UK Biobank: obstacles to obtaining coded primary care data despite the explicit consent of all 500,000 participants

Background to UK Biobank: a prospective cohort of 500,000 consented participants

UK Biobank was established as a charity by the Wellcome Trust and Medical Research Council in 2003. It continues to be funded by them along with the British Heart Foundation, Cancer Research UK and the National Institute of Health Research. The UK Biobank project has approval from the North West Multicentre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB).

The UK Biobank cohort involves just over 500,000 men and women who were aged 40-69 years when they were invited to participate in 2006-10, with the explicit support of the NHS (which provided contact details, as well as NHS numbers to help ensure correct linkage to health outcome data). All potential participants were provided with a detailed information leaflet that explained what was involved before their explicit consent was sought. This included agreement for *"permission for access to my medical and other health-related records, and for long-term storage and use of this and other information about me, for health-related research purposes (even after my incapacity or death)"*.

Since recruitment, all participants have been receiving regular updates about UK Biobank, including:

- communications about particular initiatives (such as obtaining access to their medical records);
- being asked to provide additional information (e.g. completing web-questionnaires, wearing activity monitors, and attending imaging assessments);
- findings emerging from research conducted using the resource via annual newsletters; and
- being invited to in-person and virtual meetings about UK Biobank.

Any participant can withdraw from UK Biobank at any time – and they are regularly reminded of this – but fewer than 1,000 have done so to date.

Use, storage and sharing of participant data

UK Biobank has been receiving information about deaths, cancers, hospitalisations and other health-related information about the consented participants from central NHS systems in England, Scotland and Wales for many years. These data are held securely in UK Biobank's database (with participant identifiers separated from all of the other data) behind firewalls which are subject to regular external penetration testing (last conducted September 2022 with no material issues identified). The NHS also audits the systems and processes: for example NHS Digital's Data Sharing Audit gave UK Biobank a clean bill of health in Q3 2021 (full report available here). UK Biobank makes continual investments in information security and is certified to the internationally recognised ISO27001 standard.

Researchers who wish to access UK Biobank data or samples must provide evidence of their research credentials. Applications for access are reviewed, and any that require further consideration are brought to the attention of the Access Subcommittee of UK Biobank's Board, which includes external scientific members and ethics oversight. De-identified data or samples for an approved application are made available only after UK Biobank's Material Transfer Agreement has been signed by the applicant's institution. Annual reports are required on the research that has been conducted and any findings are required to be published, with the underlying results returned to UK Biobank.

Research value of primary care data

Health conditions that are managed largely outside of hospital by general practitioners (such as arthritis and other causes of pain, dementia and other neurodegenerative conditions, impaired vision or hearing, respiratory conditions, heart failure and mental health problems) have been systematically under-represented in large-scale epidemiological studies. Consequently, securing access to the coded primary care data for the participants in UK Biobank offers an unparalleled opportunity to redress this imbalance, enabling greater understanding of ways to prevent and treat a wide range of conditions. For example, the inclusion of primary care data would result in an approximate doubling of cases of

depression and dementia that would be identified (see Figure), as well as allowing less severe cases to be detected at an earlier stage (with the ability to identify cases across the full spectrum of disease severity furthering our understanding of disease progression).

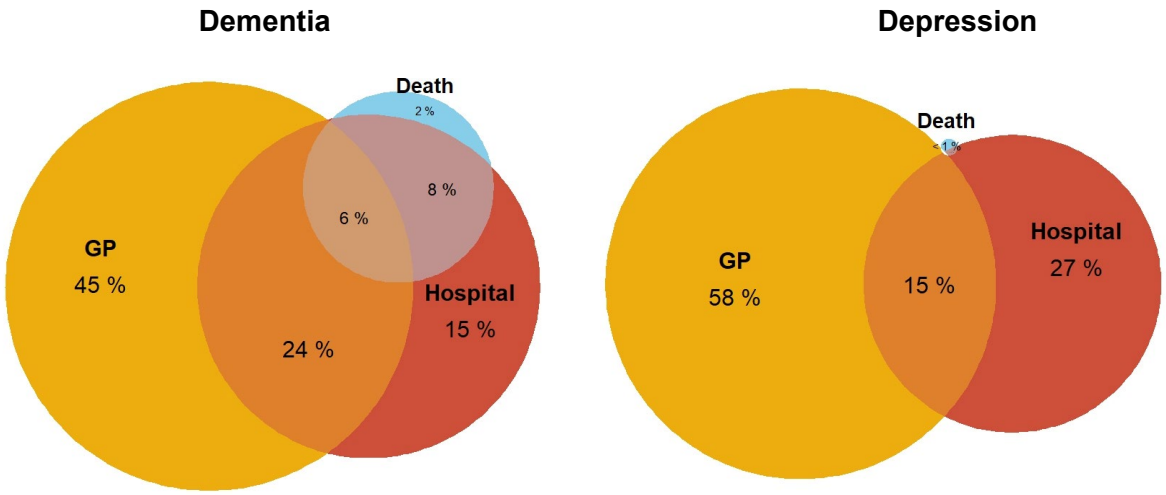

Figure: Exemplar health outcomes for which primary care data dramatically increase the numbers of cases that can be detected compared with other sources of healthcare data

Pre-pandemic approaches to obtaining coded primary care data for UK Biobank (see link)

Coded primary care data (which contains data on diagnoses, prescriptions, referrals, laboratory test results, but not any free-text or attachments) are available to UK Biobank for participants in Scotland and Wales through central systems. For the participants registered with General Practices in England, a series of efforts have been made to obtain the coded primary care data:

- Coded primary care data were obtained in 2017 from TPP, a provider of GP practice management systems (which acts as data processor on behalf of the individual GP practices as data controllers). A letter was first sent to all TPP Practices, endorsed by the RCGP, setting out the basis for the proposed extraction, with Practices freely able to opt out (only one chose to do so).
- Following a request from EMIS (the other main GP system provider), which was prompted by the intervention of the RCGP and BMA Joint GP IT Committee (JGPITC), UK Biobank sought the advice of the Information Commissioner's Office, who confirmed in writing the legitimacy of the process.
- UK Biobank then met with the JGPITC, whose recommendation was that UK Biobank seek explicit agreement from General Practitioners to release data for participants registered at their Practice. Two pilot studies were run in 2018/2019, with the materials jointly agreed between the JGPITC and UK Biobank. The first study involved 600 GP Practices, and the follow-up pilot involved 100 GP practices, with considerable efforts made to make the process straightforward by providing each Practice with:
  - a list of consented participants registered at their Practice;
  - access to copies of those participants' signed consent forms;
  - a proforma completed Data Protection Impact Assessment [DPIA]; and
  - an offer to cover the Practice's costs for dealing with the request for the data.
- Despite JGPITC endorsement of the process, the percentage of Practices that agreed to the data release never got above 20%. The main reason is uncertain but, as well as being busy and having other priorities, it may well be due to Practice concerns about their data sharing responsibilities.

Obtaining coded primary care data during the pandemic

During the pandemic, the Secretary of State issued a Control of Patient Information (COPI) notice to GP Practices (as data controllers) and to EMIS and TPP (as data processors). Only a small number of Practices sought further information from UK Biobank (which was provided) and 100% acceptance was achieved, with secure provision of the coded data for all consented participants.

The ability to combine the coded primary care data with the other sources of health outcome data (e.g. hospitalisation, cancer and death records) demonstrated their value for COVID-19 research. Over 200 papers have been published using these linked healthcare records, many of which have used the primary care data to investigate the role of co-morbidities and medications as determinants of severe COVID-19 and to identify ways to mitigate its impact.

The COPI notice expired in June 2022. In October 2022, UK Biobank met again with the JGPITC to describe the previous lack of success of the targeted GP 'opt-in' approach, and to discuss whether a central solution (based on the success of the COPI notice) could be implemented that would avoid General Practitioners having to take responsibility for the data sharing decision.

Following that meeting, UK Biobank met with the Chair of the RCGP and the President of the BMA (along with their policy advisors) to seek support for a central instruction. It was agreed that one more pilot study among 100 randomly selected Practices would be performed, using a more streamlined request with endorsement from both the RCGP and the BMA. It achieved a success rate of about 10%.

NHS England-coordinated mailing to General Practitioners requesting the primary care data

Following the failure of the third pilot study, UK Biobank engaged with the National Director of Transformation at NHS England (NHSE) and his colleagues, who proposed that a further request should be sent to all Practices in England directly through the NHS systems. A revised draft letter was created with input from the NHSE Communications team, as well as input from the RCGP Chair and BMA President. It was suggested that we also seek support from the BMA's GP Committee.

The new Chair of the BMA GP Committee indicated that the BMA was not able to consider signing the letter to GPs at that time due to clinical commitments, so the BMA logo and signature were removed before NHSE sent it to all Practices. However, because a single mention of BMA support was left in the text (albeit the current and immediate past BMA Presidents had provided such support), a revised letter was sent the next day at the request of the GP Committee Chair.

The response to this request for coded data is being monitored via the EMIS and TPP systems providers. Anecdotal evidence indicates that many GPs would be supportive of the data being made available to UK Biobank, but that many of them have either not seen the request and/or are being actively discouraged by some Integrated Care Boards because of misinformation about the process. As such, it is highly unlikely that this approach will be successful.

In parallel, UK Biobank has been informing all participants specifically about the request that is being made to their GPs for access to their coded primary care data. As was found during the focus groups run by UK Biobank (through its Ethics Advisory Committee) over the past year, there is surprise and disappointment that these data are not already available (with fewer than 0.1% of participants raising concerns). This confirms the persistence of the participants' consent for provision of their data.

Proposal that NHSE issue a statutory notice for UK Biobank

The UK Biobank data are being widely used to help improve disease prevention and treatment. More than 9,000 peer-reviewed papers have been published based on access to these data, with over 2,000 published in 2022 alone. Many of these findings have had direct implications for the NHS in identifying ways to improve the prevention and treatment of a wide range of conditions. Adding the coded primary care data to all the other health outcome data already made available by the NHS would increase UK Biobank's capacity to support research that improves population health and patient care.

Following the termination of UK Biobank's COPI notice for COVID-related research, no further coded primary care data are being made available for the approximately 450,000 consented participants registered in General Practices in England. Practices have been reminded by their system suppliers to turn off the data sharing agreement for UK Biobank, and no further data are currently being sought.

Access to primary care data of 59 million non-consented patients for COVID-related research is still occurring within OpenSAFELY following the issue of a Data Provision Notice (DPN). It is understood that the DPN may be extended for research on all types of health conditions. This is a welcome development as OpenSAFELY allows some types of analyses to be conducted that are of enormous value for improving health care delivery. However OpenSAFELY is not able to address the large majority of the needs of consented cohorts (as detailed elsewhere).

Consequently, UK Biobank recommends:

- **Given the consistent lack of success in obtaining agreement from individual General Practices (despite guidance and support from the RCGP, BMA and JGPITC over a long period of time), and the demonstrated success and acceptability of the COPI notice, a suitable central instruction for the release of coded primary care data is needed for UK Biobank;**
- **Such a central instruction from NHSE could come in the form of an extension of the COPI notice, the issue of DPN (as with OpenSAFELY) or some other suitable instrument. It would provide a clear and unambiguous instruction to General Practices, which they could be appropriately reassured had been properly considered and reviewed;**
- **General Practices should also be reassured that they would have no liability or exposure related to such an instruction, which could come through a statutory confirmation or an indemnity from UK Biobank (which it is willing to provide).**

UK Biobank would welcome the opportunity to discuss this proposal jointly with representatives of NHSE and other relevant parties (including the RCGP, BMA and JGPITC) in order to find a way to honour the wishes of the UK Biobank participants for their health data to be used for research.