Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology 2008;**37**:234–244
doi:10.1093/ije/dym276

UK BIOBANK: METHODS

The UK Biobank sample handling and storage protocol for the collection, processing and archiving of human blood and urine

Paul Elliott[1]* and Tim C Peakman[2] on behalf of UK Biobank

Accepted December 2007

Background UK Biobank is a large prospective study in the UK to investigate the role of genetic factors, environmental exposures and lifestyle in the causes of major diseases of late and middle age. Extensive data and biological samples are being collected from 500 000 participants aged between 40 and 69 years. The biological samples that are collected and how they are processed and stored will have a major impact on the future scientific usefulness of the UK Biobank resource.

Aims The aim of the UK Biobank sample handling and storage protocol is to specify methods for the collection and storage of participant samples that give maximum scientific return within the available budget. Processing or storage methods that, as far as can be predicted, will preclude current or future assays have been avoided.

Methods The protocol was developed through a review of the literature on sample handling and processing, wide consultation within the academic community and peer review. Protocol development addressed which samples should be collected, how and when they should be processed and how the processed samples should be stored to ensure their long-term integrity. The recommended protocol was extensively tested in a series of validation studies. UK Biobank collects about 45 ml blood and 9 ml of urine with minimal local processing from each participant using the vacutainer system. A variety of preservatives, anti-coagulants and clot accelerators is used appropriate to the expected end use of the samples. Collection of other material (hair, nails, saliva and faeces) was also considered but rejected for the full cohort. Blood and urine samples from participants are transported overnight by commercial courier to a central laboratory where they are processed and aliquots of urine, plasma, serum, white cells and red cells stored in ultra-low temperature archives. Aliquots of whole blood are also stored for potential future production of immortalized cell lines.

[1] Department of Epidemiology and Public Health, Faculty of Medicine, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK.

[2] UK Biobank, Spectrum Way, Adswood, Stockport, Cheshire SK3 0SA, UK.

* Corresponding author. Department of Epidemiology and Public Health, Faculty of Medicine, Imperial College London, St Mary's Camps, Norfolk Place, London W2 1PG.
E-mail: p.elliott@imperial.ac.uk

A standard panel of haematology assays is completed on whole blood from all participants, since such assays need to be conducted on fresh samples (whereas other assays can be done on stored samples). By the end of the recruitment phase, 15 million sample aliquots will be stored in two geographically separate archives: 9.5 million in a −80°C automated archive and 5.5 million in a manual liquid nitrogen archive at −180°C. Because of the size of the study and the numbers of samples obtained from participants, the protocol stipulates a highly automated approach for the processing and storage of samples. Implementation of the processes, technology, systems and facilities has followed best practices used in manufacturing industry to reduce project risk and to build in quality and robustness. The data produced from sample collection, processing and storage are highly complex and are managed by a commercially available LIMS system fully integrated with the entire process.

Conclusion The sample handling and storage protocol adopted by UK Biobank provides quality assured and validated methods that are feasible within the available funding and reflect the size and aims of the project. Experience from recruiting and processing the first 40 000 participants to the study demonstrates that the adopted methods and technologies are fit-for-purpose and robust.

Keywords Large scale prospective studies, biological sample handling and storage protocol, rationale for selection of biological samples for large scale epidemiological studies

Introduction

UK Biobank is a large, prospective national study in the UK to investigate the role of genetic factors, environmental exposures and lifestyle in the causes of major diseases of middle and late age. It is recruiting 500 000 study participants aged between 40 and 69 years with collection of baseline data from April 2007 to the end of 2010. The initial data collection is undertaken in local assessment centres, where information derived from questionnaires, physical and cognitive measurements and samples of blood and urine are obtained. The data and samples will be linked to the participant's medical and other health-related records to allow longitudinal follow-up of disease incidence and mortality. The protocol for the collection, processing and archiving of biological samples was developed through a review of current knowledge and wide consultation and peer review in the scientific community, followed by extensive validation to ensure that the proposed procedures were fit for purpose. The protocol, defined in a series of standard operating procedures, details the samples to be collected, the preliminary processing and storage temperatures, the transport of samples to a central processing facility, and their processing, aliquoting and storage. In order to facilitate rapid dissemination of these methods to the scientific community, the standard operating procedures used are on the UK Biobank website (www.ukbiobank.com).

Methods

Approach to sample collection, processing and archiving

In establishing a long-term archive of biological material, one of the key aims was to provide a resource that would have applicability to a wide range of future scientific questions (rather than a resource focused on a small number of diseases—for example, if there was a particular focus on diabetes, fasting samples would have been collected). Consequently, whilst there is no perfect solution and there are many possible solutions, development of the protocol for the collection, processing and archiving of biological samples in UK Biobank was informed, within the constraints of a finite budget, by a number of key principles (Box 1).

In order to implement these principles into an operating biobank, various options were evaluated:

Process development

The scientific needs of the project determined the protocol development which was then tested against

Box 1 Summary of the principles used in developing the UK Biobank sample handling and storage protocol

Future proofing	• Collect, process and store samples to serve as wide a range of scientific investigation as can be anticipated. • Avoid, where possible, processing approaches that inherently preclude some future analyses.
Quality control and assurance	• Develop and thoroughly test protocols with rigorous quality assurance and control procedures built-in. • Centralize and standardize processing methodology and ensure fit-for-purpose operations (high-throughput, high-quality, robust data trail).
Sample security	• Store sample fractions in ultra-low temperatures to ensure long-term stability and integrity. • Ensure all sample fractions are represented in two geographically separated archives to ensure long-term integrity of the sample fractions. • Protect sample fractions from freeze-thaw degradation by storing multiple aliquots and by withdrawing samples from the archives under ultra-low temperature conditions. • Maintain a detailed and secure data audit trail through the use of a Laboratory Information Management System and robust inventory systems.
Cost	• Ensure cost effectiveness of the protocol by comparing costs and benefits of different options.

feasibility and cost. UK Biobank operates within a finite budget; the cost of each iteration of the protocol was estimated and as a result some of the early proposals were rejected. Where processes or technology were unachievable, the protocol was modified to maintain the scientific goal as closely as possible. For example, the study required that a urine sample be collected from each participant; a spot urine sample is specified because of the difficulty in collecting and processing overnight or timed 24 h urine samples in a high-throughput distributed clinic setting.

Technology development

It was apparent early on that collection and processing of blood and other samples from 500 000 people would require high-throughput industrial-scale operations. It was also clear that no off-the-shelf capacity was available that would address the specific needs of the study and refinement and development of existing technologies would be required. We have attempted to transfer best practice from manufacturing industry around the design and implementation of high-throughput operations. We adopted the principle that any solution should rely on technology that can be delivered to the required throughput and quality with existing hardware, rather than on technology that might be developed in the near future or is still in the development and testing stage. If new technologies are developed that will improve the speed, cost or throughput of the process, these will only be integrated after thorough process and equivalence testing. No in-process technology development will be done.

Central vs local processing

Previous biobanks have been less extensive than UK Biobank or have been assembled from a number of smaller collections with sometimes quite different collection, transport, processing and storage protocols; these might introduce systematic differences in the samples.[1–20] As noted in Box 1, a principle of UK Biobank is the use of standardized procedures such that each sample is collected, transported, processed and stored in the same way with strict quality assurance and quality control for the prevention and detection of errors. Because of the distributed nature of the assessment centres, samples could be processed locally or centrally. Immediate processing and cryopreservation at the local assessment centre has the advantage of preserving any highly labile molecules in the sample. This is only possible at the point of collection but maintaining data structures and quality control in a distributed assessment centre setting is very difficult. Experience from pilot studies showed that samples from about 80 participants per day could be routinely and accurately manually processed by a team of about five or six technicians. Beyond this number, the data structures and logistics become disproportionately complex with concurrent risk of error or processing delay. For UK Biobank, the scientific need for high-quality data was the overriding consideration. Furthermore, the cost of a local process was estimated to be twice that of a centralized one, requiring more staff, hardware and, overall, more facility space. Therefore, early on it was decided that samples should undergo minimal local processing in the assessment centres before being shipped to a central high-throughput facility as soon as practically possible after collection. This introduces an element of compromise since an ideal sample handling and processing protocol would spin down, separate and freeze the samples immediately after collection. Delayed processing may result in highly labile molecules not being adequately represented, and it is possible that contamination from cell lysis may occur despite efforts to separate the cellular component from plasma and serum. An extensive series of validation studies was therefore undertaken to test

Box 2 Samples considered in the UK Biobank sample handling and storage protocol and the rationale for their inclusion or exclusion

Sample type	Inclusion criteria
Blood	• Variety of fractions: plasma, serum, white cells (including peripheral blood lymphocytes), red cells. • Wide range of biomolecules: DNA, RNA [5′ ends], proteins, products of metabolism (including xenobiotics). • Wide physiological coverage: genome, proteome, metabonome, haematological parameters. • Suitable for a wide range of assays. • Ease and low cost of collection.
Urine	• Wide range of biomolecules: electrolytes, products of metabolism (including xenobiotics). • Wide physiological coverage: metabonome (including gut microbiome). • Suitable for many assays. • Ease and low cost of collection (spot urine).
Sample type	**Exclusion criteria**
Faeces	• Limited additional information (e.g. gut microbiome). • Difficulty in collecting/processing. • Potential impact on recruitment. • Complexity and cost of storage.
Hair	• Limited additional information (e.g. exposure to environmental heavy metals). • Complicating effects of cosmetics and toiletries.
Nails	• Limited additional information (e.g. exposure to environmental heavy metals). • Complicating effects of cosmetic products. • Consistency of sample collection. • Possible impact on recruitment of clipping nails.
Saliva	• Limited additional information (e.g. indicators of periodontal disease and oral cancer).

the robustness of the protocol and the impact of the delay in sample collection on cryopreservation.[21]

Sample collection, stabilizing agents and anti-coagulants

Consideration was given to the biological samples that should be collected from the participants at the initial assessment centre visit and what preservatives should be used to protect the samples from degradation prior to cryopreservation at reasonable cost. The aim was to include readily accessible and data-rich biological samples (blood and urine) (Box 2).

During the baseline assessment visit, 45 ml of blood and 9 ml of urine are collected. This amount of blood is a trade off between sufficient volume to enable a wide range of additives to be used, what can reasonably be requested from a participant, and the cost of collection, processing and archiving of the samples. With blood and urine included, other sample types were also considered (faeces, hair, nails and saliva: Box 2), but were not thought likely to provide sufficient additional information to warrant the potential inconvenience to participants and cost of collection from all 500 000 participants. For example, bacterial gut fermentation by-products in faeces are biomarkers of a number of diseases of the gut (such as irritable bowel syndrome and, possibly, Crohn's disease).[22] These markers include hydrogen, methane, alkanes, methyl alkanes, phenols and organic acids, but they can also be measured accurately in urine. Furthermore, the gut microbiome can be profiled in urine using NMR approaches.[22] Hair and nails may be used to assess medium-term exposure to heavy metals. But, a study of the toxicokinetics of methylmercury exposure concluded that hair and blood levels are of questionable value as indicators of both body and target organ concentrations of mercury.[23] Moreover, some forms of arsenic (such as arsenobetaine, the major organic arsenic compound in seafood) do not accumulate in hair.[24] In addition, measures of environmental arsenic in hair and nails are influenced by external contaminants (such as air, water, soaps and shampoos), and such exposure is better measured in urine.[25]

Sample collection conditions

We reviewed the stability of a wide range of bioanalytes and cells as a component of whole blood in different anti-coagulant media, at varying temperatures and under differing transport conditions. Bioanalytes can be divided into two general categories: (i) known biochemicals, bioanalytes (such as DNA, defined proteins and specific metabolites), haematological parameters and cell types and (ii) unknown biomolecular entities, such as the constituent plasma/serum proteome and metabonome, as well as any as yet undefined bioanalytes.[1] Where possible, best practice for ensuring stability of known bioanalytes (e.g. cytokines) was extrapolated to include unknown biomolecular entities, and the validity of these assumptions was tested in validation studies.[21]

Design and testing of the sample handling protocol considered key factors that affect the stability of biological samples, including: anti-coagulants, stabilizing agents, temperature, elapsed time from collection to initial processing and endogenous degrading properties (enzymes, cell death). In addition, sterility of the samples may be important for high rates of virally mediated cell transformation. Finally, we aimed for cost-efficiency by avoiding collecting multiple sources of material for the same analyte. Thus, we decided against storage of whole blood on paper

media as a source of DNA despite demonstrating good recovery of DNA in a validation study.[26] Rather, a source of high molecular weight DNA suitable for whole genome amplification can be obtained from white cells and an inexhaustible supply is potentially available from immortalized B-cell lines.

Anti-coagulants

The selection of the preservatives and additives used in the collection of blood is important in determining future applicability of the samples.[2] For example, the collection of whole blood in any type of anti-coagulant-containing tubes may induce cytokine production *in vitro* and result in artificially elevated concentrations.[27] Certain anti-coagulants are recommended or even required for analytical purposes, whilst others may be contra-indicated.[1] Tubes containing an acid citrate dextrose (ACD) mix need to be used if the peripheral blood lymphocytes are required for the production of immortalized cell lines. Heparin-stabilized blood affects T-cell proliferation, binds to some proteins, and affects the performance of DNA in subsequent polymerase chain reaction-based assays. However, in metabonomic studies plasma from tubes containing lithium–heparin is preferred. Better quality RNA and DNA can be extracted from blood stabilized with citrate than with other anti-coagulants, and a higher yield of lymphocytes is produced for culture. The use of sodium citrate is recommended if functional clotting factor assays are required, though fibrinogen antigen can be measured in a sample collected into EDTA by using an enzyme immunoassay for immunologically intact fibrinogen.[28,29] Although EDTA influences Mg^{2+} concentration and poses problems for cytogenetic analysis (increases sister chromatid exchanges, decreases mitotic index, etc.), collection tubes coated with EDTA provide blood fractions suitable for a wide range of DNA-based and protein assays. EDTA and lithium–heparin are, therefore, being used in UK Biobank as broad purpose anti-coagulants, for high-quality plasma in line with the principle of future proofing in Box 1. In addition, we are collecting blood into tubes containing ACD as a source of peripheral blood lymphocytes for future immortalization.

Clot accelerators

As well as the collection of plasma and cell samples, serum is required as it is more suitable for certain types of assay (e.g. most clinical biochemistry and metabolomic studies) and provides additional assurance for future proofing. Whole blood left at room temperature will clot in glass collection vessels but this was rejected because there is a risk of breakage. Since whole blood will clot only partially or very slowly in untreated plastic collection vessels, plastic serum collection tubes contain a clot accelerator. There are two types of clot accelerators commonly used: silica or thrombin. Thrombin produces the quickest clotting (typically being used in cardiac setting where rapid effects are needed), but it is prohibitively expensive for a study the size of UK Biobank. Instead, in UK Biobank we are using silica because it is affordable and data show that the serum is unaffected for a range of assays and technologies.[30–33]

Degradation

Enzymatic degradation affects many biochemical biomarkers.[1] For example, EDTA and ascorbic acid are stabilizing agents for folate in blood, and should be added as soon as possible after blood collection to assure the accuracy of the analysis.[1,34–36] Proteins are sensitive to degradation by proteases, particularly if cell integrity has been compromised; protease inhibitors can be added to the sample immediately after collection, but they are toxic to live cells (so must not be added if cell viability is required) and may also interfere with antibody-based or functional assays. RNA is particularly sensitive to degradation by abundant and ubiquitous RNAses; RNA integrity is maintained with RNAse-free handling and the addition of commercially available RNAse inhibitors. However, as with protease inhibitors, RNAse inhibitors may compromise the utility of the samples for other assays. In UK Biobank, we are currently not including reagents in the blood samples that actively prevent degradation because of their potential to interfere with future assays, their toxicity and, specifically for RNA, intra-individual variability and their lability during long-term cryopreservation. In addition, the cost of collecting and processing the blood in a way that would protect the RNA is prohibitively expensive for the whole cohort but might be considered for a sub-sample. Instead, a dynamic representation of the transcriptome can be obtained from immortalized lymphocytes derived from the blood collected into the ACD tube, albeit in a B-cell background. Concerning the urine sample, various preservatives (such as boric acid) have been used to prevent bacterial contamination. However, to avoid the potential for the formation of chemical complexes in the samples we are not including boric acid in the urine collected in UK Biobank (by contrast with some other studies).[37] The use of azide as a preservative for urine that does not form protein complexes was also carefully considered but was rejected because of the difficulty in maintaining consistency and quality of sample handling in the assessment centres, the potential for variable dilution factors through different initial sample volumes, health and safety concerns over use of this toxic compound and the potential impact on unknown metabolites.[38]

Temperature and time to initial processing

Temperature may affect sample stability during the time between sample collection and sample

Table 1 Sample preservative, collection priority and volume

Type of sample	Collection priority	Volume collected (ml)	Transport temperature (°C)	Notes
EDTA	1	9	4	Spray dried to give a final concentration of 1.8 mg/ml blood
LH (PST)	2	8	4	Spray dried to give a final concentration of 17 IU/ml – Proprietary gel plug.
Clot activator (SST)	3	8	4	Silica clot accelerator used in plastic collection tubes. Proprietary gel plug.
EDTA	4	9	4	
Acid citrate dextrose	5	6	18	Tube contains 1.0 ml of solution giving 2.2 mg/ml blood sodium citrate/0.8 mg/ml citric acid/2.45 mg/ml dextrose
EDTA	6	4	4	
Urine	–	9	4	

Also shown is the temperature the samples are maintained at during transport to the central processing centre. (PST, plasma separation tube; SST, serum separation tube; LH, lithium heparin.)

processing, and archiving if the samples are not processed immediately after collection. Live cells are stable at room temperature for up to 48 h, but must be either cultured or cryopreserved in liquid nitrogen in order to remain viable. Freezing the sample without separation is, however, incompatible with maintaining viable cells for isolation [as cells will rupture if frozen without dimethylsulphoxide (DMSO)]. For biomolecules, when the samples are not being analysed immediately, they should be maintained at low temperature because room temperature results in degradation of labile protein biomarkers (e.g. cytokines), anti-oxidants (ascorbic acid, uric acid, α-tocopherol), and other analytes (such as folate and vitamin B_{12}).[39] Ideally, the different temperature requirements for the stability of each biomarker would be addressed, but this is not practicable in a large-scale project. Four degrees centigrade is a good compromise between immediate freezing and room temperature: cells can remain viable (though viability is reduced compared to room temperature), while the samples are protected, at least to some extent, against enzymatic degradation of sensitive protein biomarkers.[1] In a study comparing the stability and recovery in blood of several cytokines [tumour necrosis factor (TNF), interferon-α (IFN-α), IFN-γ, interleukin-lα (IL-lα), IL-Iβ, and IL-6] under varying collection protocols (e.g. anti-coagulants, stabilizing agents, preservatives, transport temperatures, times and conditions), stability was improved by storage at 4°C.[40] In that study, among the anti-coagulants tested, samples treated with EDTA performed most consistently with least assay problems. For UK Biobank, the sample stability and cell viability at 4°C and 18°C were tested in a series of validation studies that showed integrity of the samples for a range of biomarkers was maintained at least up to 24 h and the cells remained viable for up to 36 h post-collection.[21]

Sample collection protocol

In UK Biobank, blood and urine is collected from participants into vacutainer tubes according to the schedule in Table 1. During venepuncture, the hypodermic needle is connected to these vacutainer tubes, and the slight internal vacuum draws sufficient blood to fill each tube which contains the required additives. A similar system is used to transfer the participant's urine into a vacutainer from the urine collection vessel.

Sample processing at the assessment centres

As noted, processing of blood and urine samples at the assessment centres is kept to a minimum. When blood is collected from a participant, the vacutainers are inverted 10 times to mix the anti-coagulant/preservative with the whole blood. One lithium–heparin plasma separation (LH—PST) and one clot activator serum separation tube (SST) are used to increase the protection against changes with delayed separation that might affect certain assays (increasing levels of potassium and homocysteine). These tubes contain a proprietary, chemically inert, gel plug that forms a barrier to cellular material but allows plasma/serum to pass through when the tubes are centrifuged at low speeds, so producing sample separation.

After collection of a complete set of vacutainers (or as many as possible according to a pre-defined collection order shown in Table 1), the unique bar code on each one is scanned into the assessment centre IT system to link each vacutainer with the particpant's unique identifier number. Bar coded vacutainers are not pre-assigned to participants: this avoids participant identification errors and prevents empty blood collection tubes being logged into the Laboratory Information Management System (LIMS). The scanned bar code also automatically initiates a timer built into the assessment centre IT system to

facilitate accurate measurement of clotting time for the serum separator tube (see below). The blood in the plasma separation tube is immediately centrifuged at 2500*g* for 10 min at 4°C and the time of centrifugation recorded in the assessment centre IT system. The blood in the serum separator tube is allowed to clot for 25–30 min at room temperature before centrifugation at 2500*g* for 10 min at 4°C; the time of centrifugation is recorded in the assessment centre IT system. All vacutainers are held at 4°C (with the exception of the acid citrate dextrose tube which is held at 18°C) until the end of the day when they are packed (with temperature logging devices) and dispatched by a commercial courier to the central processing laboratory in temperature-controlled shipping boxes.

Sample processing at the central processing laboratory

When the vacutainers arrive on the following morning at the central processing laboratory, they are processed as soon as possible and aliquots obtained as described in Table 2. Sample aliquots are held at 4°C during processing and are transferred to long term ultra-low temperature archives as soon as possible after processing with the aim that all samples are cryopreserved not >24 h after collection.

The bar codes of all the vacutainers that arrive are scanned and compared against the LIMS data file from the assessment centres to ensure the correct tubes have arrived and the laboratory data file can be linked to the other participant data. Because of the high-throughput in UK Biobank (up to 5000 vacutainers of blood and urine per day) the samples are processed using automated systems,[41] with times and temperatures of all operations and operator identifiers logged in the LIMS:

EDTA vacutainers (9 ml)

The blood in the two vacutainers coated with EDTA is centrifuged at 2500*g* for 10 min at 4°C. Aliquots of plasma, buffy coat and red cells are transferred to 1.4 ml 2D bar coded cryostorage tubes with split-septum seals arrayed in 96 position racks (Society for Biomolecular Sciences standard footprint http://www.sbsonline.org/msdc/pdf/lsc9904-9908.pdf) and the LIMS data record is updated.

LH (plasma separator) and clot activator (serum separator) vacutainers

The plasma and serum fractions produced by the gel separator plug are transferred directly to labelled cryostorage tubes as specified in Table 2, and the LIMS data record updated.

Acid citrate dextrose vacutainer

A 500 µl aliquot of whole blood is mixed with 500 µl of sterile 20% DMSO (diluted with RPMI (Roswell Park Medical Institute) growth medium) in labelled cryostorage tubes in a sterile Laminar airflow cabinet and the LIMS data record updated. The aliquots are placed into a programmable cell freezer (Planer Products UK) held at 4°C and frozen using a standard programme (reduction of 1°C/min to −30°C and then reduction of 10°C/min to −100°C). On completion of the programme, the cryotubes are transferred to storage in the vapour phase of liquid nitrogen.

Table 2 Fractions and aliquots of blood and urine samples stored either at −80°C or in vapour phase liquid nitrogen

		Number of aliquots	
Vacutainer tube	**Fractions**	−80°C	Liquid N_2
EDTA (9 ml) × 2	Plasma	6	2
	Buffy coat	2	2
	Red cells	1	1
LH (PST)	Plasma	3	1
Clot activator (SST)	Serum	3	1
ACD	DMSO blood	–	2
EDTA (4 ml)	Haematology (immediate)	–	–
Urine	Urine	4	2
Total Aliquots		19	11

PST, plasma separation tube; SST, serum separation tube; LH, lithium heparin.

Box 3 Haematological assays being performed on whole blood from 4 ml EDTA vacutainers

Haemoglobin	Platelet count
Packed cell volume	White cell count
Red cell count	Neutrophil count
Mean cell volume	Lymphocyte count
Mean cell haemoglobin	Monocyte count
Mean cell haemoglobin concentration	Eosinophil count
	Basophil count

EDTA (4 ml) vacutainers

Only those assays that cannot be performed on cryopreserved samples are being done on fresh samples. The whole blood in the 4 ml EDTA vacutainers is used for a range of standard haematological assays (Box 3) on a Beckman automated haematology analyser. Data are attributed to the vacutainer bar-code and the LIMS data set updated.

Urine vacutainers

Aliquots of urine are transferred to labelled cryostorage tubes as specified in Table 2, and the LIMS data record updated.

LIMS, data structures and data security

In keeping with UK Biobank's overall approach to technology implementation, a robust commercially

available LIMS system (Nautilus), initially developed for high-throughput analytical laboratories, is being used. Whilst this LIMS package does require some specialist configuration to align it with specific laboratory processes, it is ideal for well-described and repeated operations because of the logical data structure and compatibility with Oracle databases. The overall data structures and systems have been designed and implemented in parallel with the protocol according to several guiding principles. The use of bar codes (1-dimensional on the vacutainers and 2-dimensional on the 1.4 ml aliquot storage tubes) ensures accuracy and traceability of all samples from the assessment centre to the archive. Because the processes are fully automated, additional information (such as time and date of collection and archiving; temperature log from collection to archiving; and real-time logging of sample volume) can be attributed to each aliquot and linked to the individual participant. Participant confidentiality is of paramount concern: all samples and their aliquots are identified only by bar codes and can be linked back to the participant only through a unique 12-digit participant identifier. The data in the LIMS are held separately from both the data from the assessment centres and the data that identify each individual. Access to the entire database structure is strictly restricted.

Process automation and quality assurance

The processing of the various samples in the central laboratory is highly automated. This greatly increases the achievable throughput, reduces the scope for operator error, and increases consistency and reproducibility. It also enables the complex data structure to be managed. Using an approach common in manufacturing industry, careful process design, prototyping testing and integration have been used to develop robust, affordable, fully redundant, automated methods that prevent unexpected process bottlenecks or technology and data incompatibility. An essential requirement of the UK Biobank resource for future users will be the ready availability of high quality samples. Therefore, each process step includes quality assurance procedures and the data outputs have an associated component that identifies the machine and operator involved in processing the sample. Daily monitoring enables the rapid identification and rectification of quality problems using foolproof devices or measures built into the process to detect defects. For example, the first three digits of the 12-digit bar code on the vacutainer tubes specify the tube type to the automation and prevent the tubes being incorrectly processed. The bar codes also allow the mixed batching of samples to reduce total processing time as much as possible because different tube types can be loaded onto, and recognized by, the same automated processing platform. The emphasis is thus shifted from detecting and reducing the number of defects (completing an inaccurate or low quality procedure on a blood or urine tube) and errors (completing the wrong procedure on a blood or urine tube) to designing in-process quality assurance that reduces or eliminates errors *per se*. Any defects are controlled either by halting the process (shut out type) or by raising an alert to the operator (attention type). Batch quality control procedures have been built into the process and it is the responsibility of the laboratory manager to ensure the process outputs are of the required quality and consistency.[41]

Long-term sample storage

Stability studies of a range of biomolecules over long periods under controlled conditions are rare and restricted to a limited number of pre-defined proteins and analytes.[1] Several studies have looked at a number of cytokines,[40,42] which are sensitive indicators of sample degradation as they represent only a small fraction of the total plasma proteome. Those studies examined the impact of storage at various temperatures, including the comparison between storage at about −70°C in freezers and at −180°C or below in liquid nitrogen over periods up to 6 years. The influence of storage temperature and freeze–thaw cycles on plasma levels of β2M, sIL-2R, neopterin, IFN-γ, sTNF-RII and TNF-α was examined at temperatures ranging from room temperature to −70°C over a period of 20 days; −70°C proved the optimal temperature in terms of future sample stability, although a number of these labile proteins did not suffer degradation even at room temperature.[43] A 6-year study aimed at comparing the stability of coagulation, fibrinolysis and inflammatory factors in stored plasma samples[44] showed no significant difference in decay when comparing storage at −70°C or in liquid nitrogen. Another study examined the stability of albumin, apolipoprotein A-1, apolipoprotein B, cholesterol, creatinine kinase, creatinine, fibrinogen, HDL-cholesterol, LDL-cholesterol, total protein and triglycerides in plasma stored at −20°C, −40, −80 and −180°C or below for up to 6 years (S Clark, Oxford, personal communication). Whilst degradation was detected in some analytes in samples stored at −20°C and −40°C, no such effect has been detected with storage at −80°C or −180°C. A quality control programme is being designed in UK Biobank to assess the stability of a broad range of analytes during long-term cryopreservation throughout the lifetime of the study.

By the end of recruitment, UK Biobank will be storing about 15 million 1.4 ml aliquot tubes. For reason of security (Box 1), samples from each participant will be stored in two geographically separate locations. One location will house the 'working' archive that will typically be used first for any research project and the other location will house the 'back-up' archive (as well as the aliquots of whole blood mixed with DMSO for potential cell immortalization) that will be used when samples in the working archive have been exhausted (or in case

certain analytes are only stable in long term storage at much lower temperatures than −80°C). At full capacity, the working archive will hold 9.5 million sample tubes at −80°C, and will use custom-built robust industrial automated processes for reliable storage and retrieval of samples. Tubes are stored in racks inside rows of cabinets maintained at −80°C by liquid nitrogen circulating in a closed system.[45] Industrial scale robots move between these rows of cabinets on linear tracks in an environment maintained at −20°C, and place or withdraw samples as required. Only those tubes that are required for assays are extracted and thawed, with the rest returned still-frozen to the −80°C cabinets (i.e. avoiding freeze–thaw events). The entire operation is controlled by an operator outside the store in normal working conditions. The back up archive will hold 5.5 million sample tubes at −180°C or below in liquid nitrogen vapour in insulated stainless steel tanks that require manual loading and retrieval of samples. The vessels are continually monitored for temperature and nitrogen vapour is delivered automatically through super-insulated pipes as required. Samples will be retrieved manually under standard safety and operating procedures.[46] For both archives, systems have been established to log times and temperatures during transfer, operator identifiers and exact archive location in the LIMS.

Discussion

The UK Biobank sample handling and storage protocol has been developed according to principles based on future proofing, quality control and assurance, sample security and cost. In parallel with the protocol, fit-for-purpose processes, facilities and technology have been established to help ensure the long-term integrity of the samples. Clearly, the protocol and processes adopted in UK Biobank offer only one solution among many; various choices and trade-offs had to be made to realize the goal of collecting, transporting and archiving samples from 500 000 participants, with high-throughput and high quality, within a specified budget.

The samples undergo minimal processing locally in the assessment centres before being shipped to the central laboratory for processing with the aim of cryopreservation within 24 h of collection. Samples are protected against degradation during shipping by being chilled at 4°C, except for those blood samples that are to be used for immortalization of peripheral blood lymphocytes, which are transported at 18°C. This centralized approach ensures high levels of quality control and is cost effective, but there are potential trade-offs compared with more localized processing (such as loss of some highly labile substances and changes in specific analytes during transit). However, results of extensive validation studies indicate that such changes are small for a wide range of analytes.[21]

Once the samples have been processed in the central laboratory, they are placed in one of two archives maintained at −80°C for the working archive and in nitrogen vapour at −180°C or below for the back-up archive. The temperature of −80°C for the working archive was chosen because previous studies (albeit limited in number and size) have shown stability of samples for long periods for a range of analytes, including highly labile molecules such as cytokines.[44] Given the importance of sample integrity, an on-going study of sample stability is being designed to check for systematic changes due to storage conditions over the lifetime of the project The nitrogen store is a geographically separate archive, providing security against physical events (such as fires) that may damage the resource. It provides the lowest possible practicable storage temperature for unknown analytes that may be labile even at −80°C over prolonged storage periods. It is also suitable for storage of whole blood for future immortalization of peripheral blood lymphocytes.

The UK Biobank resource is expected to be used to assess the relevance of different exposures for disease chiefly through a series of retrospective case-control studies of particular health outcomes 'nested' within the cohort (although other study designs are also likely). The nested case-control strategy has the advantage that most biological assays (other than haematology which cannot use stored samples) will only need to be conducted retrospectively on stored baseline blood and urine samples from cases of the particular disease and from matched controls. Consequently, it allows assays to be performed more cost effectively on a relatively small subset of the cohort (e.g. a few thousand or tens of thousands people, rather than all 500 000), which helps avoid sample depletion and facilitates good quality control (e.g. minimizing assay drift due to changes in assay conditions or reagents). Even in a cohort of 500 000 individuals, it will take several years before sufficient numbers have developed any particular disease to allow reliable statistical analyses. Consequently, this retrospective approach has the additional advantage that decisions about what assays to perform need only be made some years in the future when specific hypotheses will be clearer than at the time of collection, and the range of assays that can be conducted with available resources is much wider. It is also possible as a complementary strategy to collect additional blood from a substantial subset of the cohort for processing and measurement on fresh samples, which would allow questions related to intermediate traits such as metabolic disorders, hypercholesterolaemia etc., to be addressed using the initial cross-sectional data.

In designing and commissioning the centralized processing laboratory facility, it was important to use technology and processes that are established and robust, transferring best practice from manufacturing

industry, as failure could lead to loss of the resource. Although the main working archive is highly automated, it is the configuration of the various technology elements that is new rather than the technology itself. All of the laboratory consumables and supporting LIMS structures are standard, off-the-shelf products; for example, the 1.4 ml cryotubes for the storage of sample aliquots are available at low cost and in a format that fits all standard automation. The bar codes are four times redundant (i.e. they can still be read even when damaged) and are moulded into the tube construction to avoid loss. Use of the split-septum seal allows high throughputs on the automation because it avoids the need to cap the tubes manually, and reduces the potential for bacterial contamination. The tubes and seals are tested and rated for the ultra-low temperature conditions that are used in the archives. Further, the labels on the vacutainers are attached with a permanent adhesive and contain a check sum, which reduces the read failure rate to 1 per 5.5 million bar codes read.

The team structure in the laboratory is also based around manufacturing principles. We are using cellular teams of technicians on two shifts per day, 6 days per week, to complete the sample processing and archiving with clear handover periods and protocols. Each team is managed by a supervisor and is responsible for a defined part of the process. They are also responsible for quality control and quality and process improvement initiatives in their area. We have not implemented automation with the aim of reducing headcount; rather the automation ensures high-throughput, high quality and a robust data trail, while the technical staff ensures that the processes are completed satisfactorily, and are on hand to address any breakdowns or quality alerts.

To date (mid April 2007 to end November 2007) we have recruited 37 000 participants and processed 180 560 samples of blood and 36 250 samples of urine producing 971 445 aliquots. The average time to archiving (collection of blood and urine to storage at ultra-low temperature) is 22.75 h (SD = 2.88). During this period, the processing and archiving operation has operated well within capacity and there have been no occasions when the laboratory has not been available to process samples. Although we are currently operating at 70% of projected peak throughput, the laboratory has been designed to operate at full capacity to the same stringent targets for processing time and quality.

In summary, UK Biobank has adopted a sample handling and storage protocol that is fit for purpose for the size and aims of the project within the available funding. However, we recognize that other equally valid solutions are possible. Ultimately, the selected protocol for other biobanks will reflect the aims, size and design of the particular study, but is likely to be developed using similar guiding principles.

Acknowledgements

This protocol is the result of wide consultation with a number of groups and individuals: UK Biobank Steering Committee: R. Collins (chair), V Beral, P. Burton, P. Elliott, J. Gallacher, J. Pell, A. Silman. *UK Biobank Sample Handling and Storage Sub-group: P. Elliott (chair), M. Caulfield, N. Davies, A. Dominiczek, M. McCarthy, B. Ollier, T. Peakman, N. Samani, J. Sampson*

Others: S. Clarke, K. Kourellias, J. Nicholson, P. Vadgama, P. Vineis, I. Young, P. Downey, A. young. The UK Biobank project is funded by the Medical Research Council, The Wellcome Trust, Department of Health for England and Wales, North West Regional Development Agency and the Scottish Executive.

Conflict of interest: None declared.

References

[1] Holland NT, Smith MT, Eskenazi B, Bastaki M. Biological sample collection and processing for molecular epidemiological studies. *Mutat Res* 2003;**543:**217–34.

[2] Landi MT, Caporaso N. Sample collection, processing and storage. *IARC Sci Publ* 1997;**142:**223–36.

[3] Lum A, Le Marchand L. A simple mouthwash method for obtaining genomic DNA in molecular epidemiological studies. *Cancer Epidemiol Biomarkers Prev* 1998;**7:**719–24.

[4] Muti P. Biological specimen banks: epidemiological tools and scientific enterprises. *Epidemiol Prev* 2003;**27:**121–25.

[5] Steinberg K, Beck J, Nickerson D *et al*. DNA banking for epidemiologic studies: a review of current practices. *Epidemiology* 2002;**13:**246–54.

[6] Pero RW, Olsson A, Bryngelsson C *et al*. Quality control program for storage of biologically banked blood specimens in the Malmo Diet and Cancer Study. *Cancer Epidemiol Biomarkers Prev* 1998;**7:**803–8.

[7] Schunemann HJ, Stanulla M, Trevisan M, Aplan PD, Freudenheim JL, Muti P. Short-term storage of blood samples and DNA isolation in serum separator tubes for application in epidemiological studies and clinical research. *Ann Epidemiol* 2000;**10:**538–44.

[8] Boixes Sana D, Badia Mallorqui M. Peripheral specimen collecting centers. *Rev Enferm* 1998;**21:**13–16. [Spanish].

[9] Taioli E, Bonassi S. Pooled analysis of epidemiological studies involving biological markers. *Int J Hyg Environ Health* 2003;**206:**109–15.

[10] Royster MO, Lobdell DT, Mendola P *et al*. Evaluation of a container for collection and shipment of semen with potential uses in population-based, clinical, and occupational settings. *J Androl* 2000;**21:**478–84.

[11] Jewell SD, Srinivasan M, McCart LM *et al*. Analysis of the molecular quality of human tissues: an experience from the Cooperative Human Tissue Network. *Am J Clin Pathol* 2002;**118:**733–41.

[12] Harty LC, Shields PG, Winn DM, Caporaso NE, Hayes RB. Self-collection of oral epithelial cell DNA under instruction from epidemiologic interviewers. *Am J Epidemiol* 2000;**151:**199–205.

[13] Wagener DK. Ethical considerations in the design and execution of the National and Hispanic Health and

Nutrition Examination Survey (HANES). *Environ Health Perspect* 1995;**103(Suppl 3):**75–80.

[14] Norgaard-Pedersen B, Simonsen H. Biological specimen banks in neonatal screening. *Acta Paediatr Suppl* 1999;**88:** 106–9.

[15] Gunter EW. Biological and environmental specimen banking at the Centers for Disease Control and Prevention. *Chemosphere* 1997;**34:**1945–53.

[16] Tourtellotte WW, Rosario IP, Conrad A, Syndulko K. Human neuro-specimen banking 1961-1992. The National Neurological Research Specimen Bank (a donor program of pre- and post-mortem tissues and cerebrospinal fluid/blood; and a collection of cryopreserved human neurological specimens for neuroscientists). *J Neural Transm Suppl* 1993;**39:5**–15.

[17] Rossbach M, Giernich G, Emons H. Representative sampling and sample preparation in biological environmental monitoring using spruce shoots. *J Environ Monit* 2001;**3:**330–34.

[18] van der Hel OL, van der Luijt RB, Bueno de Mesquita HB *et al*. Quality and quantity of DNA isolated from frozen urine in population-based research. *Anal Biochem* 2002;**304:**206–11.

[19] Nederhand RJ, Droog S, Kluft C, Simoons ML, de Maat MP; Investigators of the EUROPA trial. Logistics and quality control for DNA sampling in large multicenter studies. *J Thromb Haemost* 2003;**1:**987–91.

[20] Riboli E, Kaaks R. The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol* 1997;**26(Suppl 1):**S6–14.

[21] The UK Biobank sample handling and storage validation studies. *Int J Epidemiol* 2008;**37(Suppl 1):**i2-i6.

[22] Nicholson JK, Holmes E, Wilson ID. Gut microbes, mammalian metabolism and personalized healthcare. *Nature Reviews, Microbiology* 2005;**3:**431–38.

[23] Nielsen JB, Andersen O, Grandjean P. Evaluation of mercury in hair, blood and muscle as biomarkers for methylmercury exposure in male and female mice. *Arch Toxicol* 1994;**68:**317–21.

[24] Vahter M, Marafante E, Dencker L. Metabolism of arsenobetaine in mice, rats and rabbits. *Sci Total Environ* 1983;**30:**197–211.

[25] Harrington JM, Middaugh DL, Housworth J. A survey of a population exposed to high concentrations of arsenic in well water in Fairbanks, Alaska. *Am J Epidemiol* 1978;**108:**377–85.

[26] Halsall A, Ravetto P, Reyes Y *et al*. The quality of DNA extracted from liquid or dried blood is not adversely affected by storage at 4°C for up to 24 hours. *Int J Epidemiol* 2008;**37(Suppl 1):**i7–i10.

[27] House RV. Cytokine measurement techniques for assessing hypersensitivity. *Toxicology* 2001;**158:**51–58.

[28] Sobel JH, Wu HQ, Canfield RE. The development of assays for the detection of fibrin(ogen)olysis based on COOH-terminal A alpha chain epitopes. *Blood* 1994;**84:** 535–46.

[29] Kohek M, Leme C, Nak I *et al*. Effects of EDTA and sodium citrate on hormone measurements by fluorometric (FIA) and immunofluorometric (IFMA) methods. *BMC Clinical Pathology* 2002;**2:**2. Available at www.biomedcentral.com/1472-6890/2/2/ (last accessed 8th January).

[30] Becton Dickinson technical product paper (VS7173). *Comparison of Becton Dickinson vacutainer SSTII Advance tubes with Becton Dickinson vacutainer serum glass tubes for routine chemistry analytes*.

[31] Becton Dickinson technical product paper (VS7351). *Comparison of adjusted Becton Dickinson vacutainer SSTII Advance tubes with Becton Dickinson vacutainer serum glass tubes for cortisol, total T3, total T4 and TSH on the DPC Immulite 1000 analyzer*.

[32] Becton Dickinson technical product paper (VS7352). *Comparison of adjusted Becton Dickinson vacutainer SSTII Advance tubes with Becton Dickinson vacutainer serum glass tubes for cortisol, total T3, total T4 and TSH on the Abbott AxSYM analyzer*.

[33] Becton Dickinson technical product paper (VS7192). *Comparison of Becton Dickinson vacutainer SSTII Advance tubes with Becton Dickinson vacutainer SSTII plus tubes for selected chemistry analytes*.

[34] Kerkay CM, Coburn D, McEvoy D. Effect of sodium ascorbate concentration on the stability of samples for determination of serum folate levels. *Am J Clin Pathol* 1977;**68:**481–84.

[35] MacGregor JT, Wehr CM, Hiatt RA *et al*. 'Spontaneous' genetic damage in man: evaluation of interindividual variability, relationship among markers of damage and influence of nutritional status. *Mutat Res* 1997;**377:**125–35.

[36] Fenech M, Aitken C, Rinaldi J. Folate, vitamin B12, homocysteine status and DNA damage in young Australian adults. *Carcinogenesis* 1998;**19:**1163–71.

[37] Stamler J, Elliott P, Dennis B *et al*. INTERMAP: background, aims, design, methods, and descriptive statistics (nondietary). *J Hum Hypertens* 2003;**17:**591–608.

[38] Barton RH. *Kinetic and Stereoselective Characterisation of Lamb Pregastric Lipase Catalysed Reactions.* PhD thesis. New Zealand: University of Auckland, May 1997.

[39] Komaromy-Hiller G, Nuttall KL, Ashwood ER. Effect of storage on serum vitamin B12 and folate stability. *Ann Clin Lab Sci* 1997;**27:**249–53.

[40] Thavasu PW, Longhurst S, Joel SP, Slevin ML, Balkwill FR. Measuring cytokine levels in blood. Importance of anticoagulants, processing, and storage conditions. *J Immunol Methods* 1992;**153:**115–24.

[41] McQuillan A. C, Sales S. Designing an automated blood fractionation system. *Int J Epidemiol* 2008;**37(Suppl 1):** i51–i55.

[42] Aziz N, Nishanian P, Taylor JM *et al*. Stability of plasma levels of cytokines and soluble activation markers in patients with human immunodeficiency virus infection. *J Infect Dis* 1999;**179:**843–48.

[43] Aziz N, Nishanian P, Mitsuyasu R, Detels R, Fahey JL. Variables that affect assays for plasma cytokines and soluble activation markers. *Clin Diagn Lab Immunol* 1999;**6:**89–95.

[44] Lewis MR, Callas PW, Jenny NS, Tracy RP. Longitudinal stability of coagulation, fibrinolysis, and inflammation factors in stored plasma samples. *Thromb Haemost* 2001;**86:**1495–500.

[45] Owen JM, Woods P. Designing and implementing a large scale automated −80°C archive. *Int J Epidemiol* 2008; **37(Suppl 1):**i56–i61.

[46] Fagan M, Ball P. Design and implementation of a large-scale liquid nitrogen archive. *Int J Epidemiol* 2008; **37(Suppl 1):**i62–i64.