Centre for Longitudinal Studies

Biosamples Strategy

Document Information	
Document Name	CLS_Biosamples_Strategy_v1.7.docx
Author(s)	Tom Murphy, Alissa Goodman, Aida Sanchez
Version	1.7
Issue Date	December 2020
Approved By	CLS Data Access Committee
Review frequency	Yearly

Document History			
Version	Author	Date	Summary of change
0.0	Tom Murphy	12/07/2017	First draft
0.1	Alissa Goodman	21/07/2017	Comments added by Alissa
0.2	Tom Murphy	16/08/2017	Implemented changes suggested by Alissa's comments, split NCDS sample table into 2 tables on saliva and blood
0.3	Tom Murphy	26/09/2017	Implemented Alice's changes. Comments added by Emla.
1.0	Tom Murphy	09/11/2017	Minor changes to introduction (wording, timeline info). This is the first draft to send to Sue Ring to look at
1.5	Tom Murphy	17/01/2018	Implemented changes as per comments from Sue Ring, Wendy McArdle, Aida Sanchez
1.6	Matt Brown	26/02/2019	Minor changes following meeting between Alice and Matt
1.7	Aida Sanchez	15/12/2020	Full re-formatting and addition of 2013 NCDS sample approval process in Appendix

Table of Contents

1. Introduction

The aim of this document is to provide an overview of the biological samples and data available for the CLS cohort studies, and to document the strategy for governance of BCS70 and MCS depletable biological samples going forward.

In developing this strategy, CLS intends to:

1. Consult funders and CLS DAC on the following strategy (January 2018).
2. Review ethical approval for all CLS cohort samples.
3. Launch a consultation with potential users of the BCS70 depletable resource to decide on a priority list of depletions and strategy for BCS70 samples (late Spring 2018).
4. Enact strategy for depletion jointly with CLS DAC, and publish the sample collections on the UKCRC tissue directory.
5. Hand over governance of the BCS70 and MCS samples to CLS DAC, which will then take applications for sample access.

2. Sample summaries and inventories

2.1 NCDS

The 2002/3 Biomedical Survey for NCDS collected whole blood and saliva from cohort members. The samples were processed and both original aliquots and residues are held at the University of Bristol who are the custodians of these biological samples. These are banked at the Bristol Bioresource Laboratories, Population Health Sciences, Bristol Medical School, University of Bristol. Arrangements are currently being made to transfer custodianship of these samples to CLS, with the intention of continued biobanking at Bristol. Various biological samples have been collected, some analysis was completed at the time of collection and remaining samples were aliquoted and stored for future use. Further details are provided in Tables 1-6.

Lymphoblastoid cell lines (LCLs) were generated by transforming lymphocytes with Epstein Barr virus. A DNA bank was created from two sample types, extraction from whole blood which is a finite resource and from the transformed LCLs, a renewable source. Both of these collections have been extensively genotyped.

Table 1 – Summary of saliva samples collected and assays included in biomedical sweep from September 2002 to March 2004

Sample type	Early Morning Saliva (Sarstedt salivettes)	Late Morning Saliva (Sarstedt salivettes)
Maximum number of aliquots remaining	1 (varying volumes)	1 (varying volumes)
Number of cases with at least 1 500 µL sample remaining	6618	6618
Processing protocol	Transported by post at ambient temp. Frozen - at -80°C in temporary storage, Shipped at ambient temperature to Germany for analysis. Refrozen on arrival. No information regarding how samples were shipped back but currently stored at -80°C.	Transported by post at ambient temp. Frozen - at -80°C in temporary storage, Shipped at ambient temperature to Germany for analysis. Refrozen on arrival. No information regarding how samples were shipped back but currently stored at -80°C.
Processing Location	St George's Hospital Medical School then Germany	St George's Hospital Medical School then Germany
Existing analytes	Cortisol	Cortisol
Data available from sample	6467 cases	6506 cases
Current location	University of Bristol	UK Biocentre

Table 2 – Summary of blood samples collected and assays included in biomedical sweep from September 2002 to March 2004

Sample type	**Citrated Plasma Residue**	**Plain Serum Residue**	**EDTA Plasma**	**CPDA Plasma (citrate-phosphate-dextrose-adenine)**
Maximum number of aliquots remaining	1 (varying volumes)	1 (varying volumes)	Up to 6 x 500 µL * +1 varying volume	Up to 6 x 500 µL ** +1 varying volume
Number of cases with at least 1 500µL sample remaining	7597	6400	8063	7848
Processing protocol	Shipped by post at ambient temp. 0.5mlof whole blood removed for analysis of glycosylated haemoglobin. Remainder centrifuged, aliquots frozen at -70°C, transported frozen to Glasgow Royal Infirmary for	Shipped by post at ambient temp. Centrifuged and the supernatant serum used for analysis in Newcastle. Residue retained at -80°C	Shipped by post at ambient temp. Centrifuged and supernatant plasma stored in 0.5ml individually barcoded aliquots at -80°C. Cell residues frozen and	Specific blood tube for production of lymphoblastoid cell lines. Shipped to Bristol by post at ambient temp. Centrifuged and plasma removed. Peripheral blood lymphocytes separated on a Ficoll gradient and cryopreserved for subsequent transformation into immortalised cell

Sample type	Citrated Plasma Residue	Plain Serum Residue	EDTA Plasma	CPDA Plasma (citrate-phosphate-dextrose-adenine)
	analysis. Residue retained at -80°C		transported frozen to Bristol for DNA extraction.	cultures. The supernatant plasma was sent to St George's Hospital Medical School (SGHMS) for aliquoting into 0.5ml individually barcoded tubes which were frozen at -80°C for long-term storage.
Processing Location	Royal Victoria Infirmary, Newcastle	Royal Victoria Infirmary, Newcastle	St George's Hospital Medical School	ALSPAC, University of Bristol then St George's
Days from taking sample to arrival in lab	1 day 18.9% 2 days 47.1% 3 days 24.2% 4 days 7.2% 5 days 1.5% >5 days 1.0%	1 day 18.9% 2 days 47.1% 3 days 24.2% 4 days 7.2% 5 days 1.5% >5 days 1.0%		Time to reach ALSPAC 1 day 17.8% 2 days 45.6% 3 days 24.9% 4 days 7.8% 5 days 1.9% >5 days 2.0%
Existing analytes	Glycosylated haemoglobin fibrinogen Tissue	Triglycerides Total and HDL cholesterol Total	N/A	N/A

Sample type	Citrated Plasma Residue	Plain Serum Residue	EDTA Plasma	CPDA Plasma (citrate-phosphate-dextrose-adenine)
	plasminogen Activator Von Willebrand factor C-reactive protein.	and allergen- specific immunoglobulin E Insulin-like growth factor 1		
Data available from sample	8226 cases	8226 cases	See table 6	See table 6
Current location	University of Bristol	UK Biocentre	University of Bristol and UK Biocentre	University of Bristol and UK Biocentre

* See table 3 for more details

** See table 4 for more details

Table 3 EDTA Plasma – Further details of number of 500 µL aliquots

Number of 500 µl aliquots available	Number of cases
6	5110
5	2270
4	488
3	122
2	54
1	19

Table 4 – CPDA Plasma – Further details of number of 500 µl aliquots

Number of 500 µL aliquots available	Number of cases
6	7137
5	380
4	123
3	84
2	65
1	59

Table 5 – Summary of blood derived DNA and lymphoblastoid cell line samples available for NCDS

Sample Type	**Blood derived DNA (finite)**	**Lymphoblastoid cell lines (renewable)**
Amount available	Samples for 8017 individuals	Samples for 7526 Individuals
Current location	Bristol Bioresource Laboratories, University of Bristol.	Main collection banked at Bristol Bioresource Laboratories with backup storage at European Collection of Authenticated Cell Cultures (ECACC). LGC genomics hold aliquots for genotyping.

Table 6 – NCDS genotype data deposited at the European Genome-phenome Archive

Study	Chip used	EGA reference	Number of cases
WTCCC1	Affymetrix 500k	EGAD00000000001	1502
	Illumina 15K Custom chip	EGAD00000000010	1476
	Infinium HumanHap 550K v1.1		1436
WTCCC2	Affymetrix v6	EGAD00000000021	2997
	Illumina 1.2M	EGAD00000000024	2922
T1DGC	Infinium HumanHap 550K v3	EGAD00000000030	2598
METABOCHIP	Illumina HumanExome-12v1_A-GenCall, zCall	EGAD00010000234	5839
GABRIEL	Illumina Human 660	EGAD00000000073	839
Immunochip	Immunochip	EGAD00010000248	6862
ICR1000	Illumina HiSeq 2500	EGAS00001000971	1000
HLA	Dynal RELI SSO assay	EGAD00000000031	6662

2.2 BCS70

The 2016/2017 Biomedical Survey for BCS70 collected whole blood samples from cohort members. Three tubes of blood were collected. One serum tube and one EDTA tube were posted by the nurse to RVI, Newcastle to conduct biomarker assays (see tables 7 and 8 below). Another EDTA tube was posted by the nurse to the University of Bristol, where a buffy coat was produced for future DNA extraction, and 5 aliquots of plasma were stored for future analyses.

CLS is the custodian of the samples collected for BCS70, which are held at the Bristol Bioresource Laboratory. The available samples and existing analytes are summarised in tables 7 and 8 below.

Table 7 – Summary of samples collected and assays included in BCS70 biomedical sweep from July 2016 to May 2018

Sample type	**Plain serum**	**EDTA Plasma**
Maximum number of aliquots remaining	0	5 x 500 µL per person plus buffy coat for DNA extraction
Number of cases with at least one 500ul sample remaining	0	Tbc
Processing protocol	4 ml tube shipped by post at ambient temp. to RVI laboratory, Newcastle for analysis. No residual serum stored.	One 4ml and one 6ml tube collected. 4 ml tube shipped by post at ambient temp. to RVI laboratory, Newcastle for analysis of HbA1c. 6 ml tube shipped by post at ambient temp. Centrifuged and supernatant plasma stored in 0.5ml individually barcoded aliquots at -80 C. Buffy coat retained for future DNA extraction.
Processing Location	RVI, Newcastle	RVI, Newcastle and Bristol Bioresource Laboratory
Days from taking sample to arrival in lab	tbc	tbc
Existing analytes	Total cholesterol HDL cholesterol Triglycerides* C-reactive protein* Insulin like growth factor* Ferritin* Cytomegalovirus (IgG and IgM)* Haemoglobin*	HbA1c

	Red blood cell count (only analysed in second half of fieldwork)	
Data available from sample	tbc	N/A
Current location	N/A	Bristol Bioresource Laboratory

*These assays only conducted on samples collected in the second half of the fieldwork, from May 2017 onwards. Efforts will be made to seek funding for these assays to be conducted on samples collected prior to this.

Table 8 – EDTA Plasma - Further details of number of 500 µL aliquots

Number of 500µl aliquots available	Number of cases
5	Tbc
4	Tbc
3	Tbc
2	Tbc
1	Tbc

Table 9 – Summary of blood derived DNA available for BCS70

Sample type	Blood derived DNA
Amount available	Currently DNA has been extracted for 240 samples for epigenetics project, further samples tbc
Data available from sample	240 cases with methylation data, to be made available in future
Current location	Bristol Bioresource Laboratory

2.3 MCS

Milk Teeth were collected at age 7 (MCS3) and onwards, in order to measure the cohort members' exposure to lead. These are stored at the Institute of Child Health, at University College London.

Oral fluid was collected at age 3 (MCS2) to enable analyses on the pattern of immunity to common childhood infections. All oral fluid samples were depleted and residues destroyed.

At age 14 (MCS6) saliva was collected from both cohort members and natural parents (where present in the home). DNA has been extracted from the saliva sample, and unlike the NCDS transformed lymphocytes collection this is non-renewable. The MCS samples are summarised in Table 9 below.

Table 9 - Summary of Samples collected and assays included in MCS.

Sample type	**Milk teeth**	**Oral fluid (age 3 sweep – depleted and residues destroyed)**	**DNA from Saliva collected in Oragene kits (cohort member) age 14 sweep**	**DNA from Saliva collected in Oragene kits (natural parents) age 14 sweep**
Amount remaining/available	3026 cases with at least one tooth collected	Samples fully depleted	9364 cases	14135 cases
Processing protocol	Transported by post to UCL Institute of Child Health for analysis.	Transported by post at ambient temperature to HPA Microbiology Services Division for analysis.	Transported by post at ambient temperature to Bristol. DNA extracted	Transported by post at ambient temperature to Bristol. DNA extracted
Processing location	UCL Institute of Child Health	HPA Microbiology Services Division	Bristol Bioresource Laboratory	Bristol Bioresource Laboratory
Existing assays	Lead level	Immunoglobulin G (IgG) Varicella-zoster virus (VZV)	GWAS data is currently being generated using Illumina global	GWAS data is currently being generated using Illumina global

		Epstein-Barr virus (EBV) Norovirus Adenovirus	screening array	screening array
Data available from sample	2755 cases with lead level data	12473 cases with data from this sample	tbc	tbc
Current location	UCL Institute of Child Health	N/A - depleted/destroyed	Bristol Bioresource Laboratory	Bristol Bioresource Laboratory

3. Governance and use of samples

Use of the samples from all CLS cohorts is covered by Research Tissue Bank Ethical Approval 19/NW/0710, and applications are governed by CLS DAC.

3.1 NCDS

The guidelines for access to the samples are extensively documented in 1958 Birth Cohort Biosample Strategy Guidelines set out in 2013 (see Appendix of this document) the summary of which states the following on recommendations for approval of any further depletion of the samples:

- Scientific strength of the proposal must justify use of 1958 cohort samples and thus all proposals are subject to peer review
- Evidence must be provided to show methodology is appropriate given the processing history of the samples. Eg. Evidence from published literature or pilot data generated on samples processed in a similar manner.
- The assay test platform should have proven quality assurance measures in place.
- The methodology should include measures to ensure the quality of any remaining sample is not jeopardised and can be used in further assays which can be used on freeze thawed samples.
- At least one aliquot of each sample type should be reserved for future global discovery projects.

If a proposal is approved samples will be provided with the following conditions:

- Costs incurred in providing samples will be covered by the applicant. These will include costs for retrieval, additional processing necessary for the specific project, shipping costs (both out and return) and linking data. Costs will be provided on a case by case basis depending on the work involved and may be subject to VAT.
- All results generated from samples must be returned for inclusion in the data resource and will be made available to other researchers.
- Where it is possible to use samples which have been thawed and refrozen, samples which have been returned from other projects will be supplied in preference to unused stock if available.
- DNA will only be extracted at a single lab as part of a managed resource. This will be at Bristol Bioresource Laboratory.
- CLS reserve the right to specify where analysis will be carried out in order to ensure results obtained are comparable to existing data.

3.2 BCS70

Governance of BCS70 samples has been transferred into the remit of the CLS DAC following a process for first determining the study's priority uses of the samples.

Applications for access to BCS70 samples will be subject to similar (but not identical) criteria as for NCDS. Priority will be given to proposals that satisfy certain criteria, which will be determined following a consultation with funders, CLS DAC, and potential users of the resource. Potential criteria include the following:

- Priority may be given to coverage of the additional analytes that were covered in NCDS, to enable comparisons between the cohorts (Immunoglobulin E, Fibrinogen, Tissue plasminogen activator antigen (t-PA) and von Willebrand factor antigen (vWF).
- Priority may be given to full coverage of analytes only covered in the second half of fieldwork (Triglicerides, C-reactive protein, insulin like growth factor, ferritin, cytomegalovirus, haemoglobin, and red blood cell count).
- Priority will be given to applications for a specified list of priority uses, which have been determined following a process of scientific consultation.
- Scientific strength of the proposal, potential impact, and novelty of the scientific aims must justify use of the BCS70 cohort samples, and thus all proposals for non-priority uses are subject to peer review.
- Usage of the samples should be specifically relevant to the BCS70 study. Applications to use the samples should clearly demonstrate that the proposed study will make use of longitudinal data and cannot be carried out in samples obtained from another source.
- Evidence must be provided to show methodology is appropriate given the processing history of the samples E.g. Evidence from published literature or pilot data generated on samples processed in a similar manner.
- The assay test platform should have proven quality assurance measures in place
- At least one aliquot of each sample type should be reserved for future global discovery projects
- The requested volume does not unreasonably deplete the resource
- The methodology should include measures to ensure the quality of any remaining sample is not jeopardised and can be used in further assays which can be used on freeze thawed samples.
- The work proposed is covered by existing CLS ethical approval and is within the scope of the consents obtained for the specific samples.

If a proposal is approved samples will be provided with the following conditions:

- Costs incurred in providing samples will be covered by the applicant. These will include costs for retrieval, additional processing necessary for the specific project, shipping costs (both out and return) and linking data. Costs will be provided on a case by case basis depending on the work involved and may be subject to VAT.
- All results generated from samples must be returned for inclusion in the data resource and will be made available to other researchers.
- Where it is possible to use samples which have been thawed and refrozen, samples which have been returned from other projects will be supplied in preference to unused stock if available.
- DNA will only be produced at a single lab as part of a managed resource. This will be at Bristol Bioresource Laboratory.
- CLS reserve the right to specify where analysis will be carried out in order to ensure results obtained are comparable to existing data.

3.3 MCS

Milk teeth collected in the MCS are not currently available, a review of the access and governance of these samples is being undertaken.

Applications to CLS DAC for access to the DNA samples from the age 14 saliva are welcome. Proposals for depletion of the DNA samples must satisfy the following criteria to be approved:

- Scientific strength of the proposal, potential impact, and novelty of the scientific aims must justify use of the MCS cohort samples, and thus all proposals for use are subject to peer review.
- Usage of the samples should be specifically relevant to the MCS study. Applications to use the samples should clearly demonstrate that the proposed study will make use of longitudinal data and cannot be carried out in samples obtained from another source.
- The assay test platform should have proven quality assurance measures in place
- A quantity of DNA should be reserved for future global discovery projects
- The requested volume does not unreasonably deplete the resource (Amount of DNA per individual is variable – depending on amount of DNA requested possible that only a subset of samples may be available).
- The work proposed is covered by existing CLS ethical approval and is within the scope of the consents obtained for the specific samples.

If a proposal is approved samples will be provided with the following conditions:

- Costs incurred in providing samples will be covered by the applicant. These will include costs for retrieval, additional processing necessary for the specific project,

shipping costs (both out and return) and linking data. Costs will be provided on a case by case basis depending on the work involved and may be subject to VAT.

- All results generated from samples must be returned for inclusion in the data resource and will be made available to other researchers.
- CLS reserve the right to specify where analysis will be carried out in order to ensure results obtained are comparable to existing data.

4. Data sharing

4.1 NCDS

Biochemical marker data from the NCDS biological samples is currently available in the National Child Development Study: Biomedical Data, 2002-2004, available under Special License from the UK Data Service. Currently there are plans to amend this so that this dataset is available under end user licence.

The majority of genotype data is held at the European Genome-phenome Archive. Access to genotype data unlinked to any other data (except region, sex and ethnicity) is through the Wellcome Trust Case Control Consortium Data Access Committee (WTCCC DAC) (https://www.ebi.ac.uk/ega/dacs/EGAC00001000205). Access to genotypes linked to other variables, applications for access to DNA, and for new uses of biological samples is via the CLS DAC.

Copies of the datasets held at EGA are also held at Newcastle University to enable genotype-phenotype linkage for applications that have been approved by the CLS DAC.

4.2 BCS70

Non-genotype data arising from the samples will be returned to CLS, which will make these data available via the UK data service under end user licence.

Genotype data will be deposited with the European Genome-phenome Archive (EGA). A copy of these datasets will also be stored at the University of Newcastle to enable genotype-phenotype linkage under similar arrangements that apply for NCDS genotype data. Access to the genotype data will be governed by the CLS DAC. Efforts will be made to seek funding for these data management arrangements.

4.3 MCS

Data arising from the age 3 sweep oral fluid are available from the UK data service.

Lead level data from milk teeth is not currently available, a review is being undertaken around the access and governance of this data.

As with BCS70, genotype data arising from the age 14 saliva samples will be deposited with the European Genome-phenome Archive (EGA). A copy of these datasets will also be stored at the University of Newcastle to enable genotype-phenotype linkage under similar arrangements that apply for NCDS genotype data. Access to the genotype data will be governed by the CLS DAC. Efforts will be made to seek funding for these data management arrangements.

Appendix: NCDS sample sharing strategy 2013

1958 Birth Cohort Biosample Strategy

Guidelines May 2013

Naveed Sattar and Paul Welsh

University of Glasgow

Helen Colhoun

University of Dundee

Susan Ring

University of Bristol

Objective of the scientific strategy guidelines

The 1958 Birth Cohort is a unique and powerful longitudinal epidemiological study, with tissue samples stored in biobanks, which will allow further biomarker and epidemiological work. Available tissue includes saliva, plasma and serum samples which are described in detail in Appendix 1. DNA and lymphoblastoid cell lines are also available from cohort members but are not covered by this document. Use of the samples is covered by Research Tissue Bank Ethical Approval (09/H1010/12) and requests to use the material are assessed by the Access Committee for CLS Cohorts (ACCC) (see http://www2.le.ac.uk/projects/birthcohort) . The tissue samples are a finite resource and the ethical approval requires that requests to access the material is subject to peer review. The material was collected during the Biomedical Sweep in 2003 and is the remains of samples analysed at that time. There has been little interest in the samples until late 2012 when the ACCC started to receive requests to access the material.

The objective of this document is to facilitate access to the 1958 stored tissue samples so that they get the widest possible usage while ensuring that scientific rigour is applied in selecting proposals that will yield data which are i) reliable ii) epidemiologically or clinically informative iii) novel. As such, applications will be considered in light of the cohort design; successful proposal should maximise the epidemiological strengths of the cohort, whilst also recognising limitations of the biobank (in terms of blood draw protocols, processing, storage, and sample availability).

This document provides a framework for addressing and determining the scientific rationale for access issues for biomarker work. This document does not prescribe rigid criteria because it is impossible to predict the nature of access requests or long-term trends in scientific interest. This document, whilst not exhaustive, sets a framework for making relevant decisions, giving some relevant examples where appropriate. This document has been developed to reflect current best practice and will be reviewed regularly to ensure it remains in line with current guidelines. The strategy also needs to reflect current funder policy and the ACCC will consult/update funders if there are any proposed changes to the strategy.

1. Use of the samples should be specifically relevant to the 1958 study

Applications to use 1958 samples should clearly demonstrate that the proposed study will make use of longitudinal data and cannot be carried out in samples obtained from another source. All data generated from samples will be returned to the 1958 cohort and made available to other users. Samples will only be issued under the terms of a material transfer agreement which includes the statement:

“It is a condition of access to the samples that information obtained from the samples (including any derived data, for example, derived haplotypes or the results of bioassays) is submitted to the University of Bristol for inclusion in the central 1958BC database. All genotypes, and all bioassay results that are important enough to be used in a publication must be returned to the 1958BC database. “

Recipients will also be required to return or destroy any unused material at the end of the project as requested by the ACCC under the terms of the material transfer agreement.

2. Scientific strength of the proposal, and potential impact

Critically, one must always ask whether a particular biomarker to be measured will answer a relevant and meaningful question. Using longitudinal studies as a cross-sectional resource is rarely impactful (aside from Mendelian Randomisation studies). Further, use of longitudinal data to investigate associations (hazard ratios, or risk ratios) must be justified on the grounds of potential clinical (or social) relevance. Which questions are generally meaningful in biomarker studies?

i) Clinical questions which might change the guidelines for clinicians, or give a clear public health message. Examples could include:

   a. Disease diagnosis e.g. HbA1c for diabetes or LFTs for NAFLD

   b. Vitamin D status in pregnant women and BMD in their children

ii) Clinical or social questions which might risk stratify patients e.g.

   a. Does NT-proBNP add informative to existing CVD risk scores?

   b. Do novel biomarkers improve prediction of clinical or social outcomes beyond established predictors

iii) Disease pathogenesis. Observational studies tend to be poor in investigating causality, even where impressive multivariable adjustment models are built. Wherever possible, proposals of this nature should consider whether a robust approach to causal identification can be applied, for example including whether the DNA resource can be combined with the proposal to use a Mendelian randomisation approach (assuming valid genetic instrumental variables are known and measured):

   a. Do natriuretic peptides protect against diabetes?

iv) Stratifying patients for therapy based on phenotypes. Does a particular biomarker predict better or worse response to particular therapies?

3. Novelty of the scientific aims

Often the proposals with the most obvious and immediate scientific rigour will be the least novel studies; several cohorts may have conducted similar studies before. As such the balance between a proposal's strength (in terms of potential impact) and its novelty (which studies have measured the biomarker and related measures to outcomes before) is a key factor. If a proposal to measure a novel biomarker with little previous literature is interesting and potentially impactful, this must be considered in light of what is known regarding the biomarker (points below). Often, if a biomarker is particularly novel, a small pilot study may be useful prior to committing samples from the bioresource.

4. Biomarker characteristics; pre-analytical variables

Given the scarce nature of the bioresource, pre-analytical considerations as to whether a biomarker can be measured to give reliable results in the 1958 tissue samples are a key consideration (specific details for each sample type are provided in Appendix 1):

i) Sample processing: The 1958 blood samples were sent by post. The time spent with serum/plasma in contact with cells will have a significant impact on some biomarkers, but not others. Platelets release inflammatory factors, cells metabolise others, and the time spent at room temperature may adversely affect labile proteins. As such it should be noted that UK biobank have investigated pre-analytical characteristics of several of the more common biomarkers:

a. Glucose requires fast separation and assay to be conducted on first thaw.

b. C-reactive protein (CRP) is extremely robust to pre-analytical variables.

c. Limited existing data suggest metabolomics analysis may not be appropriate in samples not rapidly separated or at least within 24 hours

Given this, proposals must make it clear, with robust data to support the proposal, that the biomarkers to be measured will be reliably measured using the 1958 samples. This could be demonstrated with a pilot study, or published data, showing that sample processing time has no impact on the biomarker, or at least has a highly predictable effect (Passing-Blok regression, Bland- Altman plots etc). Pilots are always helpful before committing considerable time and money on novel biomarkers

ii) Freeze-thaw: The EDTA samples have not been previously thawed, whereas the citrate has. Many immunoassays, which measure based on antigenic structure rather than protein activity, are very robust to freeze thaw. This is likely to be the case for most biomarkers that are relatively unaffected by the sample processing time. Nonetheless, in supporting a proposal, data on the impact of freeze-thaw on a biomarker would be useful. In order to maximise use of the resource, it should be considered whether a previously thawed

aliquot would be more appropriate to use (where possible) for a biomarker known to be robust to freeze-thaw.

iii) Sample type: There is more EDTA available than serum or citrate. The remaining serum aliquot is therefore important. Therefore, biomarkers which can be measured on EDTA should be in order to save the scarce serum resource for outstanding proposals. Very few non-haematological biomarkers are routinely measured in citrated plasma samples.

iv) Sample stability: All blood samples are stored at ≤-70oC, so this issue is of limited relevance for biomarkers in the 1958 study.

5. Assay test platform

Assays should, where possible, be carried out using gold standard automated methods. In order of preference;

i) On an automated clinical chemistry/immunoassay platform in an accredited NHS laboratory, or a lab that participates in external quality assurance schemes for that assay

ii) On an automated platform in a laboratory using manufacturer recommended or internal quality control material

iii) Using single-plex assays such as ELISAs

iv) Using multiplex immunoassays

This list is intended as broad guidance, and there will be other potential assay methodologies. The gold standard for measuring vitamin D (25OHD2 and D3) is liquid chromatography tandem mass spectroscopy. Many aspects of this assay can be automated and carried out in NHS labs.

There is a broad trend towards use of multiplex assays to make optimal use of bioresources in epidemiology. Our own experience suggests that this technology should be used with caution. We have experience with Luminex (magnetic beads), Randox (bio-chips) and MSD (Multi-spot ELISA with electrochemiluminscence reporter) platforms. We have found:

i) Extra information comes at the cost of vastly reduced sensitivity and precision (higher CVs).

ii) Luminex beads system is rather sub-optimal for human blood samples; the beads tend to clog together making the assay method difficult/impossible to carry out within manufacturer recommended tolerances.

iii) The assay panel in multiplex assays are often of limited incremental value. Assaying C Reactive Protein (CRP) and Interleukin 6 (IL-6) in a study may be useful, but the incremental value of a dozen other cytokines may be limited or lack cost benefit, particularly when a

majority are below the limit of sensitivity, or have limited or uncertain biological relevance. NB: multiplex assays often lead to reduced sensitivity for some tests and tend to lower CVs. Furthermore, where assay perform better e.g. MSD platform, there may be issues with respect to external

iv) Comparisons of data since some assays give results which are not externally comparable to values obtained by gold-standard methodologies, thereby required a conversion or “fiddle” factor.

Given the above, any proposal should be able to demonstrate that the assay they propose is sensitive enough to detect a signal (<20% CV as absolute and more desirable <10%) in a majority of the samples (commensurate with the aims). Ideally the platform/manufacturer used should be established in the literature to maximise the potential impact of the results, and minimise potential referee criticisms.

6. Assay test characteristics

This is a practical consideration, once a strong scientific case for a biomarker has been made in a proposal. An automated assay will have a dead volume (often ~200uL). For all assays the volume of sample consumed by the assay should also be considered in light of the potential impact of the study. If an EDTA sample has been previously thawed, the repeated use of this sample for other assays should be considered. If the volume remaining is too small for an automated assay it may remain sufficient for use in an ELISA assay by manual pipetting by a technician.

Often, multiple tests can be run on the same sample in automated platforms thereby maximising efficiency.

7. Global Discovery Versus Specific Hypothesis

All the above refers to specific tests of hypotheses; an alternative approach would be to reserve part of the resource for a more global discovery approach; specifically, it would be of interest across a wide range of disease states and phenotypes to acquire as much data as possible on the lipidome, proteome and metabolome from high dimensional methods. For consideration might be mass spectroscopy (often semi-quantitative) and Nuclear magnetic resonance (NMR) based methods for quantitation of many small molecular weight metabolites and some peptides and proteins. Most experts in the field suggest if sufficient volume is available, the best approach for metabolomics is a combination of mass spectroscopy and NMR. Also for consideration are antibody-based arrays for high dimensional protein quantitation. Other methods to consider include proximal ligation assays for proteins, NMR

based methods and mass spectroscopy methods for molecular species lipid analyses etc. Also one might consider serum micro RNAs worth detecting and quantifying.

Many of these approaches require relatively little volume (e.g. at least 600 serum metabolites can be detected and quantified with 120 ul, whereas other Mass spectroscopy platforms can yield potentially more than 1000 metabolites on 20 ul serum). However what is also true is that for many of the available platforms there is a surprising dearth of good data on the within person repeatability over short periods of time , the test re-test repeatability, pre-analytic effects on sensitivity and specificity and so on i.e. basic QC. For protein arrays etc, sensitivities may be particularly important to check since for some specific measurements high sensitivity single-plex ELISAs are employed (e.g. IL-6 in cohort studies) since conventional assays (and potentially arrays) cannot reliably pick up such low levels. Therefore before committing such a precious resource to any of these platforms careful consideration and possibly some pilot studies with less valuable samples are to be recommended. Furthermore, for some of these techniques, the statistical analyses can be very complex and in some cases, the best bioinformatics approach to analyse data, in particular data generated from mass spectroscopy, remains unclear. Finally, in all cases, whilst new techniques allow discovery science, the linkage of any measurements to pre-defined outcomes or to answer specific questions on disease pathology will help focus analyses.

SUMMARY

The 1958 tissue samples are a valuable resource but there are limitations regarding their suitability for some assays due to the sample processing history. Recommendations for CLS DAC for approving use of the samples are:

- Scientific strength of the proposal must justify use of 1958 cohort samples.
- Evidence must be provided to show methodology is appropriate given the processing history of the samples. Eg. Evidence from published literature or pilot data generated on samples processed in a similar manner.
- The assay test platform should have proven quality assurance measures in place.
- The methodology should include measures to ensure the quality of any remaining sample is not jeopardised and can be used in further assays which can be used on freeze thawed samples.
- At least one aliquot of each sample type should be reserved for future global discovery projects.