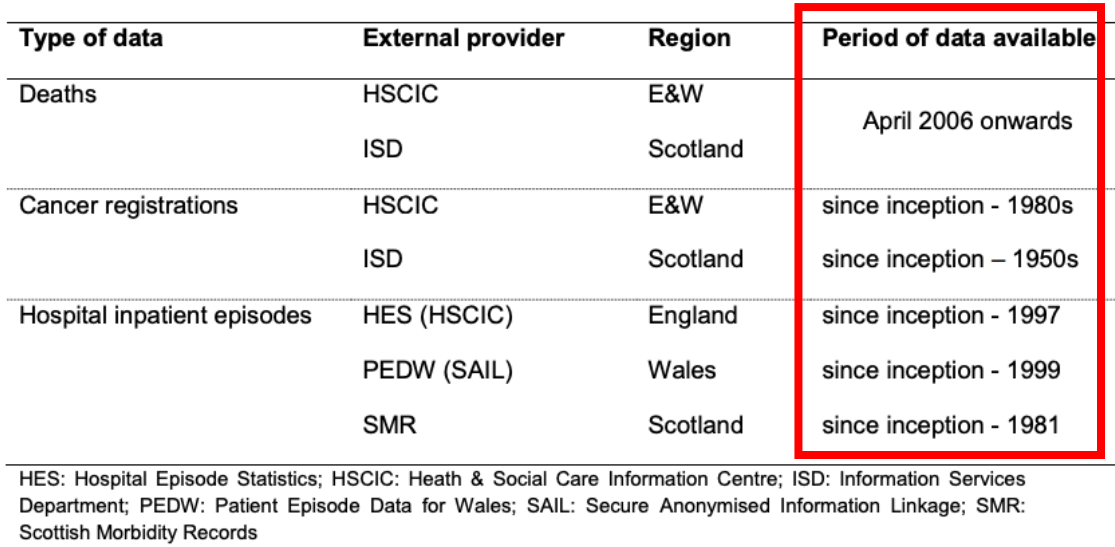
14 Data-Fields | 5 Data-Codings

Field ID	Description	Category
132122	Date N80 first reported (**endometriosis**)	Genitourinary system disorders
132123	Source of report of N80 (**endometriosis**)	Genitourinary system disorders
40002	Contributory (secondary) causes of death: ICD10	Death register
41270	Diagnoses - ICD10	Summary Diagnoses
41271	Diagnoses - ICD9	Summary Diagnoses
41202	Diagnoses - main ICD10	Summary Diagnoses
41203	Diagnoses - main ICD9	Summary Diagnoses
41204	Diagnoses - secondary ICD10	Summary Diagnoses
41205	Diagnoses - secondary ICD9	Summary Diagnoses
41201	External causes - ICD10	Summary Diagnoses
40006	Type of cancer: ICD10	Cancer register
40013	Type of cancer: ICD9	Cancer register
40001	Underlying (primary) cause of death: ICD10	Death register
20002	Non-cancer illness code, self-reported	Medical conditions

Linked Health Data

- Linkage with electronic health record databases from England, Wales and Scotland
- Files sent to UKB quarterly for death and cancer, annually for hospital activity data

- Primary care data
  - available for ~230,000 UKB Participants up to 2016 or 2017
  - GP system
  - Consultations, diagnoses, procedures & laboratory tests
  - Read v2 and Read CTV3 codes (since 1985)
- Hospital inpatient data
  - Available for full cohort
  - Hospital admission data: date, diagnosis & procedures
  - ICD9 & ICD10 codes
- Cancer data
  - ICD9 & ICD10 codes & cancer histology code
- Death data
  - Date & cause of death
  - ICD10 codes



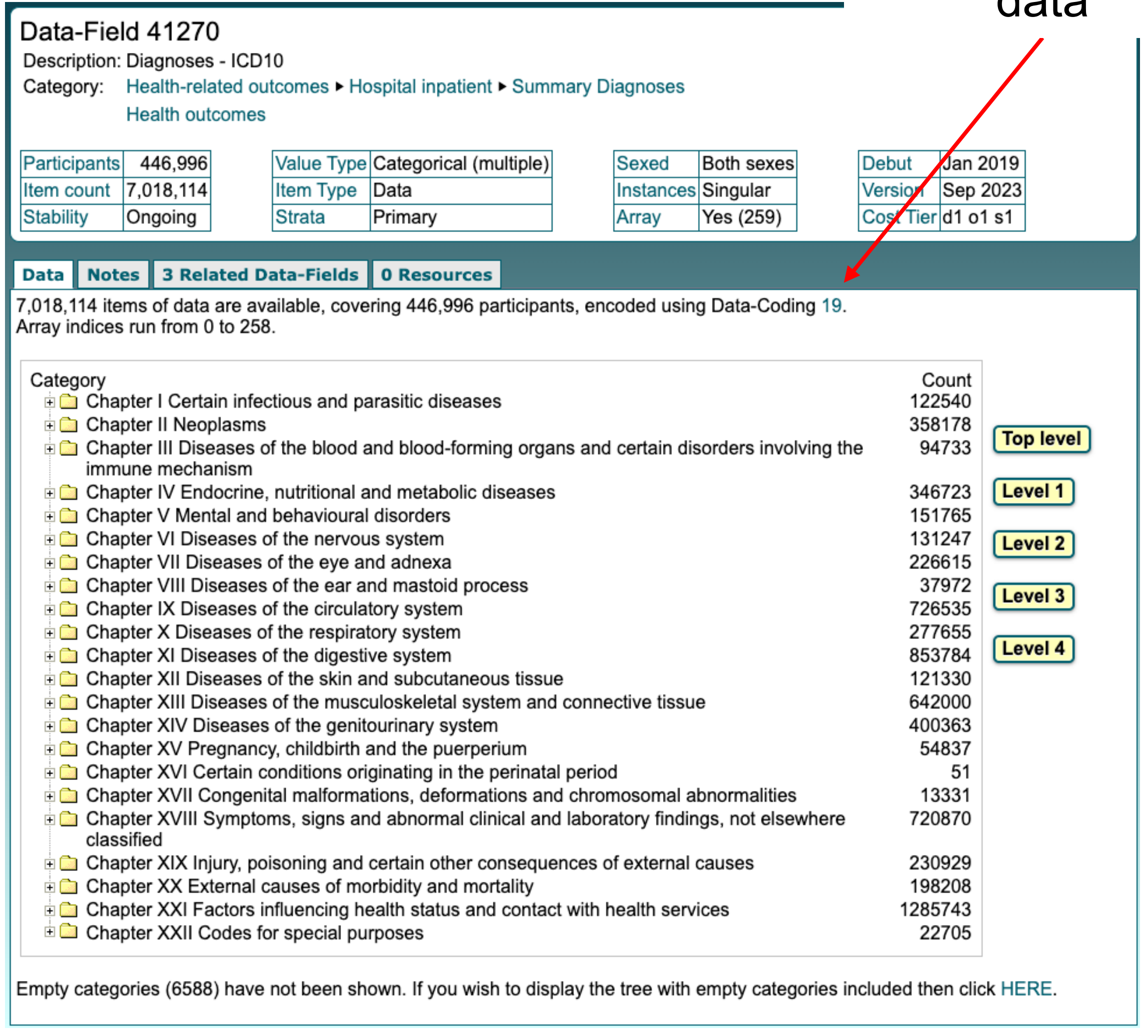
Type of data	External provider	Region	Period of data available
Deaths	HSCIC	E&W	April 2006 onwards
	ISD	Scotland	
Cancer registrations	HSCIC	E&W	since inception - 1980s
	ISD	Scotland	since inception – 1950s
Hospital inpatient episodes	HES (HSCIC)	England	since inception - 1997
	PEDW (SAIL)	Wales	since inception - 1999
	SMR	Scotland	since inception - 1981

HES: Hospital Episode Statistics; HSCIC: Heath & Social Care Information Centre; ISD: Information Services Department; PEDW: Patient Episode Data for Wales; SAIL: Secure Anonymised Information Linkage; SMR: Scottish Morbidity Records

Hospital inpatient data: ICD10 codes

- Diagnostic codes used for billing
- Grouped into 22 chapters
- Great for hypothesis-free approaches
- 4-digit codes

use data coding file to decrypt data



Data-Field 41270

Description: Diagnoses - ICD10

Category: Health-related outcomes ▸ Hospital inpatient ▸ Summary Diagnoses
Health outcomes

Participants	446,996
Item count	7,018,114
Stability	Ongoing

Value Type	Categorical (multiple)
Item Type	Data
Strata	Primary

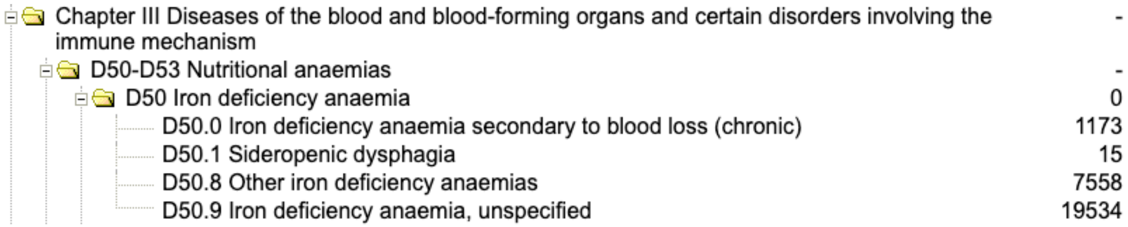
Sexed	Both sexes
Instances	Singular
Array	Yes (259)

Debut	Jan 2019
Version	Sep 2023
Cost Tier	d1 o1 s1

7,018,114 items of data are available, covering 446,996 participants, encoded using Data-Coding 19.
Array indices run from 0 to 258.

Category	Count
Chapter I Certain infectious and parasitic diseases	122540
Chapter II Neoplasms	358178
Chapter III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	94733
Chapter IV Endocrine, nutritional and metabolic diseases	346723
Chapter V Mental and behavioural disorders	151765
Chapter VI Diseases of the nervous system	131247
Chapter VII Diseases of the eye and adnexa	226615
Chapter VIII Diseases of the ear and mastoid process	37972
Chapter IX Diseases of the circulatory system	726535
Chapter X Diseases of the respiratory system	277655
Chapter XI Diseases of the digestive system	853784
Chapter XII Diseases of the skin and subcutaneous tissue	121330
Chapter XIII Diseases of the musculoskeletal system and connective tissue	642000
Chapter XIV Diseases of the genitourinary system	400363
Chapter XV Pregnancy, childbirth and the puerperium	54837
Chapter XVI Certain conditions originating in the perinatal period	51
Chapter XVII Congenital malformations, deformations and chromosomal abnormalities	13331
Chapter XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	720870
Chapter XIX Injury, poisoning and certain other consequences of external causes	230929
Chapter XX External causes of morbidity and mortality	198208
Chapter XXI Factors influencing health status and contact with health services	1285743
Chapter XXII Codes for special purposes	22705

Empty categories (6588) have not been shown. If you wish to display the tree with empty categories included then click HERE.



Chapter III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	-
D50-D53 Nutritional anaemias	-
D50 Iron deficiency anaemia	0
D50.0 Iron deficiency anaemia secondary to blood loss (chronic)	1173
D50.1 Sideropenic dysphagia	15
D50.8 Other iron deficiency anaemias	7558
D50.9 Iron deficiency anaemia, unspecified	19534

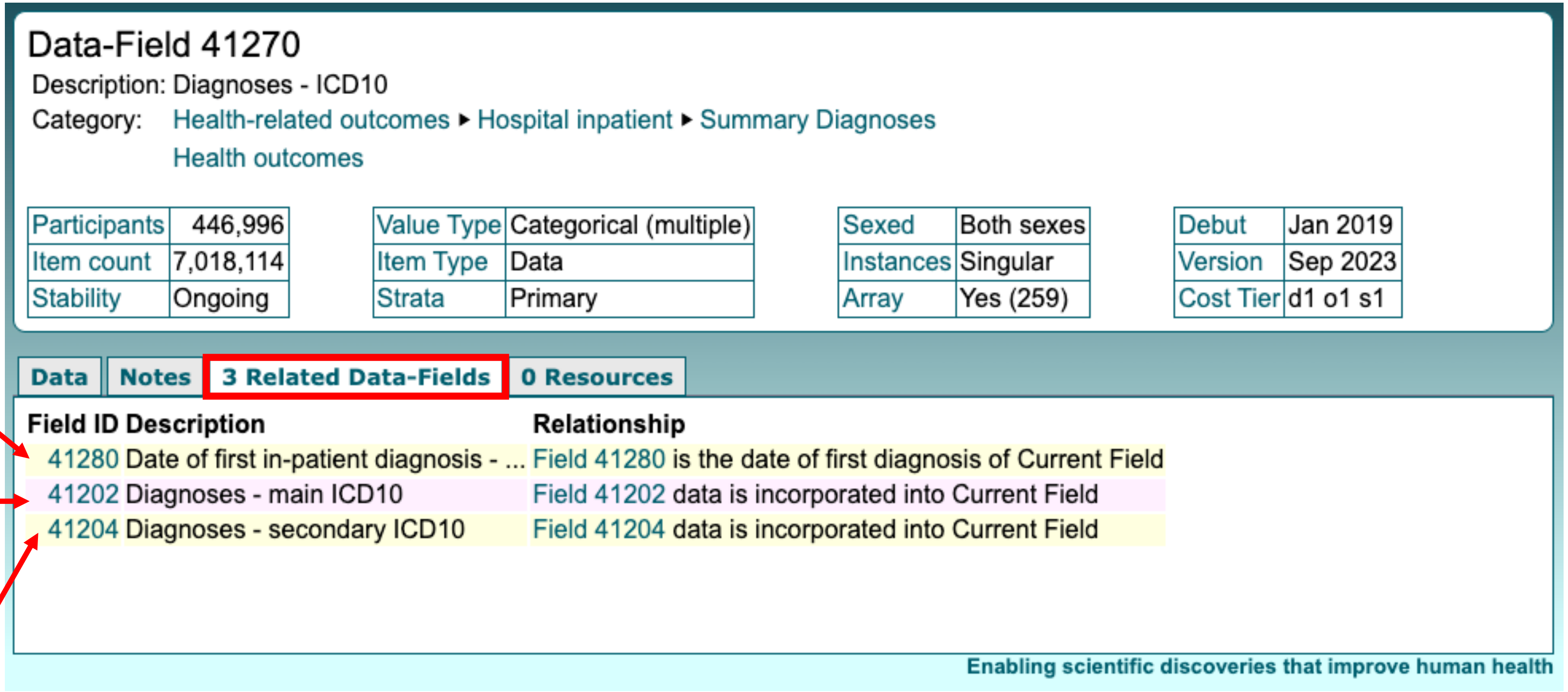
number of individuals with the code

Hospital inpatient data: ICD10 codes

Date of diagnosis available

Main reason for admission

Additional diagnoses during stay



Data-Field 41270

Description: Diagnoses - ICD10

Category: Health-related outcomes ▸ Hospital inpatient ▸ Summary Diagnoses
Health outcomes

Participants	446,996
Item count	7,018,114
Stability	Ongoing

Value Type	Categorical (multiple)
Item Type	Data
Strata	Primary

Sexed	Both sexes
Instances	Singular
Array	Yes (259)

Debut	Jan 2019
Version	Sep 2023
Cost Tier	d1 o1 s1

Field ID	Description	Relationship
41280	Date of first in-patient diagnosis - ...	Field 41280 is the date of first diagnosis of Current Field
41202	Diagnoses - main ICD10	Field 41202 data is incorporated into Current Field
41204	Diagnoses - secondary ICD10	Field 41204 data is incorporated into Current Field

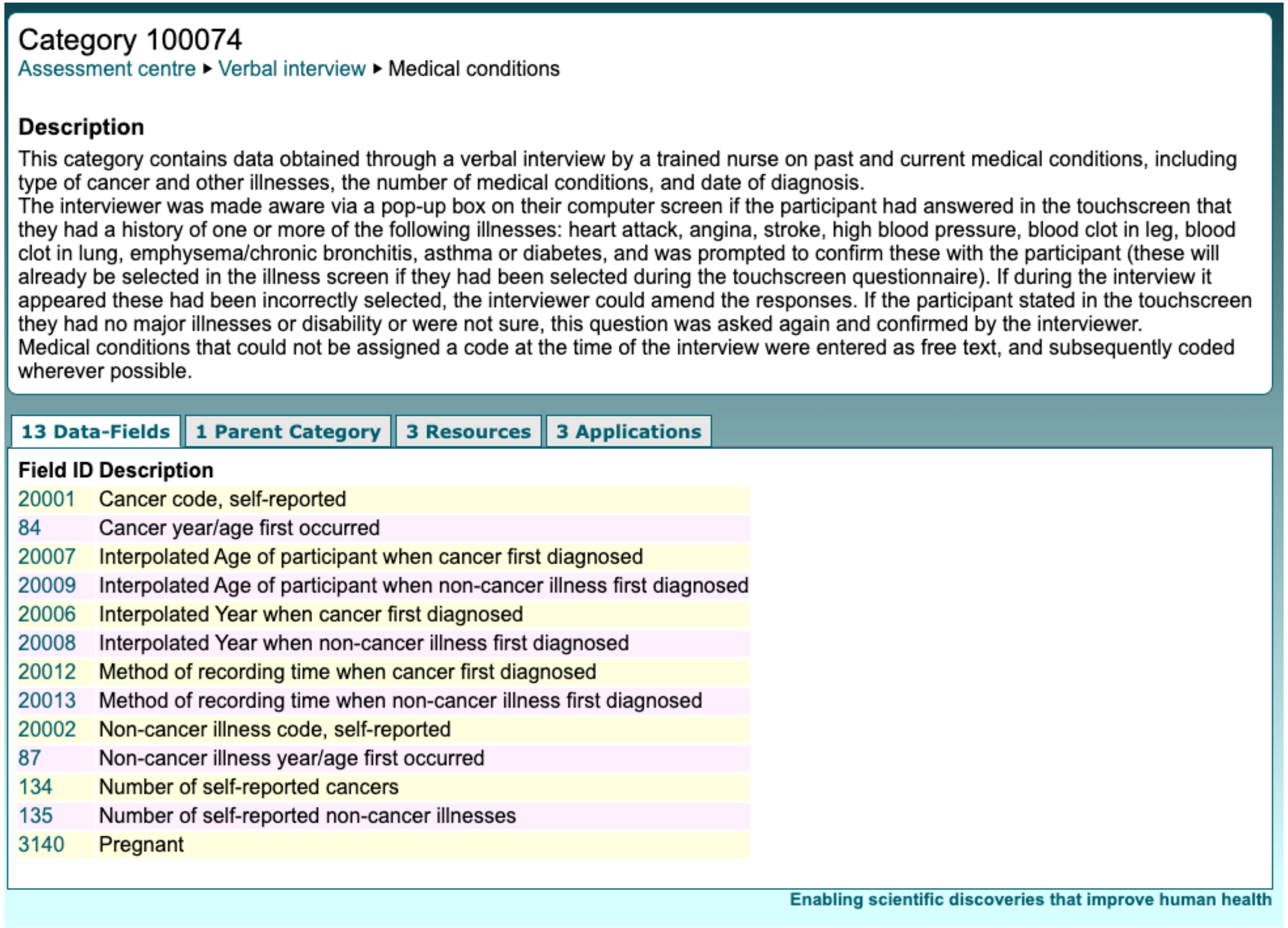
Enabling scientific discoveries that improve human health

Hospital inpatient data: ICD10 codes

Example data structure

EID	ICD10.0	ICD10.1	ICD10.2	ICD10.3	ICD10.4	ICD10.5
1	D50.1 Sideropenic dysphagia	K36 Other appendicitis				
2	N60.4 Mammary duct ectasia	N99.0 Postprocedural renal failure	O04.9 Complete or unspecified, without complication	S70.7 Multiple superficial injuries of hip and thigh	K36 Other appendicitis	
3						
4	K51.0 Ulcerative (chronic) enterocolitis	H40.9 Glaucoma, unspecified	F20.5 Residual schizophrenia	F40.0 Agoraphobia	G72.4 Inflammatory myopathy, not elsewhere classified	V10.0 Driver injured in nontraffic accident

Self Report Conditions



Category 100074

Assessment centre ▸ Verbal interview ▸ Medical conditions

Description

This category contains data obtained through a verbal interview by a trained nurse on past and current medical conditions, including type of cancer and other illnesses, the number of medical conditions, and date of diagnosis.

The interviewer was made aware via a pop-up box on their computer screen if the participant had answered in the touchscreen that they had a history of one or more of the following illnesses: heart attack, angina, stroke, high blood pressure, blood clot in leg, blood clot in lung, emphysema/chronic bronchitis, asthma or diabetes, and was prompted to confirm these with the participant (these will already be selected in the illness screen if they had been selected during the touchscreen questionnaire). If during the interview it appeared these had been incorrectly selected, the interviewer could amend the responses. If the participant stated in the touchscreen they had no major illnesses or disability or were not sure, this question was asked again and confirmed by the interviewer.

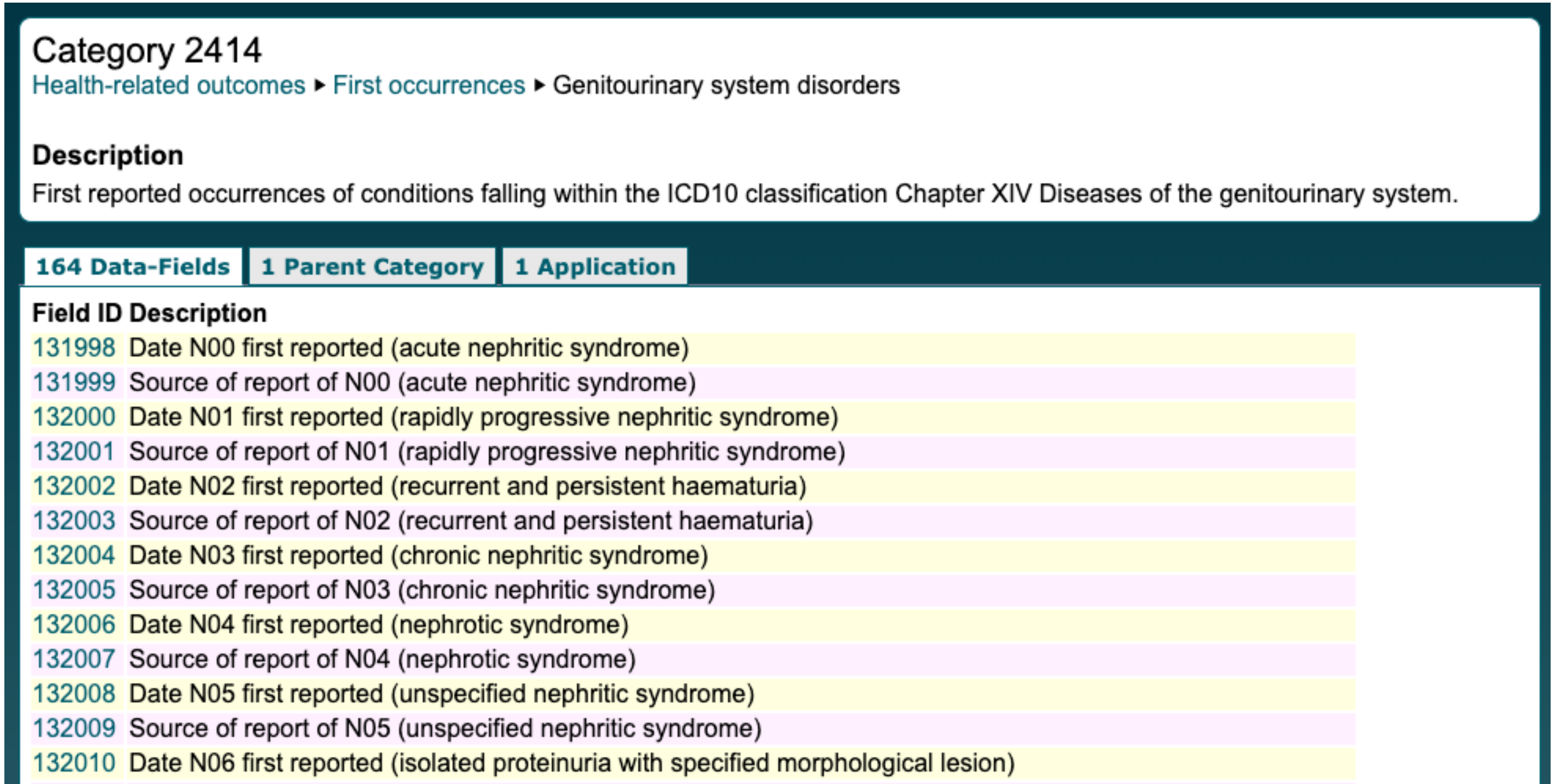
Medical conditions that could not be assigned a code at the time of the interview were entered as free text, and subsequently coded wherever possible.

Field ID	Description
20001	Cancer code, self-reported
84	Cancer year/age first occurred
20007	Interpolated Age of participant when cancer first diagnosed
20009	Interpolated Age of participant when non-cancer illness first diagnosed
20006	Interpolated Year when cancer first diagnosed
20008	Interpolated Year when non-cancer illness first diagnosed
20012	Method of recording time when cancer first diagnosed
20013	Method of recording time when non-cancer illness first diagnosed
20002	Non-cancer illness code, self-reported
87	Non-cancer illness year/age first occurred
134	Number of self-reported cancers
135	Number of self-reported non-cancer illnesses
3140	Pregnant

Enabling scientific discoveries that improve human health

First Occurrences

- Created by UKB - available for a wide range of health outcomes across self-report, primary care, hospital inpatient data and death data, mapped to a 3-character ICD10 code.
- Great for single-disease approaches



Category 2414

Health-related outcomes ▸ First occurrences ▸ Genitourinary system disorders

Description

First reported occurrences of conditions falling within the ICD10 classification Chapter XIV Diseases of the genitourinary system.

Field ID	Description
131998	Date N00 first reported (acute nephritic syndrome)
131999	Source of report of N00 (acute nephritic syndrome)
132000	Date N01 first reported (rapidly progressive nephritic syndrome)
132001	Source of report of N01 (rapidly progressive nephritic syndrome)
132002	Date N02 first reported (recurrent and persistent haematuria)
132003	Source of report of N02 (recurrent and persistent haematuria)
132004	Date N03 first reported (chronic nephritic syndrome)
132005	Source of report of N03 (chronic nephritic syndrome)
132006	Date N04 first reported (nephrotic syndrome)
132007	Source of report of N04 (nephrotic syndrome)
132008	Date N05 first reported (unspecified nephritic syndrome)
132009	Source of report of N05 (unspecified nephritic syndrome)
132010	Date N06 first reported (isolated proteinuria with specified morphological lesion)

First Occurrences

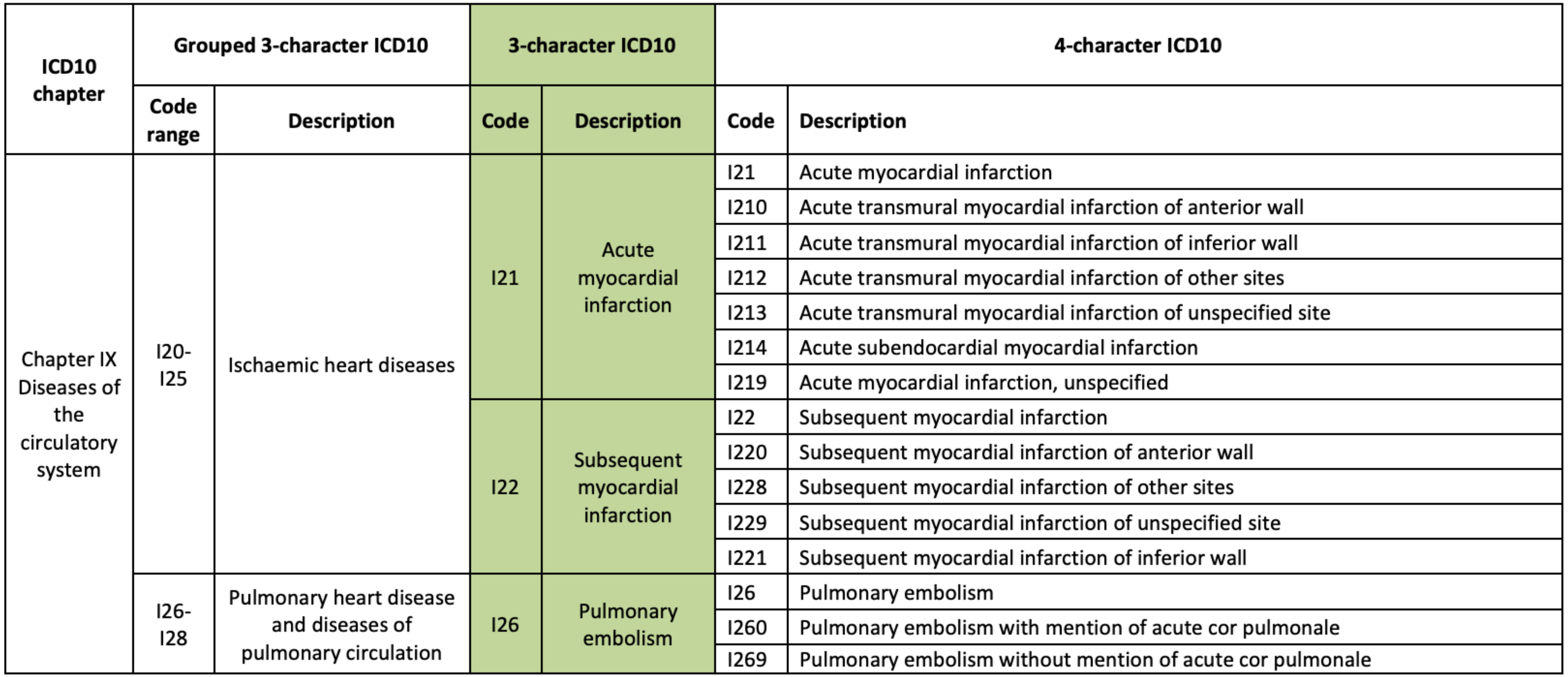
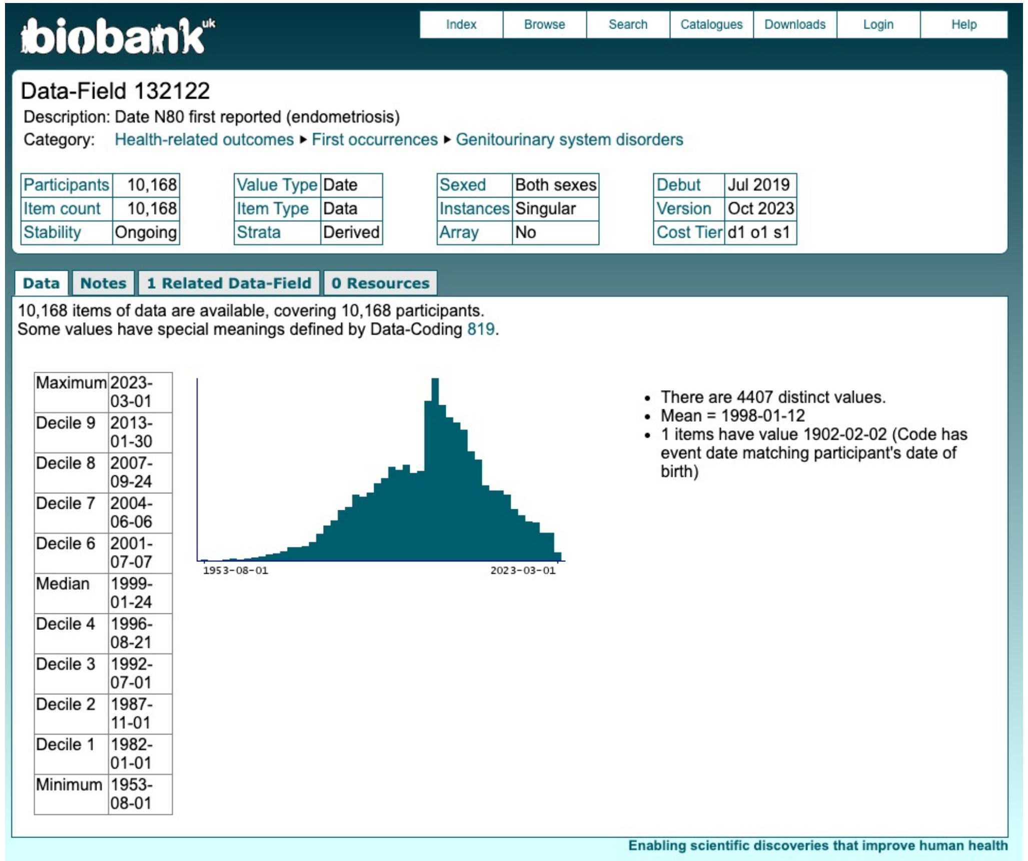


Table 1. Selected examples of 4-character ICD10 codes and relation to 3-character level code, code range, chapter and description at each hierarchical level. The green cells indicate the level at which the health outcomes for the first occurrence are defined.

ICD10 chapter	**Grouped 3-character ICD10**		**3-character ICD10**		**4-character ICD10**	
	Code range	**Description**	**Code**	**Description**	**Code**	**Description**
Chapter IX Diseases of the circulatory system	I20-I25	Ischaemic heart diseases	I21	Acute myocardial infarction	I21	Acute myocardial infarction
					I210	Acute transmural myocardial infarction of anterior wall
					I211	Acute transmural myocardial infarction of inferior wall
					I212	Acute transmural myocardial infarction of other sites
					I213	Acute transmural myocardial infarction of unspecified site
					I214	Acute subendocardial myocardial infarction
					I219	Acute myocardial infarction, unspecified
			I22	Subsequent myocardial infarction	I22	Subsequent myocardial infarction
					I220	Subsequent myocardial infarction of anterior wall
					I228	Subsequent myocardial infarction of other sites
					I229	Subsequent myocardial infarction of unspecified site
					I221	Subsequent myocardial infarction of inferior wall
	I26-I28	Pulmonary heart disease and diseases of pulmonary circulation	I26	Pulmonary embolism	I26	Pulmonary embolism
					I260	Pulmonary embolism with mention of acute cor pulmonale
					I269	Pulmonary embolism without mention of acute cor pulmonale

https://biobank.ndph.ox.ac.uk/ukb/ukb/docs/first_occurrences_outcomes.pdf

First Occurrences



Data-Field 132122

Description: Date N80 first reported (endometriosis)

Category: Health-related outcomes ▸ First occurrences ▸ Genitourinary system disorders

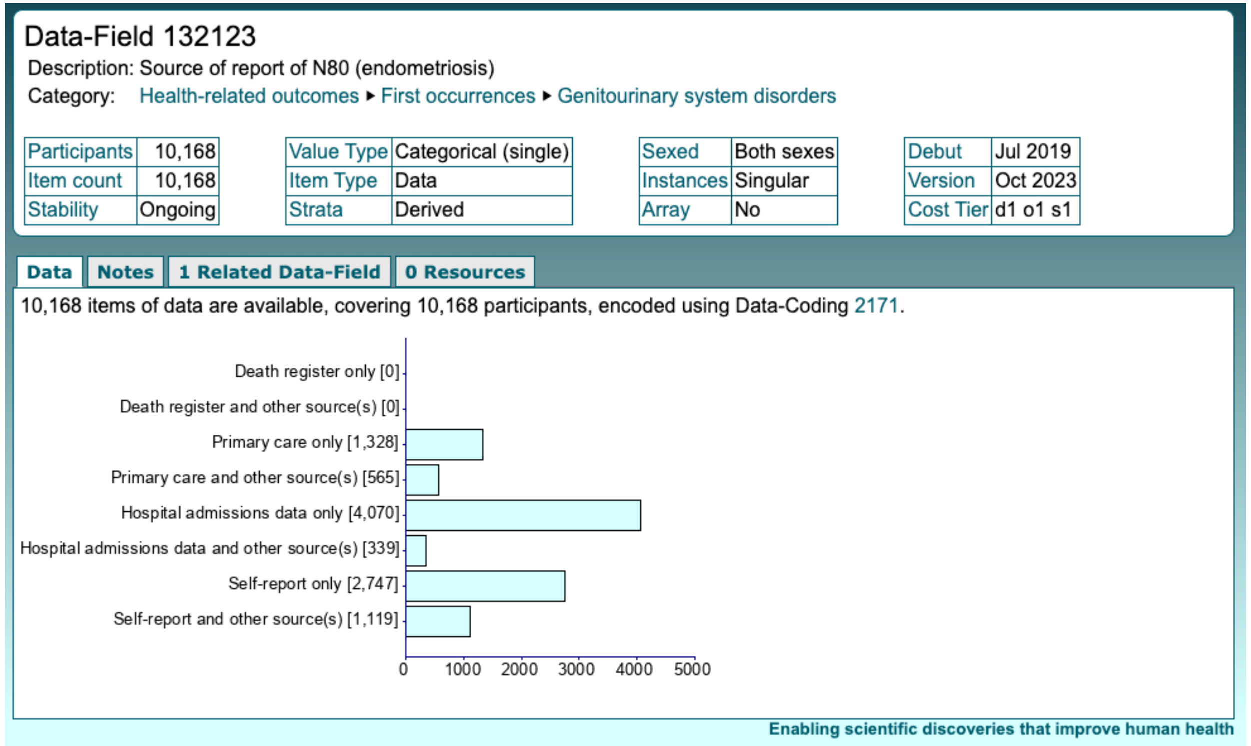
Participants	10,168	Value Type	Date	Sexed	Both sexes	Debut	Jul 2019
Item count	10,168	Item Type	Data	Instances	Singular	Version	Oct 2023
Stability	Ongoing	Strata	Derived	Array	No	Cost Tier	d1 o1 s1

10,168 items of data are available, covering 10,168 participants.
Some values have special meanings defined by Data-Coding 819.

Maximum	2023-03-01
Decile 9	2013-01-30
Decile 8	2007-09-24
Decile 7	2004-06-06
Decile 6	2001-07-07
Median	1999-01-24
Decile 4	1996-08-21
Decile 3	1992-07-01
Decile 2	1987-11-01
Decile 1	1982-01-01
Minimum	1953-08-01

- There are 4407 distinct values.
- Mean = 1998-01-12
- 1 items have value 1902-02-02 (Code has event date matching participant's date of birth)

Enabling scientific discoveries that improve human health


Data-Field 132123

Description: Source of report of N80 (endometriosis)

Category: Health-related outcomes ▸ First occurrences ▸ Genitourinary system disorders

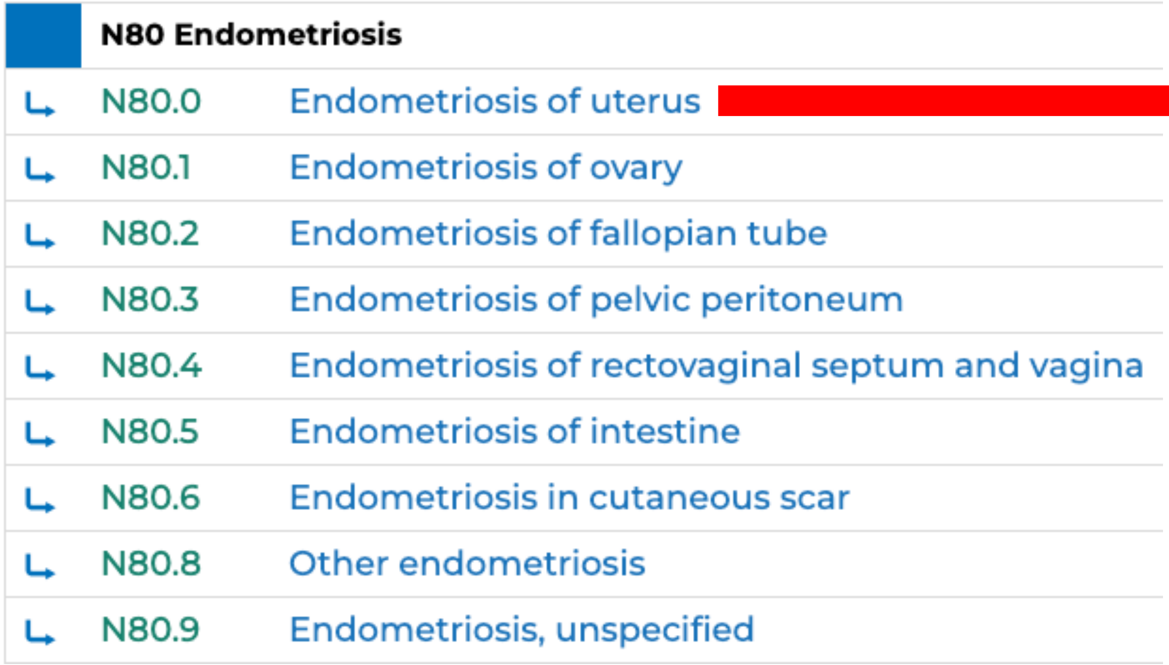
Participants	10,168	Value Type	Categorical (single)	Sexed	Both sexes	Debut	Jul 2019
Item count	10,168	Item Type	Data	Instances	Singular	Version	Oct 2023
Stability	Ongoing	Strata	Derived	Array	No	Cost Tier	d1 o1 s1

10,168 items of data are available, covering 10,168 participants, encoded using Data-Coding 2171.

Death register only [0]
Death register and other source(s) [0]
Primary care only [1,328]
Primary care and other source(s) [565]
Hospital admissions data only [4,070]
Hospital admissions data and other source(s) [339]
Self-report only [2,747]
Self-report and other source(s) [1,119]
0
1000
2000
3000
4000
5000

Enabling scientific discoveries that improve human health

Same 3-character ICD10 code ≠ Same Disease



	N80 Endometriosis	
↳	N80.0	Endometriosis of uterus
↳	N80.1	Endometriosis of ovary
↳	N80.2	Endometriosis of fallopian tube
↳	N80.3	Endometriosis of pelvic peritoneum
↳	N80.4	Endometriosis of rectovaginal septum and vagina
↳	N80.5	Endometriosis of intestine
↳	N80.6	Endometriosis in cutaneous scar
↳	N80.8	Other endometriosis
↳	N80.9	Endometriosis, unspecified

Endometriosis of uterus → Adenomyosis

Phecode Mapping

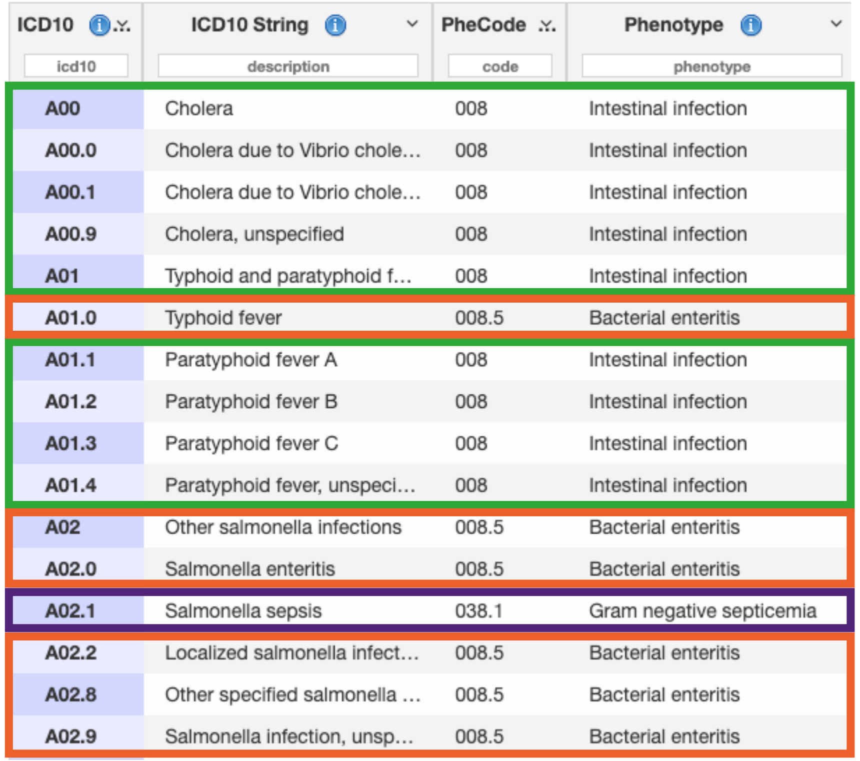
Lots of specific ICD10 codes → low power for discovery

Phecodes: a high-throughput strategy for defining phenotypes using ICD10 codes

Wu P, Gifford A, Meng X, Li X, Campbell H, Varley T, Zhao J, Carroll R, Bastarache L, Denny JC, Theodoratou E, Wei W

Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation

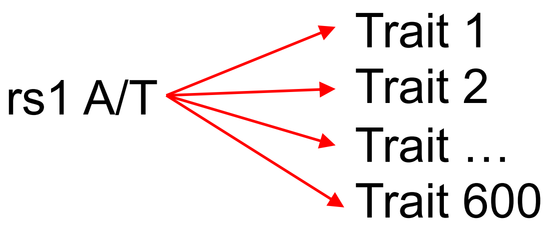
JMIR Med Inform 2019;7(4):e14325
doi: 10.2196/14325 PMID: 31553307 PMCID: 6911227



ICD10	ICD10 String	PheCode	Phenotype
icd10	description	code	phenotype
A00	Cholera	008	Intestinal infection
A00.0	Cholera due to Vibrio chole…	008	Intestinal infection
A00.1	Cholera due to Vibrio chole…	008	Intestinal infection
A00.9	Cholera, unspecified	008	Intestinal infection
A01	Typhoid and paratyphoid f…	008	Intestinal infection
A01.0	Typhoid fever	008.5	Bacterial enteritis
A01.1	Paratyphoid fever A	008	Intestinal infection
A01.2	Paratyphoid fever B	008	Intestinal infection
A01.3	Paratyphoid fever C	008	Intestinal infection
A01.4	Paratyphoid fever, unspeci…	008	Intestinal infection
A02	Other salmonella infections	008.5	Bacterial enteritis
A02.0	Salmonella enteritis	008.5	Bacterial enteritis
A02.1	Salmonella sepsis	038.1	Gram negative septicemia
A02.2	Localized salmonella infect…	008.5	Bacterial enteritis
A02.8	Other specified salmonella …	008.5	Bacterial enteritis
A02.9	Salmonella infection, unsp…	008.5	Bacterial enteritis

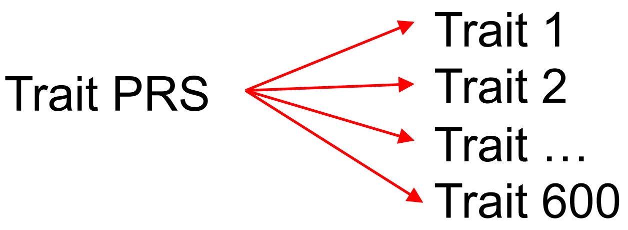
- Maps can be downloaded from the PheWAS catalog
  - https://phewascatalog.org/files/Phecode_map_v1_2_icd10_beta.csv.zip
  - https://phewascatalog.org/files/phecode_definitions1.2.csv.zip.

Phenome-wide association study (PheWAS)



rs1 A/T

Trait 1
Trait 2
Trait …
Trait 600



Trait PRS

Trait 1
Trait 2
Trait …
Trait 600

Why?

- Explore pleiotropic SNP effects
- Investigate the pleiotropic effects of genetic liability to a condition/trait in individuals with or without the condition
- Utilisation of a cohort that hasn't been phenotyped for the condition if interested in genetic liability

What do you need?

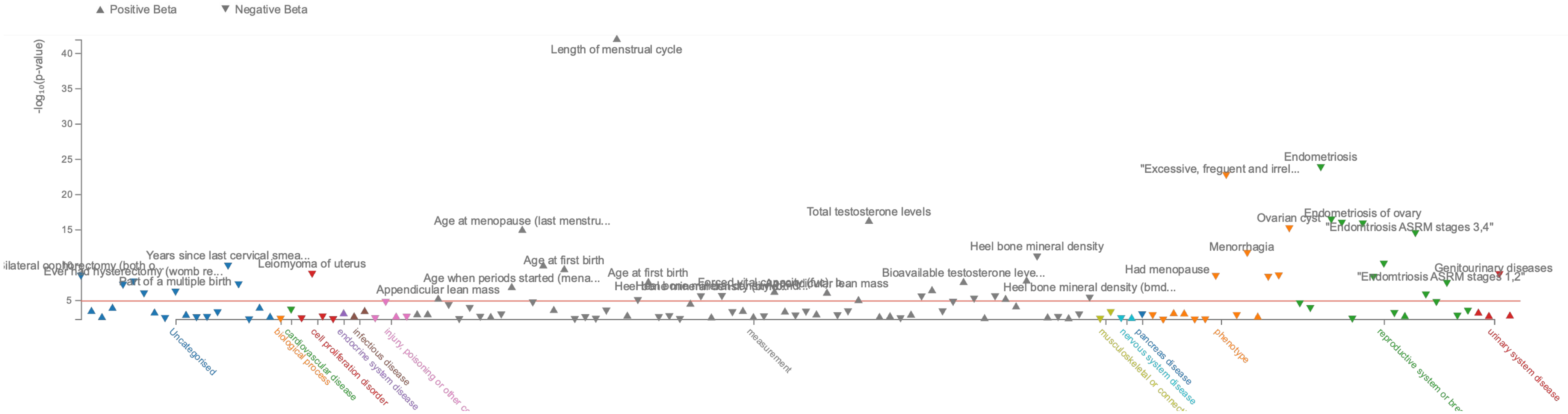
- Large cohort of individuals phenotyped for many traits with matched genotype data
- If running single SNP PheWAS, a SNP of interest
- If running PRS PheWAS, GWAS summary statistics derived from a different cohort

Readily Available PheWAS Results

OpenTargets Genetics

PheWASs performed in UKB, FinnGen and GWAS Catalog available

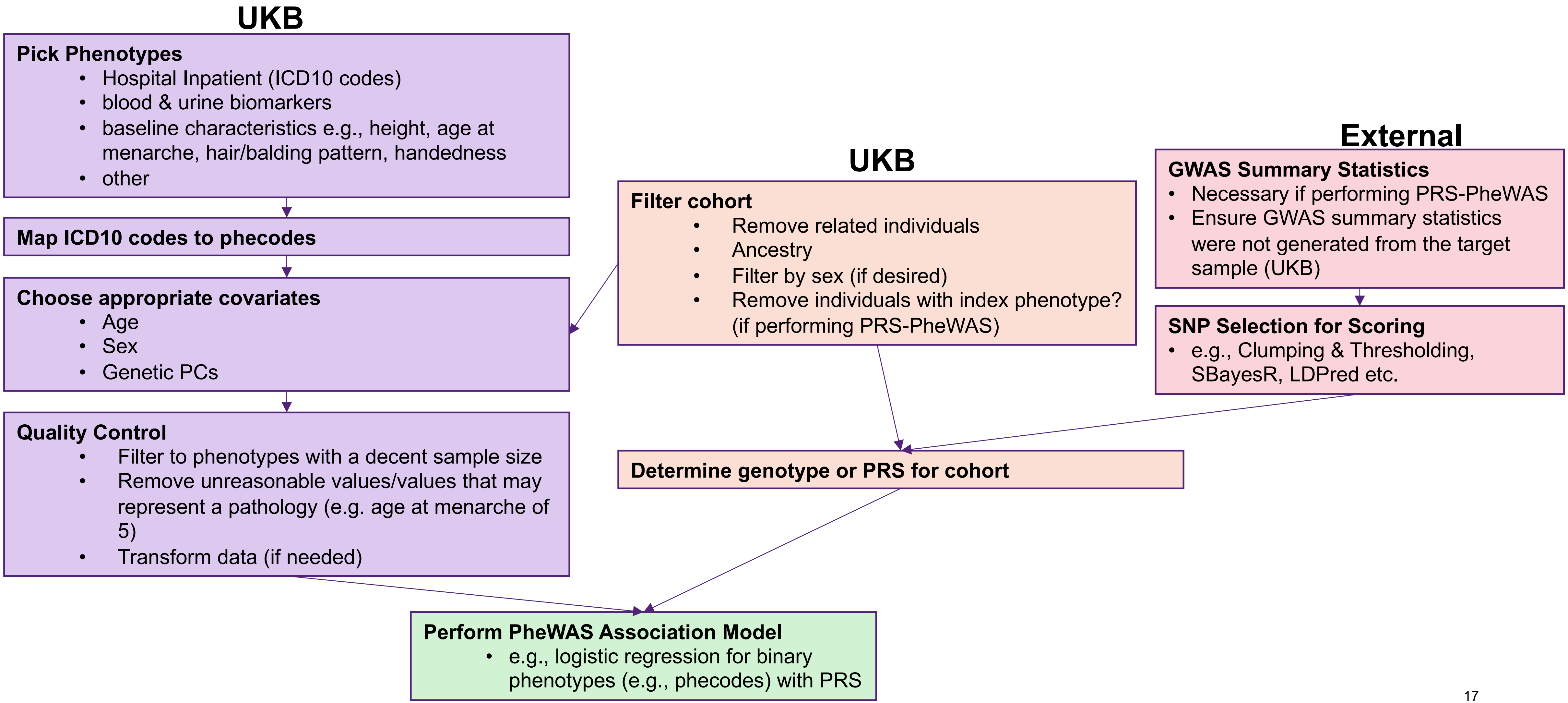
rs11031005 (T/C) PheWAS



Only traits with P-value < 0.005 are shown

https://genetics.opentargets.org/

Performing a PheWAS in the UKB



UKB

Pick Phenotypes

- Hospital Inpatient (ICD10 codes)
- blood & urine biomarkers
- baseline characteristics e.g., height, age at menarche, hair/balding pattern, handedness
- other

Map ICD10 codes to phecodes

Choose appropriate covariates

- Age
- Sex
- Genetic PCs

Quality Control

- Filter to phenotypes with a decent sample size
- Remove unreasonable values/values that may represent a pathology (e.g. age at menarche of 5)
- Transform data (if needed)

UKB

Filter cohort

- Remove related individuals
- Ancestry
- Filter by sex (if desired)
- Remove individuals with index phenotype? (if performing PRS-PheWAS)

Determine genotype or PRS for cohort

External

GWAS Summary Statistics

- Necessary if performing PRS-PheWAS
- Ensure GWAS summary statistics were not generated from the target sample (UKB)

SNP Selection for Scoring

- e.g., Clumping & Thresholding, SBayesR, LDPred etc.

Perform PheWAS Association Model

- e.g., logistic regression for binary phenotypes (e.g., phecodes) with PRS

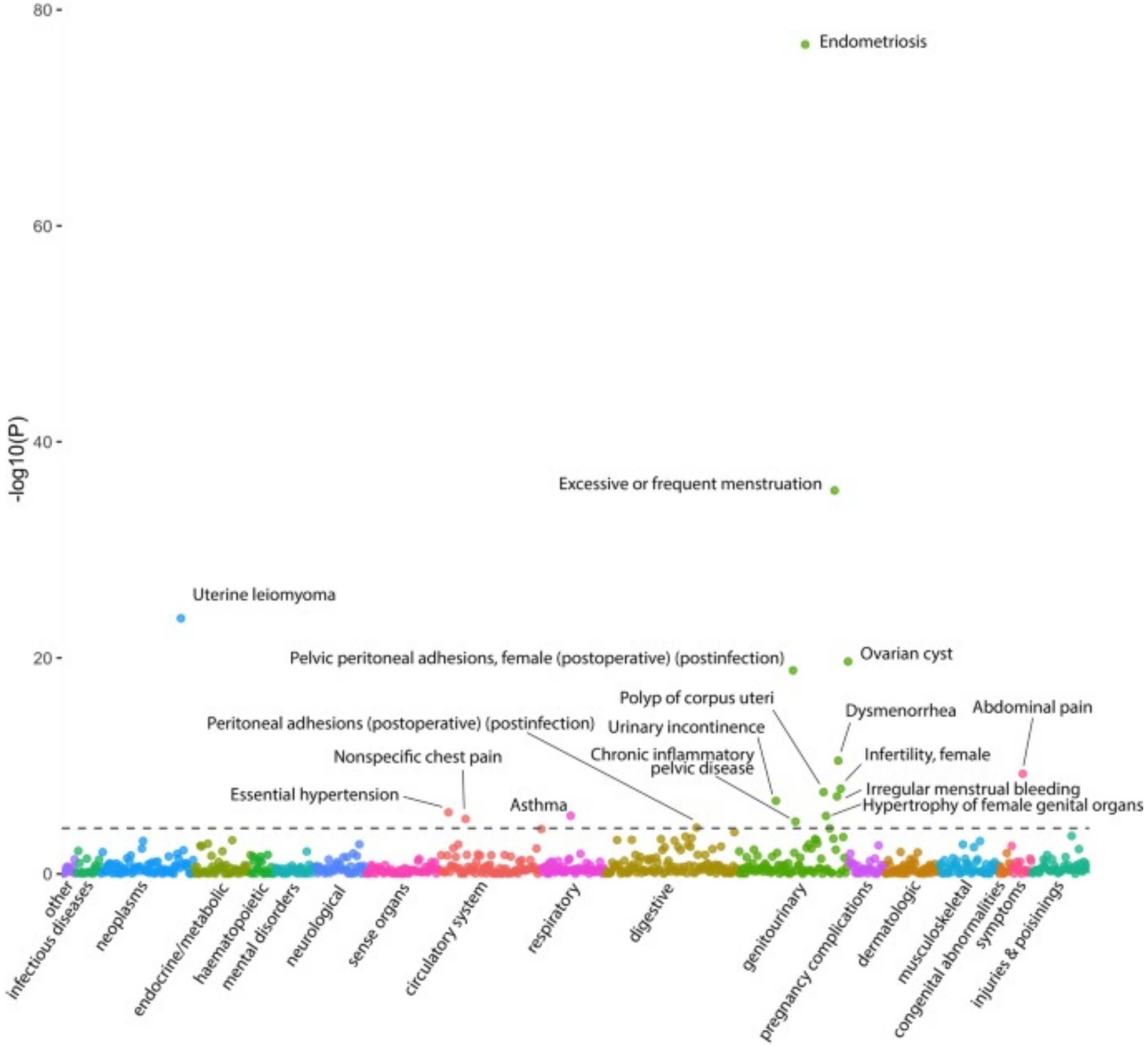
Running a PheWAS in UKB

EID	PRS	Age	PC1	…	PC10	Phecode1	Phecode2	…	Phecode600
1	0.61	65	0.345		-0.235	1	0		1
2	0.05	82	0.214		0.521	0	1		1
3	1.24	76	-0.531		0.951	0	0		0
4	-0.85	72	0.001		-0.315	0	0		0
5	-0.41	89	-0.817		0.128	1	0		0
6	0.29	91	0.412		-0.469	0	0		1

```
glm(Phecode1~PRS+Age+PC1+PC2+PC3+PC4+PC5+PC6+PC7+PC8+PC9+PC10, data=dataframe1, family="binomial")
```

- Repeat for all phecodes
- Correct for multiple testing

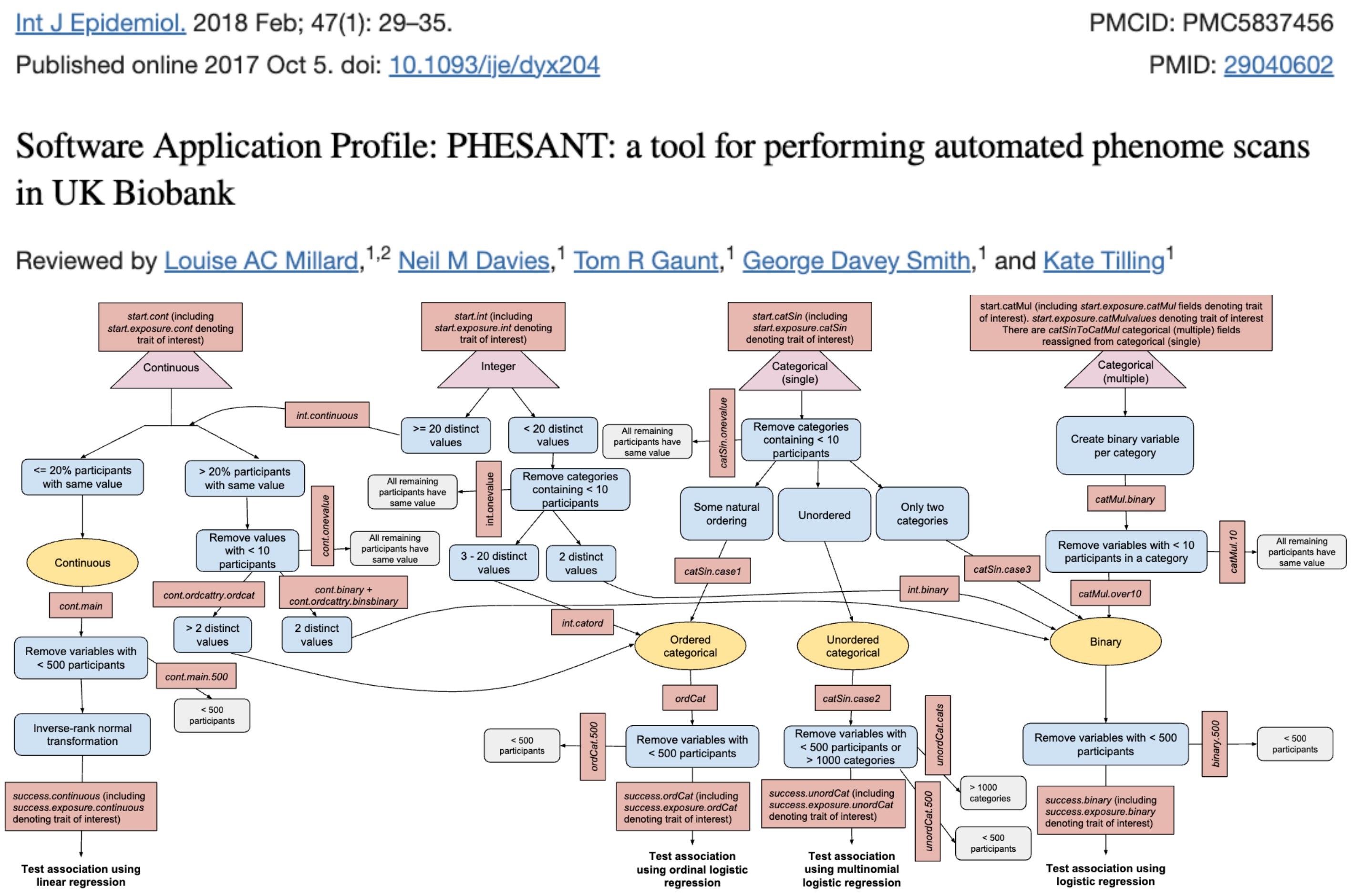
PRS-PheWAS for Endometriosis



Plotted: -log10(P) of logistic regression testing endometriosis PRS ~ Phenotype + age + 10 genetic PCs

McGrath, I.M., International Endometriosis Genetics Consortium., Montgomery, G.W. *et al.* Polygenic risk score phenome-wide association study reveals an association between endometriosis and testosterone. *BMC Med* **21**, 482 (2023). https://doi.org/10.1186/s12916-023-03184-z

Tools available for PheWAS



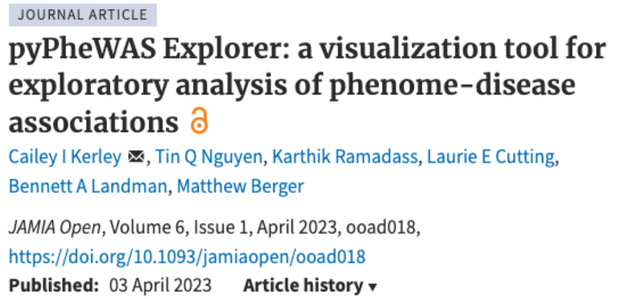
Int J Epidemiol. 2018 Feb; 47(1): 29–35.
Published online 2017 Oct 5. doi: 10.1093/ije/dyx204

PMCID: PMC5837456
PMID: 29040602

Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank

Reviewed by Louise AC Millard,[1,2] Neil M Davies,[1] Tom R Gaunt,[1] George Davey Smith,[1] and Kate Tilling[1]

start.cont (including start.exposure.cont denoting trait of interest)
Continuous
int.continuous
<= 20% participants with same value
> 20% participants with same value
Continuous
cont.main
Remove variables with < 500 participants
cont.main.500
< 500 participants
Inverse-rank normal transformation
success.continuous (including success.exposure.continuous denoting trait of interest)
Test association using linear regression
Remove values with < 10 participants
cont.onevalue
All remaining participants have same value
cont.ordcattry.ordcat
cont.binary + cont.ordcattry.binsbinary
> 2 distinct values
2 distinct values
start.int (including start.exposure.int denoting trait of interest)
Integer
>= 20 distinct values
< 20 distinct values
All remaining participants have same value
int.onevalue
Remove categories containing < 10 participants
3 - 20 distinct values
2 distinct values
int.catord
int.binary
start.catSin (including start.exposure.catSin denoting trait of interest)
Categorical (single)
catSin.onevalue
All remaining participants have same value
Remove categories containing < 10 participants
Some natural ordering
Unordered
Only two categories
catSin.case1
catSin.case3
Ordered categorical
Unordered categorical
ordCat
catSin.case2
Remove variables with < 500 participants
ordCat.500
< 500 participants
success.ordCat (including success.exposure.ordCat denoting trait of interest)
Test association using ordinal logistic regression
Remove variables with < 500 participants or > 1000 categories
unordCat.cats
> 1000 categories
unordCat.500
< 500 participants
success.unordCat (including success.exposure.unordCat denoting trait of interest)
Test association using multinomial logistic regression
start.catMul (including start.exposure.catMul fields denoting trait of interest). start.exposure.catMulvalues denoting trait of interest There are catSinToCatMul categorical (multiple) fields reassigned from categorical (single)
Categorical (multiple)
Create binary variable per category
catMul.binary
Remove variables with < 10 participants in a category
catMul.10
All remaining participants have same value
catMul.over10
Binary
Remove variables with < 500 participants
binary.500
< 500 participants
success.binary (including success.exposure.binary denoting trait of interest)
Test association using logistic regression

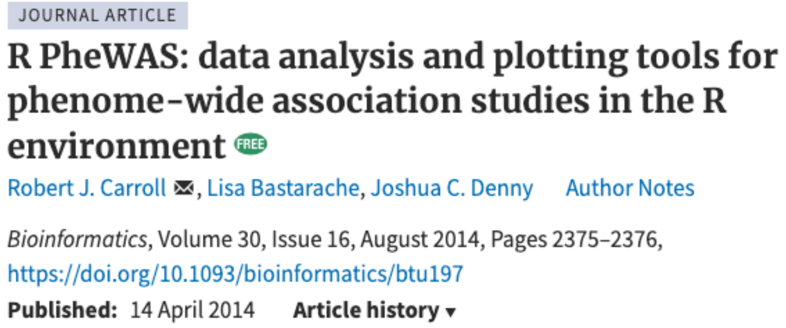


JOURNAL ARTICLE

pyPheWAS Explorer: a visualization tool for exploratory analysis of phenome-disease associations

Cailey I Kerley, Tin Q Nguyen, Karthik Ramadass, Laurie E Cutting, Bennett A Landman, Matthew Berger

JAMIA Open, Volume 6, Issue 1, April 2023, ooad018,
https://doi.org/10.1093/jamiaopen/ooad018
Published: 03 April 2023 **Article history ▾**



JOURNAL ARTICLE

R PheWAS: data analysis and plotting tools for phenome-wide association studies in the R environment FREE

Robert J. Carroll, Lisa Bastarache, Joshua C. Denny Author Notes

Bioinformatics, Volume 30, Issue 16, August 2014, Pages 2375–2376,
https://doi.org/10.1093/bioinformatics/btu197
Published: 14 April 2014 **Article history ▾**

Summary

- Multiple sources of health data in UKB
- When interested in a particular phenotype, use search & browse functions to find relevant fields to maximise sample size
- ICD10 codes useful for high-throughput screens
- PheWAS approach requires a cohort with matched phenotype and genotype data, and can reveal novel pleiotropic effects
- PheWAS results should be considered exploratory and require validation

biobank uk

Enabling scientific discoveries that improve human health