Title: Scientific and clinical impacts of UK Biobank in cardiovascular medicine

Short title: UK Biobank and cardiovascular medicine

Manuscript reference: CIRCULATIONAHA/2024/072449

Authors: [1,2]Adam J. Lewandowski, DPhil; [2,3,4]Martin K. Rutter, MD; [1,2]Rory Collins, FRS FMedSci

Affiliations:

[1]Nuffield Department of Population Health, University of Oxford, Oxford, UK

[2]UK Biobank Ltd, Stockport, Greater Manchester, UK

[3]Division of Endocrinology, Diabetes & Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

[4]Diabetes, Endocrinology and Metabolism Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

Corresponding author: Professor Sir Rory Collins, Richard Doll Building, Roosevelt Drive, Old Road Campus, Oxford, UK. OX3 7DG. Email: rory.collins@ndph.ox.ac.uk

Word Count: 1150 words

Benefits of UK Biobank to cardiovascular science

Cardiovascular diseases account for ~30% of deaths worldwide and remain a leading cause of morbidity and mortality globally, driving the need for ongoing major research investments. These diseases often arise from a complex interplay of genetic factors, lifestyle choices, and environmental exposures, each of which may only have a small to moderate individual impact, making them difficult to study comprehensively. In this context, UK Biobank offers unique opportunities as a large prospective cohort study, with detailed, extensive longitudinal data facilitating research into the determinants of cardiovascular diseases.[1] Its open access model for both academic and commercial researchers worldwide has already resulted in ~4,200 cardiovascular-related publications.

Study design and enhancements

Between 2006-2010, UK Biobank enrolled 500,000 men and women aged 40-69 from across the UK. Baseline data collection included biological samples; information on lifestyle and environmental exposures; family and personal medical history; physical and physiological measures; and consent to recontact participants for further assessments and permission to access their health records.

UK Biobank has since undergone major enhancements (Figure 1), including additional clinical assessments, new and repeated questionnaires, and extended assays on stored samples enabled by substantial academic and commercial partnerships. By 2023, genotyping, whole exome sequencing, and whole-genome sequencing data for all half million participants were available to researchers. These enhancements have facilitated the creation of robust genetic instruments for Mendelian randomization studies, which help to elucidate causal relationships and overcome the main limitations of traditional epidemiology studies (such as confounding and reverse causality).

The extensive genetic data has also enabled development and validation of genome-wide polygenic risk scores for diseases and quantitative traits, which were previously unattainable due to the limited sizes of

earlier genome-wide association studies (GWAS). Polygenic risk scores for common conditions including coronary artery disease, atrial fibrillation, and type 2 diabetes have been developed using UK Biobank data. These scores have been shown to identify substantially larger portions of the population at comparable or greater disease risk than those identified through having rare monogenic mutations.[2] To date, polygenic risk scores for around 50 diseases and quantitative traits, many originally derived using UK Biobank GWAS data, are available to the research community through the UK Biobank resource and are starting to be incorporated into routine healthcare.

In 2014, UK Biobank launched the world's largest multi-organ imaging study, with 100,000 of the 500,000 UK Biobank participants to undergo comprehensive imaging assessments. This includes cardiac, brain, and abdominal magnetic resonance imaging (MRI), dual-energy X-ray absorptiometry, 12-lead electrocardiogram, and carotid ultrasound, as well as repeat baseline assessment measurements. In 2023, UK Biobank initiated a repeat imaging study, aiming to re-image 60,000 of these participants between two and nine years after their initial scans. This effort seeks to document prospective phenotypic changes, their relevance to disease onset, and their lifestyle and genetic determinants. The MRI scans have already been used to derive population reference ranges, automated segmentation algorithms, and quality control pipelines.[3] This has led to creation of image-derived phenotypes available to the research community and development of commercial image analysis software used in clinical practice. Furthermore, combining imaging data with clinical and genetic data from UK Biobank participants has enabled development of clinical and genetic risk prediction tools. For instance, Pirruccello *et al.* used UK Biobank cardiovascular magnetic resonance (CMR) images and deep learning models to determine aortic size, identify participants with thoracic aortic aneurysms, and demonstrate disease prediction using polygenic risk scores. The authors went on to use clinical data and UK Biobank CMR images to develop and validate a robust clinical risk score for thoracic aortic aneurysms, which was

subsequently externally validated using echocardiography and computed tomography[4] and will be trialed clinically.

As the resource has grown, the open access model has accelerated health-related discoveries. While this may sometimes result in duplicated efforts, it offers valuable opportunities for replication studies and fosters greater collaboration. The increasing size of the resource, along with changes to data security and governance, have prompted the creation of the *Research Analysis Platform* (RAP); a cloud-based platform that will enhance accessibility to UK Biobank data for the growing global research community. Major efforts are underway to ensure that the RAP's functionality meets diverse researcher needs. For certain groups of individuals, including minority ethnic groups, there are too few UK Biobank participants to reliably examine exposure-disease associations.[1] However, UK Biobank findings can be compared and validated in other large, longitudinal cohorts where these populations are not minorities, such as the *Mexico City Prospective Study* and *China Kadoorie Biobank*.[5]

Future directions and opportunities

The scientific and clinical value of UK Biobank data continues to increase over time. By 2020, there were 31,000 observed cases of diabetes, 15,000 cases of myocardial infarction, and 12,000 cases of stroke among UK Biobank participants. It is projected the number of incident cases for these conditions will roughly double by 2027, with cases of myocardial infarction and stroke expected to triple by 2032.[1] There is strong scientific value in collecting additional data on these participants nearer to the time of disease emergence, recognizing that the cohort, now in their early 70s on average, is at an optimal age for assessing determinants of age-related health and disease. To capitalize on this opportunity, we are planning a cohort-wide repeat assessment of the baseline measures. This will enable researchers to study 20-year changes in risk factors and their prediction of subsequent diseases, as well as associations between baseline exposures and changes in functional measures. This initiative also provides an

opportunity to introduce new clinic assessments, collect additional biological samples, and extend data capture of physiological and behavioral parameters using digital health technologies.

Enhancing health outcome characterization for UK Biobank participants is a key priority, with plans to link the full cohort to primary care data, national audits, registries, and prescribing records, along with imaging and digital pathology data for specific conditions. However, heterogeneity of underlying pathology for some clinical conditions and inadequate capture of disease subtypes in health records are barriers to understanding pathophysiology and developing new clinical interventions. These issues have been identified as particular challenges for studying neurodegenerative diseases. To address this, we have designed a *Brain Health Study* to be offered to UK Biobank participants who show symptoms of, or have been diagnosed with, a neurodegenerative disease. This assessment will include comprehensive clinical phenotyping, brain magnetic resonance imaging, blood testing, and data collection using digital health technologies. Learnings from this approach could then be applied to other heterogeneous diseases that are difficult to characterize from health records alone, such as distinguishing heart failure subtypes.

Conclusions

The scale, depth, and accessibility of UK Biobank data collected over nearly two decades make it one of the most unique and valuable biomedical resources for health-related discovery science. The groundbreaking knowledge generated from UK Biobank data is expected to lead to transformative advances in the prevention and treatment of cardiovascular diseases and other common conditions, benefiting populations worldwide.

References

1. Allen NE, Lacey B, Lawlor DA, Pell JP, Gallacher J, Smeeth L, Elliott P, Matthews PM, Lyons RA, Whetton AD, Lucassen A, Hurles ME, Chapman M, Roddam AW, Fitzpatrick NK, Hansell AL, Hardy R, Marioni RE, O'Donnell VB, Williams J, Lindgren CM, Effingham M, Sellors J, Danesh J, Collins R. Prospective study design and data analysis in UK Biobank. *Sci Transl Med*. 2024;16:eadf4428. doi: 10.1126/scitranslmed.adf4428

2. Khera A V, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, Natarajan P, Lander ES, Lubitz SA, Ellinor PT, Kathiresan S. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50:1219–1224. doi: 10.1038/s41588-018-0183-z

3. Raisi-Estabragh Z, Harvey NC, Neubauer S, Petersen SE. Cardiovascular magnetic resonance imaging in the UK Biobank: a major international health research resource. *Eur Heart J Cardiovasc Imaging*. 2021;22:251–258. doi: 10.1093/ehjci/jeaa297

4. Pirruccello JP, Lin H, Khurshid S, Nekoui M, Weng L-C, Vasan RS, Isselbacher EM, Benjamin EJ, Lubitz SA, Lindsay ME, Ellinor PT. Development of a Prediction Model for Ascending Aortic Diameter Among Asymptomatic Individuals. *JAMA*. 2022;328:1935–1944. Available from: https://doi.org/10.1001/jama.2022.19701

5. Collins R, Balaconis MK, Brunak S, Chen Z, De Silva M, Gaziano JM, Ginsburg GS, Jha P, Kuri P, Metspalu A, Mulder N, Risch N. Global priorities for large-scale biomarker-based prospective cohorts. *Cell Genomics*. 2022;2. doi: 10.1016/j.xgen.2022.100141

Funding Sources

UK Biobank has core funding from the MRC, Wellcome, British Heart Foundation, Cancer Research UK, and the National Institute for Health Research.

Disclosure Statement

A.J. Lewandowski and M.K. Rutter are members of the UK Biobank scientific team. R. Collins is the UK Biobank Principal Investigator and Chief Executive.

Acknowledgements

Special thanks to A. Motley for constructing Figure 1. We sincerely thank the 500,000 participants in the UK Biobank study for their remarkable generosity, selflessness, and continued engagement, support, and commitment.

Figure legends

Figure 1. Cardiovascular data available in UK Biobank. Shown are data collected at baseline and follow-up clinical assessments, including the imaging and repeat imaging studies, as well as data captured from health records and questionnaires. Further details on all available data can be found at https://biobank.ndph.ox.ac.uk/showcase/. *Primary care data are currently only available for 45% of the UK Biobank cohort for general research purposes (which represents complete coverage from one primary care system supplier, up to 2016/2017). EKG stands for electrocardiogram; MRI, magnetic resonance imaging; NMR, nuclear magnetic resonance.