

UK Biobank: scale, depth, duration ... but, most importantly, accessibility

Ben Lacey

UK Biobank, University of Oxford

I have no conflicts of interests that need to be disclosed

WCE

WORLD CONGRESS OF EPIDEMIOLOGY 2024





Enabling scientific discoveries that improve human health

UK Biobank: scale, depth, duration ... but, most importantly, accessibility

Ben Lacey

UK Biobank, University of Oxford

Session: UK Biobank: democratising
health data for epidemiological research
globally

World Congress of Epidemiology

Cape Town, 25th Sept 2024

www.ukbiobank.ac.uk



What is UK Biobank?

Prospective cohort study of 500,000 adults, conducted to further understanding the genetic, physiological, lifestyle, and environmental determinants of disease

The study combines **scale**, **depth**, and **duration** of follow-up

But, most importantly, the data is made **readily accessible** to researchers globally



The graphic features a world map with several teal hexagonal icons connected by lines, representing global research and data access. The icons include a heart, a DNA double helix, a stethoscope, a pill, and a person. The UK Biobank logo is prominently displayed on the left. Below the logo, the text reads 'Research Analysis Platform' in pink, followed by the tagline 'Enabling scientific discoveries that improve human health'. A pink button labeled 'ACCESS PLATFORM' is positioned below the tagline. At the bottom left, there are links for 'SUPPORT | DOCUMENTATION | COMMUNITY' and a copyright notice: '© 2021 DNAnexus, Inc. All Rights Reserved'. At the bottom right, the logos for 'DNAnexus' and 'aws' are displayed.

biobank
Research Analysis Platform

Enabling scientific discoveries that improve human health

ACCESS PLATFORM

SUPPORT | DOCUMENTATION | COMMUNITY

© 2021 DNAnexus, Inc. All Rights Reserved

DNAnexus® aws

Overview of recruitment

- Recruited between 2006-2010
- 500,000 adults aged 40-69 years
- Registered with the NHS
- 22 assessment centre located around the UK to enhance heterogeneity of the cohort



Data collected at recruitment

Questionnaire and verbal interview

Wide-ranging: sociodemographic, lifestyle, personal and family medical history (cognitive function and hearing tests)

Physical measures

Blood pressure, anthropometry, carotid ultrasound, arterial stiffness, hand grip strength, heel bone ultrasound, spirometry; eye measures (2009 onwards)

Sample collection

Blood, urine and saliva; total >15M aliquots

Consent for linkage and re-contact

Consent for linkage to health-related records and for re-contact for further assessments



UK Biobank baseline characteristics

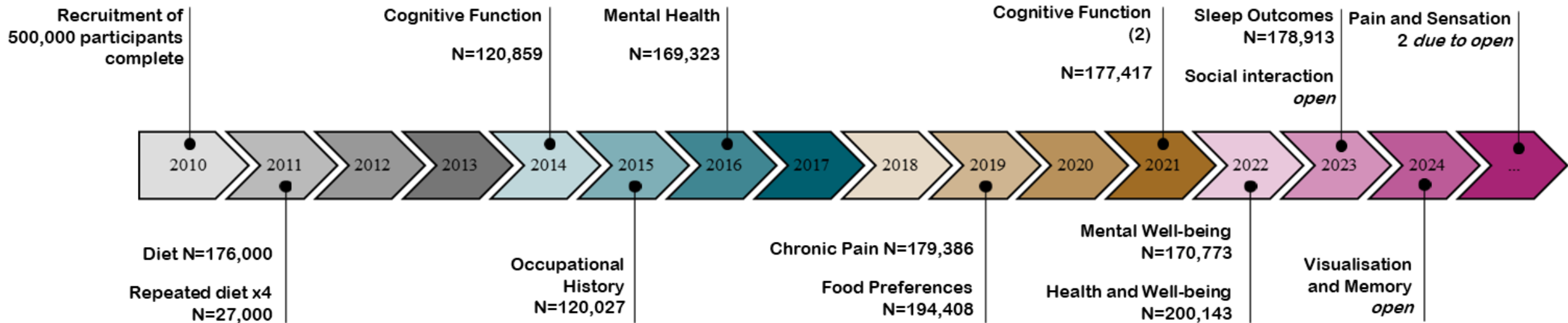


Characteristic	Category	Participants, n (%)
Age	40-49 yrs	119,000 (24%)
	50-59 yrs	168,000 (34%)
	60-69 yrs	213,000 (42%)
Sex	Male	230,000 (46%)
	Female	270,000 (54%)
Ethnicity	White	473,000 (95%)
	Other	27,000 (5%)
Deprivation	More	92,000 (18%)
	Average	166,000 (33%)
	Less	241,000 (46%)
Total		500,000

Enhancements: Web-based questionnaires

2011-ongoing

Among 330,000 participants for whom we have an email address



Enhancements: Physical activity monitor study

2013-16

- 100,000 participants; wrist-worn accelerometer for 7 days (2013-16)
- 2,500 people repeated 4 times over the course of 1 year (2018-19)



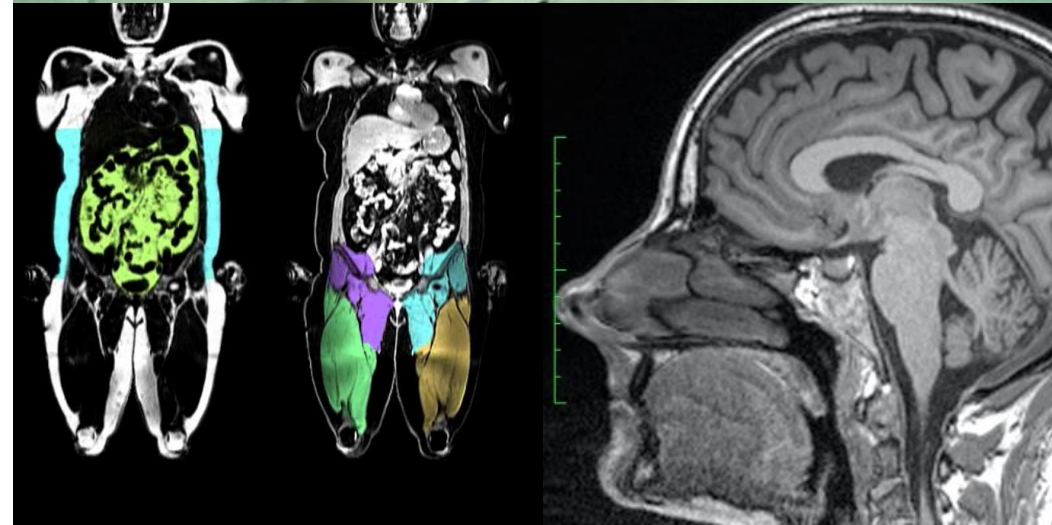
Enhancements: Resurvey and imaging

Repeat assessment (20k participants; 2012-13)

Multi-modal imaging (80k of 100k; 2014-)

- MRI (heart, brain, abdomen)
- Full-body DEXA
- Carotid ultrasound
- 12-lead ECG
- (2weeks of cardiac monitoring, 65+ years;2019-)

Repeat imaging underway (target of 60k; 2019-)



Whole cohort assay data already available

- Genotyping, whole exome sequencing, whole genome sequencing
- Haematology and biochemistry markers
- Leukocyte telomere length

Subset assay data available

- NMR metabolomics (~275,000 participants)
- Olink 3,000 proteins (~55,000 participants)
- Infectious markers (~10,000 participants)



Follow-up of health outcomes

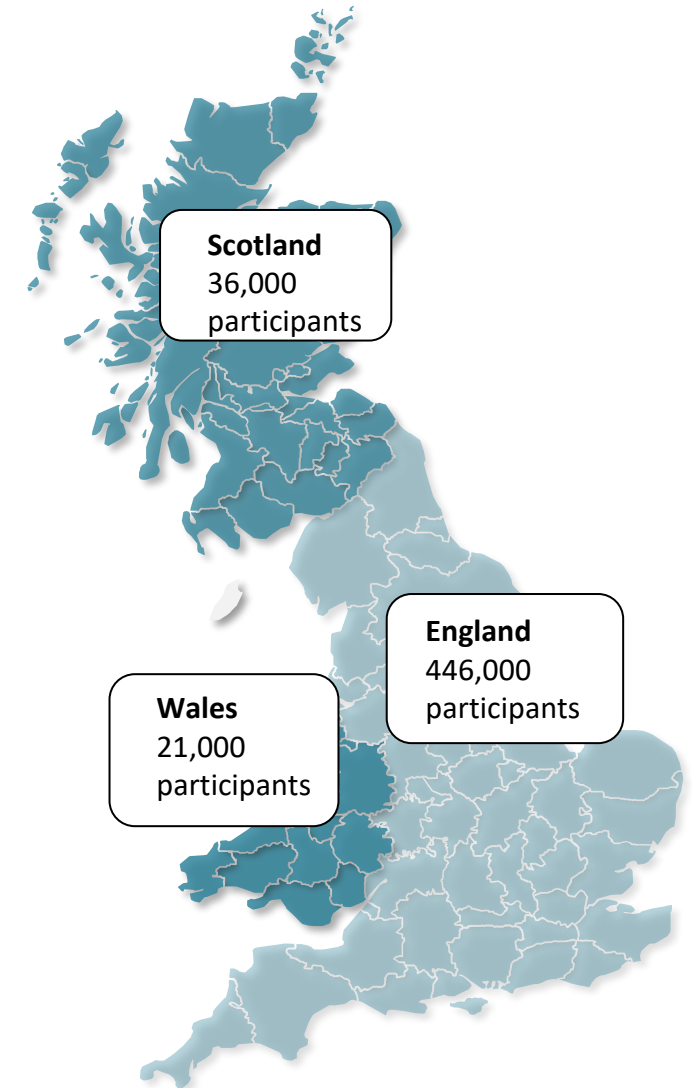
Regularly updated information on a wide range of diseases from NHS and other datasets in all 3 countries:

Deaths (2006-2022; causes of death)

Cancers (1957-2020; cancer code, stage and grade)

Hospitalisations (1997-2022; diagnoses and procedures)

Primary care (~45% of participants, 1938-2017; diagnoses, prescription and test results)



Cumulative number of incident case over time

Condition	Year		
	2025	2030	2035
Diabetes	35,000	40,000	45,000
Heart attack	20,000	30,000	45,000
Stroke	15,000	20,000	35,000
Breast cancer	18,000	22,000	25,000
Prostate cancer	15,000	25,000	40,000
Colorectal cancer	8,000	11,000	13,000
Lung cancer	5,000	8,000	10,000
Alzheimer's disease	10,000	20,000	40,000
Parkinson's disease	6,000	15,000	40,000

Potential for further enhancements



Enhanced risk factor characterisation

Extended assays conducted on stored biological samples

Repeat assessments on the surviving participants

First-degree relative recruitment

Enhancing health outcome characterisation



Extended linkages such as cancer histopathology

Specific assessments of participants with particular conditions (e.g. dementia)

Develop novel measures of health over time (apps)

Readily accessible to scientists worldwide

Registration: rapid central check of the *bona fides* of researchers

Data access: Streamlined process for requesting data to conduct any type of health-related research that is in the public interest

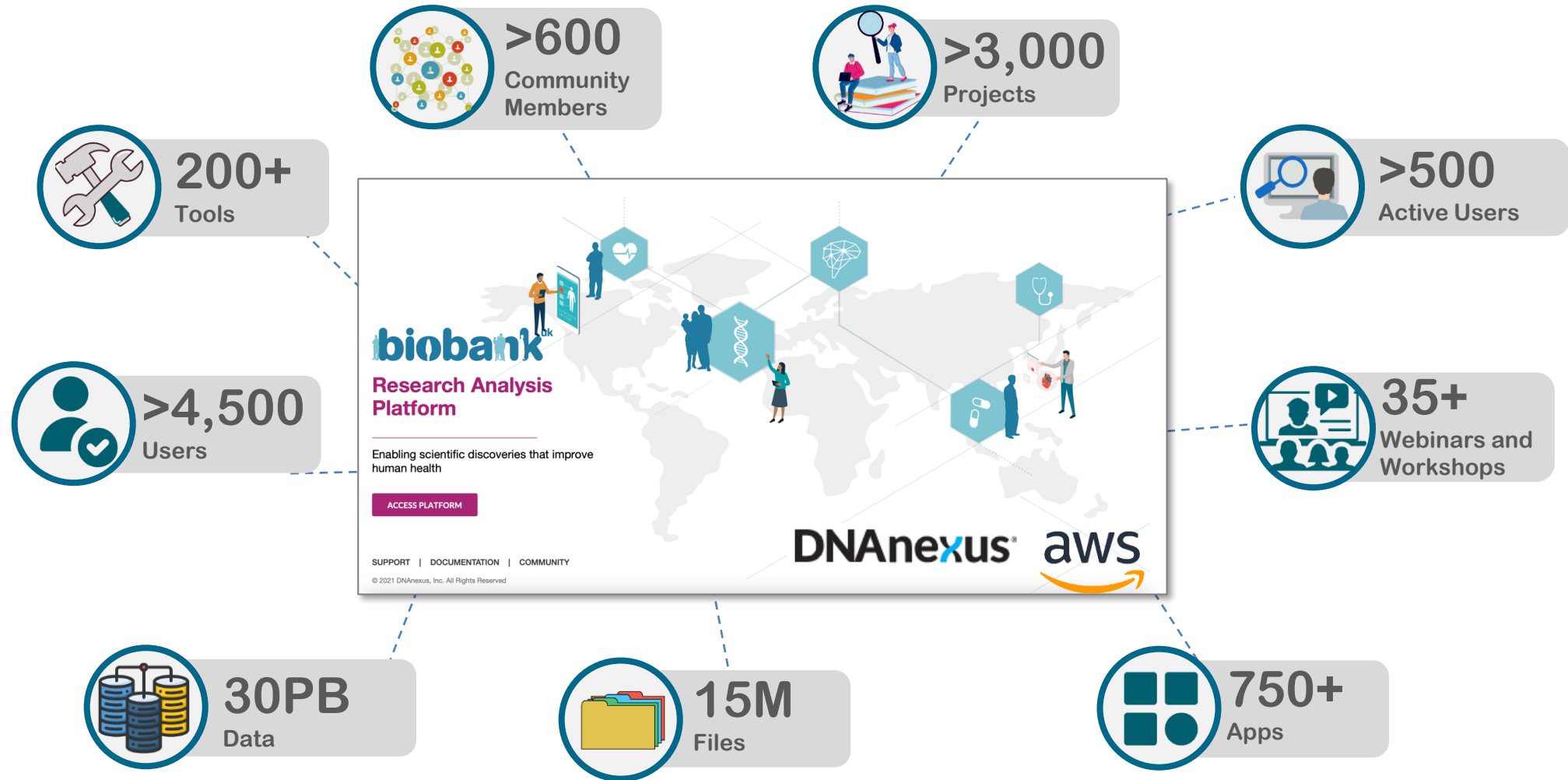
Sample access: Requirement for well-validated assays that use limited sample volume from whole cohort, or substantial subset

Standard MTA: agreement requiring data be kept secure (and now transitioning towards conducting analyses on our secure Research Analysis Platform)

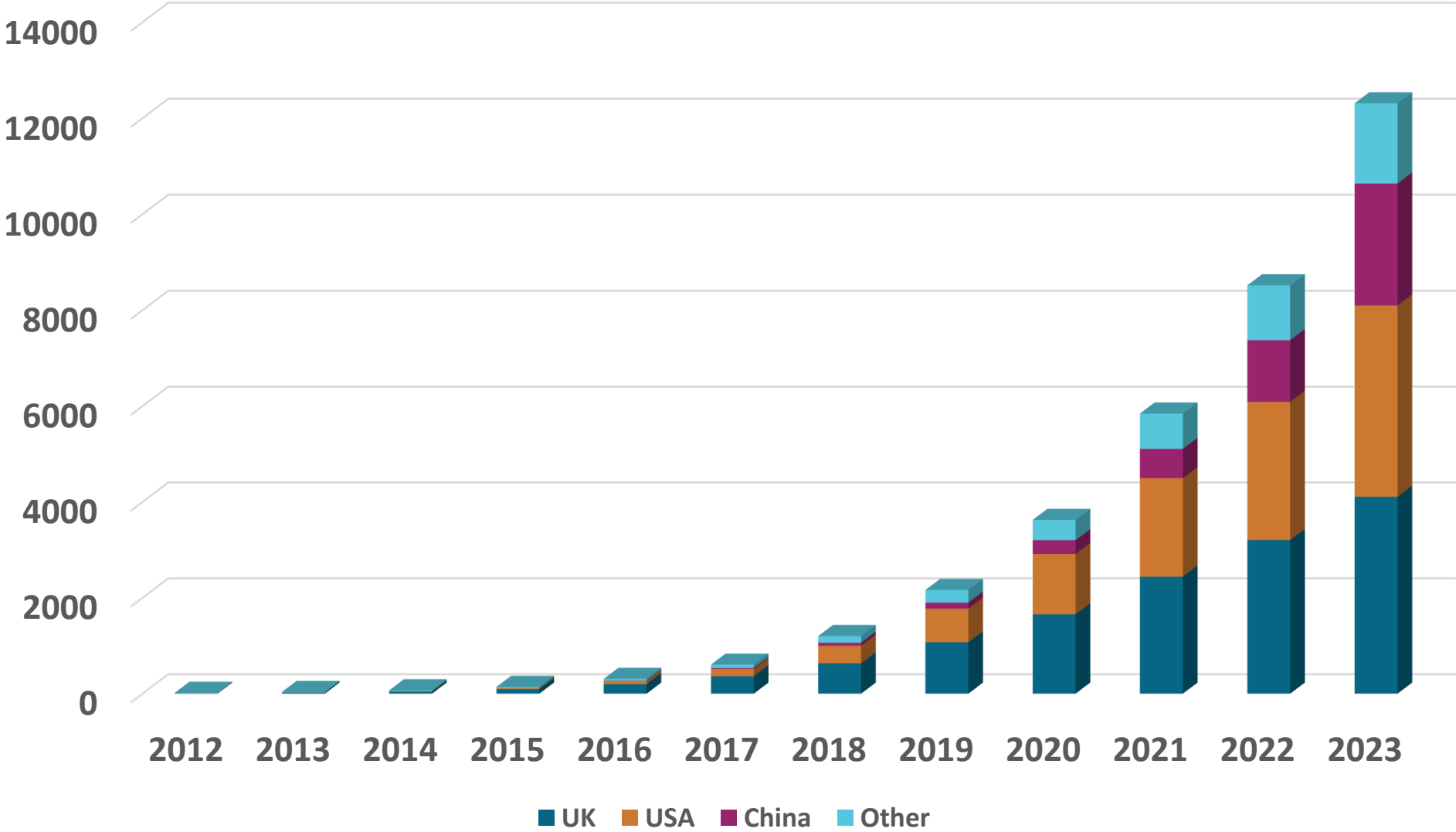


Fees: depending on 'tier' of data, 3k-9k + VAT for 3 years access

Democratising access to the data



Rapidly increasing number of publications



Access Fund and UKB Platform Credits

Student researchers from any country are eligible for reduced **access fees** (£500+VAT).

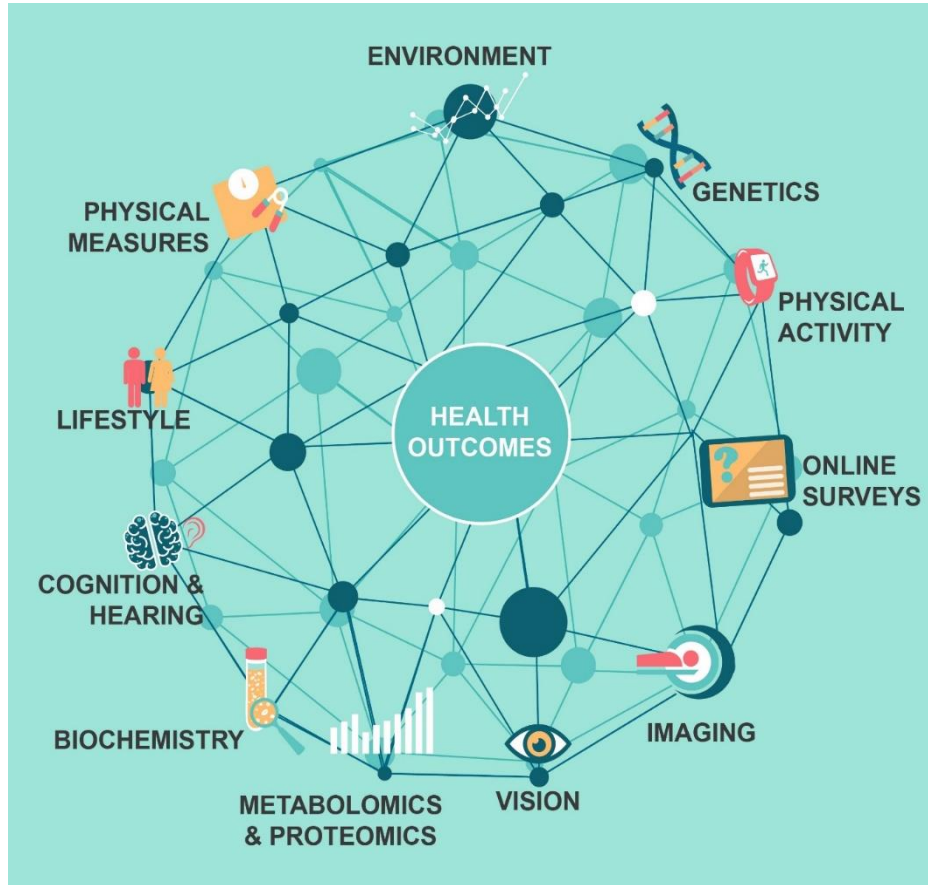
Approved researchers from less-well-resourced countries can have their **access fees** paid for by UK Biobank's Global Researcher Access Fund.

Students, early career researchers and researchers from less-well-resourced countries can also receive support with **computing costs** associated with UK Biobank's RAP.

Funded researchers can use the entire UK Biobank data set on 500,000 participants. For further information please contact UK Biobank on researchercommunications@ukbiobank.ac.uk



Conclusion



- **SIZE:** 500,000 adults
- **DEPTH:** exposure and health outcomes
- **DURATION:** ~15 years of follow-up
- **ACCESSIBILITY:** global community of researchers using the data, facilitated by the UKB Research Analysis Platform

Combination of size with increasing depth x duration x accessibility enabling cutting-edge science

Acknowledgements



Medical
Research
Council



CANCER
RESEARCH
UK



British Heart
Foundation

NIHR | National Institute for
Health and Care Research



REGENERON
SCIENCE TO MEDICINE®



GSK

AstraZeneca

abbvie



Bristol Myers
Squibb™



Department for
Science, Innovation,
& Technology



CITADEL



novo nordisk®

AMGEN

Chan
Zuckerberg
Initiative



Genentech
A Member of the Roche Group

UK Biobank team

- Executive team
- Coordinating centre staff
- Imaging staff
- Steering Committee
- International Scientific Advisory Board
- Scientific Working Groups
- Oxford University team

Our 500k participants

