NIHR BTRU grant Web-Appendix
Separately-funded assays in baseline samples of 50,000 INTERVAL donors

The Affymetrix Axiom UK Biobank Genotyping (Axiom UKBG) Array
The Axiom UKBG array is a genotyping array containing 820,967 single nucleotide polymorphism (SNP) and indel markers selected primarily for 4 different reasons:
1) 349k markers providing **genome-wide coverage ("GWAS scaffold") of common variants** down to a frequency of 5%
2) 280k markers providing **genome-wide coverage ("GWAS scaffold") of low-frequency variants** to allow imputation between 1% and 5%
3) 112k exonic coding variants that are likely to be **functional** (ie, missense and truncating variants), akin to an "Exome array"
4) 95k variants with known or possible associations with **phenotypic variation**, including all variants known to be associated with any phenotype through genomewide association studies (GWAS), variants with previously established clinical effects from the Human Mutation Database, plus 2545 variants selected for relevance to various blood phenotypes including regulation and formation of red blood cells and platelets, as well as regulation of iron homeostasis.

The genome-wide coverage of the array should allow imputation of ~20M common (frequency>5%) and low-frequency variants (1%-5%) in European populations using dense reference datasets from sequencing projects such as the UK10K project and the 1000 Genomes project. The array should be particularly informative in British populations as information used to select the markers was based on sequencing data from these same sequencing projects.

The array is being assayed at the Affymetrix genotyping service centre in Santa Clara (US) - expected completion: 2015.

NMR-based quantitative metabolomics assay
A scientific team in Finland, led by Professor Mika Ala-Korpela, has developed an automated high-throughput metabolomics assay that can provide molar concentrations of >220 measured and derived metabolites (Kettunen *et al., Nat Genet*, 2012;44:269). This platform utilizes proton nuclear magnetic resonance (NMR) spectroscopy in 3 different molecular windows to generate information on 3 classes of analyte:

1) **lipoprotein subclasses and particle concentrations**
2) **lipids**
3) **low molecular-weight metabolites** (eg, amino acids).

By contrast with other NMR platforms, this approach is fully automated and hence all 50,000 INTERVAL donors can be assayed within a year. This platform also has the advantage compared to mass spectroscopy (MS) approaches of providing absolute concentrations of each analyte rather than relative concentrations, increasing the interpretability of the data produced.

This assay has been run in other smaller datasets, which have already yielded high-impact findings in relation to the genetic determinants of metabolites (Kettunen *et al., Nat Genet*, 2012;44:269) and prediction of all-cause mortality and disease-specific deaths (Fischer *et al., PLoS Genet*, 2014;e1001606).

These assays are being performed in baseline samples at Brainshake, a Finnish company established by Prof. Ala-Korpela specifically to rapidly conduct this high-throughput assay - expected completion: 2015.

Clinical chemistry biomarkers
The following 32 biomarkers are being assayed in the baseline samples of all INTERVAL donors:

Lipids and lipoproteins
Cholesterol
HDL-cholesterol
Triglyceride
Direct LDL-cholesterol
Lipoprotein (a)
Apolipoprotein A
Apolipoprotein B

Glycaemic markers
HbA1c
Glucose

Inflammation / autoimmunity
C-reactive protein
Fibrinogen
Rheumatoid factor
D-dimer

Renal function markers
Enzymatic creatinine
Urea
Urate
Cystatin C

Liver markers
Alkaline phosphatase
Alanine aminotransferase
Aspartate aminotransferase
Gamma glutamyltransferase
Direct bilirubin
Total bilirubin

Steroid hormone
Sex hormone binding globulin
Testosterone
Oestradiol

Minerals / electrolytes
Calcium
Phosphate

Bone-related markers
IGF-1
Vitamin D

Other markers
Total protein
NT-proBNP

These biomarkers are being assayed at UK BioCentre, which is running the same panel in the 500,000 person UK Biobank study, and hence has established high-throughput assay pipelines for all assays - expected completion: 2015.

Extended haematology profiling

Assessment of haematological profiles is being measured on fresh whole blood samples taken from each participant using a Sysmex Haematology Analyser XN Series (Sysmex Corporation, Kobe, Japan). The Sysmex analyser uses a highly sensitive automated system that combines fluorescent flow cytometry and cell counting methods to provide information on ~200 blood cell parameters including measures of red cell traits, (ie, haemoglobin levels, packed cell volume, mean corpuscular volume, reticulocyte count, red blood cell distribution width, and reticulocyte haemoglobin), platelets and white blood cells.

The Sysmex assay is being run on fresh samples taken at baseline and at the 2-year re-survey (with funds requested in this bid for additional measurement at the 4-year re-survey), since these parameters cannot be accurately measured in frozen samples. The assays are being conducted at UK BioCentre, with input from Sysmex staff to ensure optimum assay performance - expected completion of 2-year assessment assays: mid-2016.