CATAPULT
Medicines Discovery

Data Access
Quick Guide to

biobank uk

April 2020

Medicines Discovery Catapult Quick Guide: UK Biobank
April 2020

UK Biobank is a national and international health resource following the health and well-being of 500,000 volunteer participants. It provides non-identifiable health information to approved researchers in the UK and overseas, from academia and industry.

The Cohort

The cohort includes around 500,000 women and men aged between 40 - 69 recruited from 2006 – 2010 across England, Scotland and Wales. All data is anonymised, and patients have consented for re-contact, but this can only happen with the permission of, and via, UK Biobank.

An overview of the data is available in the 'data showcase' on the UK Biobank website.

The data includes:

Initial assessment

- A touchscreen questionnaire and verbal interview collecting a range of health, lifestyle and sociodemographic measures.
- Physical measures including blood pressure, a hearing test, heel ultrasound, arterial stiffness test, eye measures, body composition, spirometry and an exercise test.
- Collection of biological samples including blood, saliva, and urine

Enhancements
A number of additional analyses are ongoing on the biological samples collected including:
- Biomarker analysis of blood and urine samples (see list)
- Infectious disease antibodies (currently for 10,000 participants)
- Metabolomics and telomere length are being measured by external groups, with data available within the next year

Genetic data

- Genotyping of 850k markers & imputation for all 500,000 volunteers.
- Exome sequencing on 50,000 UK Biobank participants

- Exome sequence data for the remaining 450,000 participants is ongoing with a further update of 100,000 participants in 2020. The data for the remaining 350,000 participants will be released in 2021
- A Whole Genome Sequencing pilot is ongoing – the first release for ~175,000 participants is anticipated for spring 2021.

Cohort-wide linked medical record data

- Death register (>20,000 deaths)
- Cancer registry (>120,000- site-specific cancer cases)
- In-patient hospital admissions which includes diagnoses and procedures
- Linkage to primary care data (currently 45% of participants). Linkage for the remaining 55% ongoing.

Subset data

- Repeat assessment for 20,000 participants
- Online questionnaires conducted annually (~330,000 ppts with email; ~50% response rates), including 24 hour diet, cognitive function, occupational history, mental health, digestive health and food preferences
- Objective physical activity data from wearable accelerometers for 100,000 people.
- Multimodal imaging assessment including abdomen, brain, heart MRI, carotid ultrasound results and whole body dual-energy X-ray absorptiometry of bones and joints. Data is updated bi-annually and currently available for ~40,000 participants. The aim is to have data for 100,000 participants by 2023

Applications for use

UK Biobank has an open and transparent data application and access management system to enable access to data on the cohort for health-related research in the public interest. Commercial companies and researchers are able to register and submit applications in the same way as academics.

The application process takes approximately 19 weeks to process if everything is approved.

UK Biobank is currently reviewing access procedures and Material Transfer Agreements and updated versions should be available during 2020. Improvements to the access process are helping UK Biobank move towards its goal of having a 12 week turnaround time by end of 2020.

During registration the identity and researcher status of the applicant are checked. Once the application is submitted it is assessed for rationale and feasibility of research and whether it is a) health-related and b) in the public interest.

For approved applications, the MTA must be executed, and access charges paid, before release of data and / or samples to the approved researcher.

Terms of access are identical for academic and commercial users.

Data only applications are rarely declined. The main criteria for data only applications are that the scope is clearly defined and that it is health-related research that is in the public good. Otherwise, applications can be broad in scope and do not require researchers to focus on single risk factors / diseases. Data only applications do not require peer review from subject specific experts, however they are reviewed by the scientific team (for application content) and the data analyst team (for data fields requested).

Sample use applications are more frequently declined as they have more stringent criteria to meet the use of this depletable resource.

Re-contact request applications have a high bar for approval. A recontact policy is available online and it is recommended that researchers should contact UK Biobank directly in the first instance to determine the feasibility of the recontact proposal.

Current Pricing

- £250 + VAT (where applicable) payable upon submission of an application
- £1,500 + VAT (where applicable) per application that requires access to data only
- £500 + VAT (where applicable) for access to any "bulk" data files (imaging, raw ECG data, raw HES data, genetic data, raw accelerometer data, built environment data)
- Bespoke costs for applications requesting access to samples, for re-contact requests or datasets requiring time-consuming customization

Medicines Discovery Catapult Quick Guide: UK Biobank
April 2020

Useful links

- [Detailed access procedures](#)
- [MTA describing terms of access](#)
- [Criteria for sample applications](#)

Usage

As of January 2020, 14,000 researchers have been approved. There have been around 1,800 approved applications submitted since opening for access in 2012 with more than 1,000 published papers.

The following rules and guidelines apply to use of the data:

- Users must not attempt to identify participants.
- Users may only work on the purpose described in their application - although change requests' can be submitted to extend the scope of the original application. These are reviewed by the scientific team.
- The development of commercial medicines and diagnostics is considered to be "health related" and "in the public interest".
- Users have full rights to inventions made using the data.
- Results (i.e. code, derived variables) must be returned 6 months after publication or 12 months after project completion, whichever comes first.
- There is [an outline of the results that researchers are expected to return on the UK Biobank website](#).
- Users may not use Intellectual Property Rights to unreasonably restrict, for example, naturally occurring genetic sequences / biomarkers / proteins / processes – in this case UK Biobank can force a license grant back to themselves.
- Users must update UK Biobank annually on progress through an annual update form and share a copy of any patents derived.

Data formats and platforms

Data delivery is via password-protected download. Individual level data can only be accessed by individuals listed as a collaborator on the application, although summary level data can be shared with anyone.

There are no resellers of UK Biobank data.