

Data Access: Who can access the data and how?

Rory Collins

UK Biobank Principal Investigator

BHF Professor of Medicine & Epidemiology

Nuffield Department of Population Health

University of Oxford, UK

UK Biobank: Principles of Access

- Available to academic or commercial researchers, for all types of health-related research that is in public interest
- No preferential or exclusive access; use of the Resource does not involve “collaboration” with UK Biobank
- Researchers only have to pay for the costs of using the Resource (and not for any of the costs of setting it up)
- Access to the biological samples that are limited and depletable will be carefully controlled and coordinated
- Researchers are required to publish their findings and return derived data so other researchers can use them (but any research-related IP is retained by researchers)

UK Biobank Prospective Cohort

- 500,000 UK men and women aged 40-69 years when recruited and assessed during 2006-2010
- General consent for all types of health research and follow-up through all health-related records (as well as for re-contact for specific purposes)
- Extensive baseline questions and measurements, with biological samples stored for future assays
- Enhancements in large subsets of participants, and assays of samples from all participants
- Repeat assessments over time in subsets of participants to allow for sources of variation

Baseline sample collection: different types of biological sample allowing a wide range of different assays

Sample collection tube	Fractions collected	Potential assays
Na ⁺ EDTA	<ul style="list-style-type: none"> • Plasma • Buffy coat • Red cells 	<ul style="list-style-type: none"> • Plasma proteome and metabolome • Assays of genomic DNA • Membrane lipids and heavy metals
Lithium Heparin (PST)	<ul style="list-style-type: none"> • Plasma 	<ul style="list-style-type: none"> • Plasma proteome and metabolome (without haemolysis)
Silica clot accelerator (SST)	<ul style="list-style-type: none"> • Serum 	<ul style="list-style-type: none"> • Serum proteome and metabolome (without haemolysis)
Acid citrate dextrose	<ul style="list-style-type: none"> • Whole blood 	<ul style="list-style-type: none"> • Assays of DNA extracted from EBV immortalised cell lines • B-cell transcriptome
EDTA	<ul style="list-style-type: none"> • Whole blood 	<ul style="list-style-type: none"> • Standard haematological parameters
Tempus RNA stabilisation	<ul style="list-style-type: none"> • Whole blood with lysis reagent 	<ul style="list-style-type: none"> • Blood transcriptome • Representative transcriptomes of other tissues
Urine	<ul style="list-style-type: none"> • Urine 	<ul style="list-style-type: none"> • Urine proteome and metabolome • Gut microbiome
Saliva	<ul style="list-style-type: none"> • Mixed saliva sample 	<ul style="list-style-type: none"> • Salivary proteome and metabolome • Salivary microbiome • (Mucosal proteome and metabolome)

Baseline questionnaire: assessing a wide range of socio-economic, lifestyle and environmental factors

Self-completion: topics	Median time (minutes)	Interview: topics	Median time (minutes)
Socio-demographics	1.7	Medical history/medication	3.1
Ethnicity	0.1	Occupation	0.4
Work-employment	1.4	Other	0.6
Physical activity	4.4	Total time	4.1
Smoking (non-smokers)	0.5		
(past/current smokers)	1.5		
Diet (food frequency)*	4.5		
Alcohol	1.1		
Sleep	1.2		
Sun exposure	1.3		
Environmental exposures	1.0		
Early life factors	0.8		
Family history of common diseases	1.6		
Reproductive history & screening (women)	2.4		
(men)	0.8		
Sexual history	0.4		
General health	2.1		
Past medical history & medications	1.6		
Noise exposure	1.0		
Psychological status	4.5		
Cognitive function tests	10.0		
Hearing speech-in-noise test	8.0		
Total time	52.5		

*Subset of 200,000 participants: repeated daily diet diaries conducted via the internet

Touchscreen and interview questions (plus extra enhancement questions) available at www.ukbiobank.ac.uk

Baseline assessment: Standard physical measures (with enhanced measures made in large subsets)

All 500,000 participants

- Blood pressure & heart rate
- Height (standing/seated)
- Waist/hip circumference
- Weight/impedance
- Spirometry
- Heel ultrasound

Subset: 175,000 participants

- Hearing test
- Vascular reactivity

Subset: 120,000 participants

- Visual acuity, refractive index & intraocular pressure

Subset: 85,000 participants

- Retinal images & optical coherence tomograms
- Fitness test & ECG limb leads

Some researchers (e.g. Eye and Vision Consortium) have taken the lead in turning data into information



UK Biobank Eye and Vision Consortium

Home

The Consortium

Research Groups

Group Activities

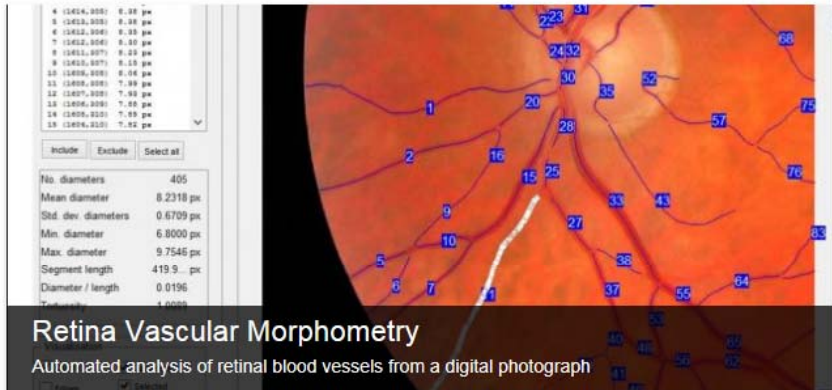
People

Funding

Publications

News

Contact



The Consortium

- We are a group of scientists and clinicians with an interest in eyes, vision and diseases which affect sight
- We represent leading UK Universities and Hospitals
- The consortium management is based at Moorfields Eye Hospital and The UCL Institute of Ophthalmology in London

[Learn More](#)

About UK Biobank



UK Biobank is a major national health resource, and a registered charity in its own right, with the aim of improving the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses.

[Read More](#)

Funding



The UK Biobank parent study was established by the Wellcome Trust medical charity, Medical Research Council, Department of Health, Scottish Government and the Northwest Regional Development Agency.

[Read More](#)

What we do



- We use data from UK Biobank to study what causes eye disease and vision loss and investigate how these can be controlled and prevented
- Research into the causes of disease help to develop new, more effective treatments

[Read More](#)

Other enhancements of UK Biobank phenotyping being driven by experts in particular topic areas

- Web-based assessments of diet completed; and currently doing cognitive function (2015)
- Wrist-worn accelerometers are being worn for one week by 100,000 participants (2013-15)
- Biobank chip to genotype (GWAS; candidate SNPs; exome) all participants (2013-15)
- Standard panel of assays (e.g. lipids; clotting) on samples from all participants (2014-16)
- Multiple imaging modalities (brain/heart/body MRI; bone/joint DEXA) and 2-week cardiac monitoring in 100,000 participants (2014-19)

Keeping participants informed about developments (email addresses for ~340,000 of the participants)

The screenshot shows a web browser window displaying the UK Biobank Imaging Study website. The browser's address bar shows the URL: <https://securehost.tvf.co.uk/opencc.co.uk/ukbiobankmicrosite/>. The website features the Biobank UK logo and the text "Imaging study". A large orange banner on the right side says "FEEDBACK". The main content area includes a "Welcome to the UK Biobank Imaging Study" heading, a paragraph of text, and a blue "Introduction" tab. Below the tab is a large graphic with a play button icon, a silhouette of a person, an MRI scanner, and icons of a brain, heart, and kidney. To the right of this graphic is a vertical menu with four items: "Introduction", "Consent", "Feedback", and "Eligibility", each with a small video thumbnail. At the bottom of the main content area, there are two orange buttons: "Find out more about the research" with a right-pointing arrow, and "Further documents". Below these buttons, the text reads "Improving the health of future generations" and "Go to the UK Biobank website" with an external link icon. At the very bottom, there is a footer with registration information: "Incorporated in England and Wales, registered number 4978912, and registered as a charity in England and Wales, number 1101332. Charity registered in Scotland, number SC039230. © UK Biobank 2014". The Windows taskbar at the bottom shows the time as 11:33 on 20/02/2014.

UK Biobank Imaging Study

<https://securehost.tvf.co.uk/opencc.co.uk/ukbiobankmicrosite/>

biobank^{uk}
Imaging study

Welcome to the UK Biobank Imaging Study

Thank you for your interest in this project. Your help in building an exciting health resource is very much appreciated. We hope you find the information on these web pages of interest, and look forward to seeing you at our assessment centre in Cheadle, Stockport soon ([click here for map](#)). If you have questions, or wish to make an appointment, please contact us on 0800 0 276 276 8am-7pm

Introduction

Introduction

Consent

Feedback

Eligibility

Find out more about the research →

Further documents

Improving the health of future generations

Go to the [UK Biobank website](#)

Incorporated in England and Wales, registered number 4978912, and registered as a charity in England and Wales, number 1101332. Charity registered in Scotland, number SC039230. © UK Biobank 2014

11:33
20/02/2014

UK Biobank: Centralised follow-up of health (but there is not a single system in the UK)

- Death and cancer registries
- In-patient and out-patient hospital episodes (including psychiatric) and related procedure registries
- Primary care records of health conditions, prescriptions, diagnostic tests and other investigations
- Other health-related: disease registries; dispensing; imaging; screening; dental; tax/benefit payments
- Direct to participants: self-reported medical conditions; treatments actually being taken; degree of functional impairment; cognitive and psychological scores

Category 100091

Health-related outcomes

Description

This category contains information related to the health outcomes of the participants provided through linkages to a range of health records. These datasets are continually being accrued and the database is updated on a regular basis.

3 Sub-Categories

1 Resource

Category ID	Description	Fields
2000	Hospital in-patient	127
100093	Death register	6
100092	Cancer register	8



Category 100093

Death register - Health-related outcomes

Description

This category contains coded data on the cause of death (International Classification of Diseases [ICD10]), obtained through linkage to national death registries.

6 Data-Fields

1 Parent Category

1 Resource

Field ID Description

40018	Source of death report
40000	Date of death
40007	Age at death
40001	Underlying (primary) cause of death: ICD10
40002	Contributory (secondary) causes of death: ICD10
40010	Description of cause of death



Participants	8,074
Item count	8,200
Stability	Accruing

Value Type	Categorical (single)
Item Type	Data
Strata	Primary

Sexed	Both sexes
Instances	Defined (2)
Array	No

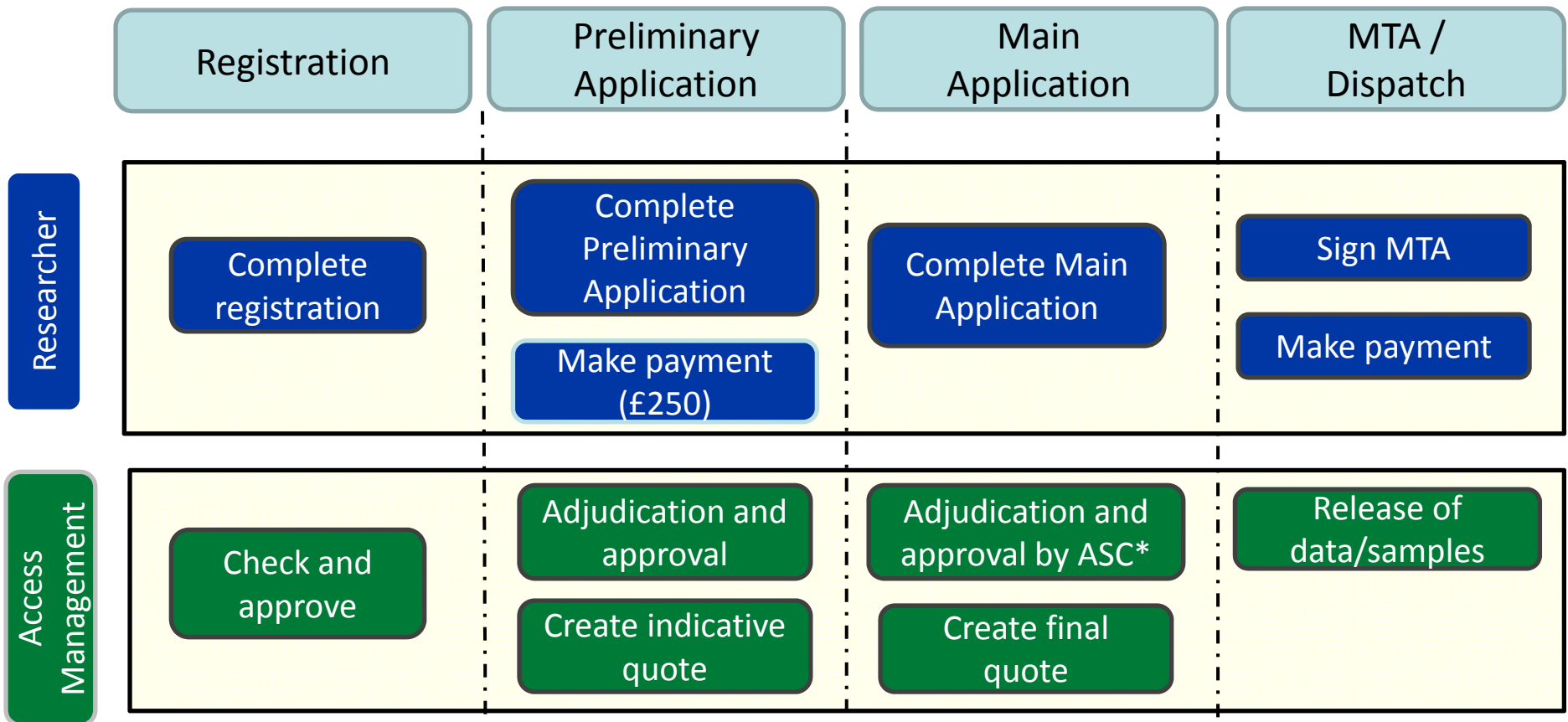
Data | **Notes** | **2 Categories** | **4 Related Data-Fields** | **1 Resource**

8,200 items of data are available, covering 8,074 participants, encoded using Data-Coding 19. Defined-instances run from 0 to 1, labelled using Instancing 9000001.

Category	Count
+ Chapter I Certain infectious and parasitic diseases	82
+ Chapter II Neoplasms	4865
+ Chapter III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	14
+ Chapter IV Endocrine, nutritional and metabolic diseases	58
+ Chapter V Mental and behavioural disorders	40
+ Chapter VI Diseases of the nervous system	209
+ Chapter IX Diseases of the circulatory system	-
+ I05-I09 Chronic rheumatic heart diseases	8
+ I10-I15 Hypertensive diseases	45
- I20-I25 Ischaemic heart diseases	-
+ I20 Angina pectoris	2
+ I21 Acute myocardial infarction	406
+ I24 Other acute ischaemic heart diseases	8
+ I25 Chronic ischaemic heart disease	595
+ I26-I28 Pulmonary heart disease and diseases of pulmonary circulation	49
+ I30-I52 Other forms of heart disease	163
+ I60-I69 Cerebrovascular diseases	296
+ I70-I79 Diseases of arteries, arterioles and capillaries	106

Cross-tabulations also available for registered researchers

UK Biobank access process: unduly complex and requires streamlining



*ASC (Access Sub-Committee): responsible for final approval of all access applications

Conditions of the Material Transfer Agreement

- Use data/samples only for the approved research project (but perhaps too restrictive of researchers' imagination)
- Restrict access to approved members of research team (but perhaps too cumbersome in collaborative research)
- Publish results and return derived data to UK Biobank (but storing such data may be costly and not useful)
- Potential actions for breaches of MTA:
 - Prevent further access by the researcher, and all other researchers based at the same Institution
 - Inform the researcher's Institution, funders and/or any relevant governing or regulatory bodies

Issues with access to cohorts established to support an extensive range of uses by different researchers

- Access process: may be made unduly complex due to concerns that proposed research uses may go beyond the original consent provided by the participants
- Depletable sample: a “resource” needs to be able to provide appropriate samples for a wide range of uses or, preferably, the results of a wide range of assays
- Data inaccessibility: researchers may not be able to handle complex data (e.g. from imaging) and, instead, need it to be converted into accessible “information”
- Insufficient specificity: lack of detailed characterisation of disease outcomes may either limit utility or result in delays before such information can be made available

Applications to use the UK Biobank resource, which has no preferential access, by country

	UK	US	Other	Overall
Registration	76%	7%	17%	1612
Preliminary application	81%	5%	14%	277
Main application	81%	5%	14%	149
Data release	80%	4%	16%	73

What might be the reasons for differential access?

Are there any known differences in the access procedures?
 does it's a key time to get to be a pie (at the level of)?