



## Biobanks



Biobanks are repositories of biological samples, physical measurements and outcome and behavioural data collected for research, clinical practice or public health monitoring. Most biobanks are established to investigate the common determinants of mortality and morbidity. This briefing updates POSTnote 180 and summarises biobank activity in the UK and the legal, ethical and practical issues.

### Background

Biobanks promise to facilitate major advances in health research. They take advantage of state-of-the-art genetics together with big data sets and individual health records to allow for complex and powerful studies. The UK offers a world-leading environment for the creation of biobanks, which has seen the establishment of UK Biobank (Box 1) and several others that have collected data from various populations (Boxes 2 and 3).

Biobanks are key in the mission to understand the genetic, environmental and lifestyle factors that cause disease.<sup>1</sup> They are a resource for storing and analysing data and samples. They collate large numbers of variables which can include biological specimens (such as tissue, blood, urine and saliva), physical measurements (such as age, height, weight and blood pressure), information obtained from behavioural questionnaires (such as physical activity and smoking habits) and personal information (such as address, employment history and medical records). This allows researchers to analyse any interactions between these factors and an individual's genetic susceptibility to disease risk.<sup>2</sup>

### Overview

- Biobanks are seen as a critical resource for research into the genetic and environmental causes of disease.
- Ethical issues pertinent to biobanks include consent and feedback for participants.
- The model of broad consent for biobanks may become incompatible with proposed changes to data protection law, if they are adopted by the European Parliament.
- Significant benefits for research could be derived from greater linkage of data from NHS records into biobanks.
- Increased data sharing between biobanks is seen as key to getting the most from scientific research.
- Participant and public engagement with biobanks is necessary for public trust but may be eroded by commercialisation or government use of personal data.

### Types of Biobank

Biobanks vary in size and scale, but can be divided into two broad categories: large cohort studies and disease specific.

#### *Large Prospective Cohort Studies*

These follow a large group of people over time and monitor who develops what diseases and when (Boxes 1 and 2). This approach can be used to link lifestyle and biological characteristics (such as biomarkers or genetic factors) with risks of developing a wide range of diseases. A minimum of tens of thousands of participants is required, as only a small proportion will develop any one condition over time and the effects of different factors on the development of most conditions are likely to be modest. These studies are common in the UK, with an estimated one in 30 people taking part in a study.

#### *Disease-Specific Biobanks*

These focus on one disease, tend to be smaller (Box 3) and recruit people affected by the particular condition. Research focuses on causes and the effectiveness of treatments. They can complement larger prospective studies.

Research using biobanks faces tension between the benefit to medical innovation and the potential to improve public health on the one hand, and an individual's right to privacy and the need to gain consent on the other. This tension includes issues of participant recruitment and feedback, the sharing of data between biobanks and the NHS as well as commercialisation of biobanks.

### **Consent and EU Data Protection Law**

The collection, storage and use of data held in European biobanks are governed by national laws based on the EU Data Protection Directive. A new data protection regulation has been proposed by the EU to update the law, which includes exemptions for medical research with certain safeguards. However, the existing research exemption from consent would be restricted if the European Parliament adopts amendments to Articles 81 and 83 of the Data Protection Regulation.<sup>3</sup>

Some UK researchers and funders argue that if these proposals go ahead, health and scientific research will be impeded. They anticipate that the amendments would make it difficult for personal health data to be used in research (such as in biobanks) without specific consent for each research activity, although participants would have already given their broad consent for such activity. If the amendments are approved, participants may need to be re-contacted for consent to individual research projects that use their data. Researchers consider that this would cause problems by creating consent fatigue and increasing the administrative component of research.

### **Feedback for Research Participants**

A widely held view is that participants donate samples to a study as an act of altruism for the public good. However, surveys have shown many people participate with the expectation of receiving information about their health status. There is no statutory obligation for UK researchers to discuss any findings with individuals. Conventionally, biobanks inform participants that any information resulting from research conducted with their data will not be discussed with them. The rationale for this approach is that in some instances the outputs from biobank research may not meet clinical standards for diagnosis.

However, another view is that participants are entitled to all information generated from analysis of their data, and participants have voiced to UK research funders that they are keen to have such findings returned to them. Offering such feedback would incur additional costs for research projects and may not be clinically relevant for the participant. Furthermore, there may be implications for participants' family members. An MRC and Wellcome Trust framework does not include a unilateral approach to participant feedback but concludes that it is best to treat each research project individually.<sup>4</sup>

### **Box 1. UK Biobank**

UK Biobank is a large scale prospective cohort study and a world leading resource for researchers. It is designed to allow detailed investigation of the combined effects of genetic and environmental determinants of health.<sup>5</sup> It recruited 500,000 participants between the ages of 40-69, each of whom provided information on their lifestyle, medical history, biological specimens as well as consent for their health to be followed-up through health records. Genome-wide genotyping and biomarker data for the full cohort will soon be available. A further subset of 100,000 participants will undergo imaging scans. Ethical oversight is provided by the Ethics and Governance Council, which advises on governance relating to data collection and research, to ensure participants' protection.

UK Biobank is an open-access resource. Researchers from academia, public and commercial sectors (in the UK and internationally) can apply to use the data for research in the public interest on prevention, diagnosis and treatment of a wide range of conditions. Since opening to researchers in 2012, applications have been submitted for a wide range of projects including those for which there were no previous large cohort studies available (such as for mental health, hearing and rare diseases). Research from UK Biobank has begun to be published and all results are returned to UK Biobank so they can be used by other researchers. The resource is funded mainly by the Medical Research Council and Wellcome Trust, and also from the Department of Health, British Heart Foundation, the former Northwest Regional Development Agency, the Scottish Government and the Welsh Assembly Government.

### **Health Related Findings**

Health related findings are results of potential health importance that arise outside of the original purpose for which the test or procedure was conducted. There is debate as to whether researchers undertaking or analysing the baseline data for biobanks have a duty of care to inform participants of any potential serious conditions.

Technological developments, such as in depth genetic research (such as genome wide association studies), make it more likely that researchers will uncover information about an individual's risk of developing certain diseases. Biobanks in the US have moved towards a fully informative feedback model, with the US Presidential Commission for the Study of Bioethical Issues concluding that information about both anticipated and unanticipated findings should be fed back to participants. This has stimulated discussion in the UK.

### **Implications of Feedback for Participants**

While some findings can be lifesaving, others can lead to uncertainty or emotional distress without any corresponding improvement to an individual's health or wellbeing; for example, finding out about a condition for which there is no treatment. There is ethical consideration as to whether the biobank community has a responsibility to protect the public against false alarms and the risk of psychological distress.<sup>6</sup> Furthermore, individual research results derived from cohort studies are often either not clinically relevant and may be misinterpreted by lay participants, or analysed a long time after sample donation, by which time participants may be aware of any condition.<sup>7</sup>

**Box 2. Examples of Cohort Studies****Life Study**

Life Study is the UK's largest national birth cohort study and will collect lifestyle information, physical measures and biological samples from over 80,000 babies born between 2014-2018. Funded by ESRC, MRC and UCL, the study will collect data from before birth through childhood and into adulthood. This study will examine how the family, social and physical environment in early life influences child development, health and wellbeing.

**Twins UK**

Twins UK is a registry of 12,000 sets of identical and non-identical twins used to distinguish between the relative effects of genetic similarities and environmental variables on complex age-related diseases. It began in 1992, funded by the Wellcome Trust, the EU, the National Institute of Health Research and other bodies. It is the most clinically detailed bank of twin data in the world, with whole genome sequencing and comprehensive data on participant's observable characteristics and lifestyle information. Through the open-access resource there has been publication of significant research on conditions such as osteoporosis, metabolic syndromes and cardiovascular disease.

**Million Women Study**

This is a national study of over one million UK women aged over 50, which assesses reproductive and lifestyle factors affecting women's health. It was established in 1996 and is funded by Cancer Research UK, the MRC and the NHS, which also helped to recruit participants through the Breast Screening Programme. Studies have involved collecting lifestyle questionnaires and biological samples in a subset of women, with linkage to individual records and regional and national data on cancer prevalence (such as cancer registries). A broad range of issues have been studied, with particular focus on hormone replacement therapy and its possible role in the development of some cancers.

**Data Sharing between NHS and Biobanks**

Linking healthcare records to biobanks allows researchers to monitor health outcomes and connect them to biological, genetic or behavioural factors and treatments. For example:

- **UK Cohort Studies** are linked to at least one source of routine medical records; for example between Office of National Statistics mortality data and Hospital Episode Statistics. Participants must provide consent to allow the linking of their biobank data with existing and future health records. Research suggests that this extensive linking did not affect the decision to take part in a biobank study.
- **Disease-Specific Biobanks** have begun to link participants' data to those held by disease registries. For instance, Public Health England is piloting the linking of the National Cancer Registration Service registries to disease-specific biobanks. Through this connection it is possible to obtain follow-up patient information, clinical indicators, medical history and outcomes, which may be used to interpret the results for research that uses these samples. It also makes it possible to trace the progress of the disease in the individuals whose data are being tested. This could improve research diagnostics.

**Health Records held by GPs**

The majority of biobanks have not been able to incorporate data from primary care records held by general

**Box 3. Examples of Disease Specific Biobanks****Children's Cancer and Leukaemia Group Bank (CCLG)**

This biobank was established in 1998, funded by Cancer Research UK and the CCLG to provide a research resource on childhood cancers. 12,000 biopsy samples have been collected from 19 centres across the UK and 109 research projects have used the data. The samples are stored with patients' clinical data in order to understand childhood cancer progression.

**Breast Cancer Campaign Tissue Bank**

This is a collaboration between five UK research institutes and the NHS that together have collected samples surplus to diagnosis requirements to create the first national breast cancer tissue bank. Funded by Breast Cancer Campaign, it is open to researchers in the UK and Ireland investigating why breast cancer develops and spreads, and possible treatment options. It also has an open data sharing platform to allow its data to be linked to other genomic and clinical data sets.

**UK Brain Banks Network**

The Medical Research Council has established a network of brain tissue banks as a neuroscience research resource, in line with the GB and UK focus on dementia research. They collect tissue from patients with a detailed clinical history of neurological disorder, alongside those from healthy individuals for comparative analysis. The tissue is analysed with findings linked to clinical and genetic information.

practitioners, despite having participants' explicit consent to do so. Access to such records would add value to biobanks since it would offer researchers a better understanding of conditions that are usually treated in primary care settings, such as mental ill health, diabetes and asthma. Incorporating GP data requires data transfer between interoperable IT systems. This is possible, but it also requires the compliance of GPs who hold these records. However recent controversies over data governance, which in a medical context centre around the NHS care.data issue, may make it much more difficult to overcome these challenges.

**Data-Sharing between Biobanks**

Some biobanks do not have sufficiently large sample sizes to produce statistically significant results. To maximise research potential, UK research funders argue that the immense array of data stored in numerous biobanks should be shared.

Shared data need to be compatible, having been collected following similar coding practices and controls. Efforts to standardise and harmonise data across biobanks are described below. Other concerns include:<sup>8</sup>

- ethical and legal issues related to confidentiality and differences in consent models
- technical challenges of creating interoperable systems, while maintaining data security
- different governance structures including those of intellectual property and research practices
- methodological, intercultural or philosophical differences in scientific knowledge production
- different priorities and expertise of the scientific community that established each biobank.

### Standardising and Harmonising Biobank Data

Each biobank has its own operating procedures that govern how to collect, process, store and share biological samples or related healthcare data. There is variation in the data collected by different biobanks as well as their governance models. Harmonisation is the process to identify these differences and develop techniques to make them compatible. Standardisation imposes consistent standards and protocols on all biobanks, in order to facilitate uniformity of samples for data sharing. While a straightforward solution to promote consistency, it may mean that data already collected cannot be used. It is preferable that operating procedures adhere to agreed standards. Harmonisation initiatives are outlined in Box 4.

### Factors Affecting Participation in Biobanks

Public and patient support groups are seen as vital to the development of biobanks. A large Europe-wide survey found that two thirds of respondents had not heard of biobanks.<sup>9</sup> However, the same research suggests that once aware of biobanking, the public is generally supportive.

### Participant Involvement

Biobanks require individuals to donate their time, biological samples and be prepared to give broad consent for the future use of their data. This is managed by the ethical oversight of each biobank's established governance principles. Participant involvement may be based on an expectation of reciprocity, whether that be to the individual participant (through individual feedback and or payment) or to society (through research development). Public dialogue between researchers, the public and potential participants may give individuals the opportunity to voice their concerns in relation to their participation. This would help biobanks to act within contemporary best practice guidelines to safeguard participants' rights.

### Commercialisation of Biobanks

Research has shown that people are more willing to contribute samples and data if they will be used for publicly-funded research. Many participants express negative opinions about commercial uses of biobanks. This has been linked to perceptions that the reciprocity arrangement between participant and biobank is challenged, or that donated biological samples and associated health data may be used in contentious ways, such as sharing data with the insurance industry, the police or a participant's employer. These suspicions vary between types of biobank. Participants in some disease specific biobanks, who are also patients affected by that condition, understand the benefits to be derived from the use of their samples by pharmaceutical companies for the development of new treatments or medicines. Under current governance models consent must be obtained at the start of the sample collection to allow for both public and commercial access to an individual's data.

#### Box 4. Harmonisation Efforts for Biobanks

##### Public Populations Project in Genomics and Society (P3G)

P3G is a global consortium established in 2007, which aims to develop and manage inter-disciplinary collaboration between biobanks to compare and merge a range of datasets and analyses from already published studies. It also aims to harmonise future epidemiological projects in genomics by:

- optimising the design, methods, tools and core variables for collection in emerging biobanking initiatives
- facilitating knowledge transfer and training among biobank scientists.

##### Biobanking and Biomolecular Resources Research Initiative

This EU-funded initiative, which reached the end of its preparatory phase in 2011, is developing infrastructure for a pan-European Biobank network and interoperable research facilities. The network includes existing and de novo biobanks and biomolecular resources. The aim is to improve European activity in biobanking by:

- building on existing infrastructure and to develop new methodological components to improve harmonisation between EU biobanks
- enhancing data sharing capabilities between organisations to improve medical research to benefit the European population.

### Public Engagement and Trust

Academic studies highlight the importance of ensuring that the public has a clear understanding of the work carried out by biobanks whether they rely on patient or population level data for their studies. Although all work carried out with biobank samples is done in the public interest, and adheres to strict ethics and governance frameworks, a European survey shows that biobanks cannot take public support for granted, and will need to cultivate public confidence. This is seen as especially important in the wake of recent data privacy scandals in the UK and elsewhere. Increased understanding of biobanking research may improve recruitment of participants. Those running and using biobanks see it is an important factor in sustaining continued funding, as most rely on public funding which is dependent on continued public and political support.

#### Endnotes

- 1 Gottweis & Petersen, *Biobanks :Governance in Comparative Perspectives*, Routledge, 2008
- 2 European Commission, *Biobanks for Europe: A challenge for governance, Report of the Expert Group on Dealing with Ethical and Regulatory Challenges of International Biobank Research*, 2012
- 3 *Protecting health and scientific research In the Data Protection Regulations (2012/0011(COD))*, position of non-commercial research organisations and academics, April 2014
- 4 Medical Research Council, Wellcome Trust, *Framework on the feedback of health related findings in research*, 2014
- 5 Collins R *What makes UK Biobank special?* *The Lancet*, Vol 379, March 31 2012
- 6 Wolf et al, *The law of incidental findings in human subjects research: establishing researcher's duties*, *Journal of Law and Medical Ethics*, 36 (2): 361-214, 2008
- 7 Cambon-Thomsen et al, *Trends in ethical and legal frameworks for the use of human biobanks*, *European Respiratory Journal*, 30; 373-382, 2007
- 8 Demir & Murtagh, *Data sharing across biobanks, epistemic values, data mutability and data incommensurability*, *New Genetics and Society*, p359, 2013
- 9 Gaskell et al, *Europeans and Biotechnology in 2010: winds of change - Eurobarometer report to the European Commission's Directorate-General for Research*, 2010