Research opportunities in UK Biobank

Cancer Outcomes Conference
June 2012

Cathie Sudlow
Reader and Honorary Consultant Neurologist, University of Edinburgh
Chief Scientist and Senior Epidemiologist, UK Biobank

What is UK Biobank?

- Very large prospective cohort study
- 500,000 UK adults age 40-69 at recruitment
- Baseline data on lifestyle, environment, personal & family medical history, physical measures & biological samples
- Follow-up for disease outcomes over 20+ years
- Establish genetic and environmental determinants of common diseases of middle and old age
- Improve prevention, diagnosis and treatment of cancer, heart disease, stroke, arthritis, dementia….

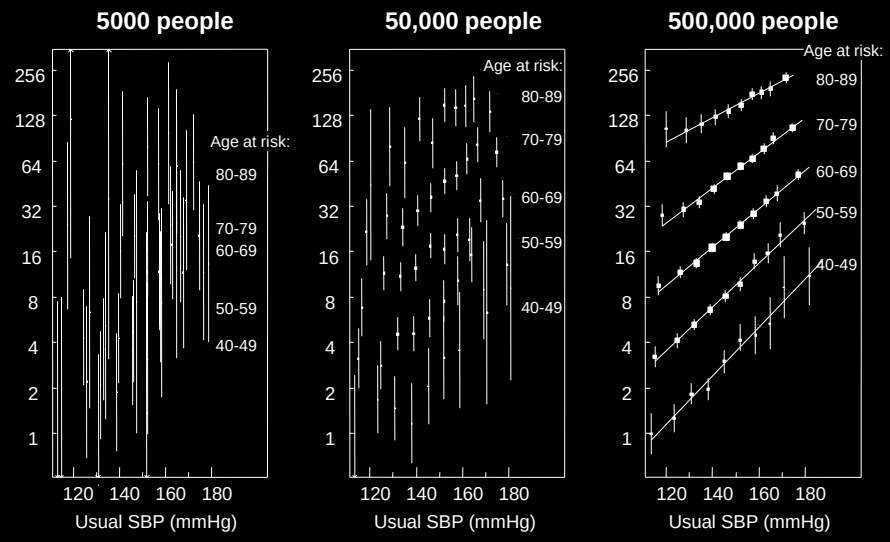
What's so special about it?

- Taking advantage of
  - lessons from previous prospective studies
  - current and future developments in high throughput laboratory technology
- Very large – will generate sufficient numbers of cases of diseases to allow adequately powered nested case-control and case-cohort studies
- Extensive and detailed exposure measures
- Comprehensive follow-up with detailed phenotyping of outcomes
- Open access resource

Key advantages of prospective studies

- Risk factors can be measured before the disease develops
- Associations can be assessed with a range of diseases
- Appropriate controls can be selected from within the same population as the disease cases

Need for observational studies to be LARGE: CHD versus SBP for 5K *vs* 50K *vs* 500K people in the Prospective Studies Collaboration



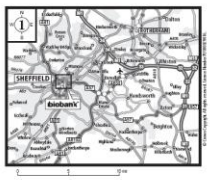
How was it done? - recruitment

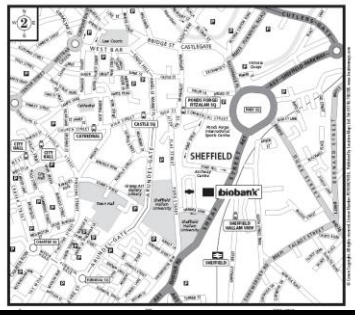
22 recruitment centres based in large UK cities

Assessment centres in rented office space

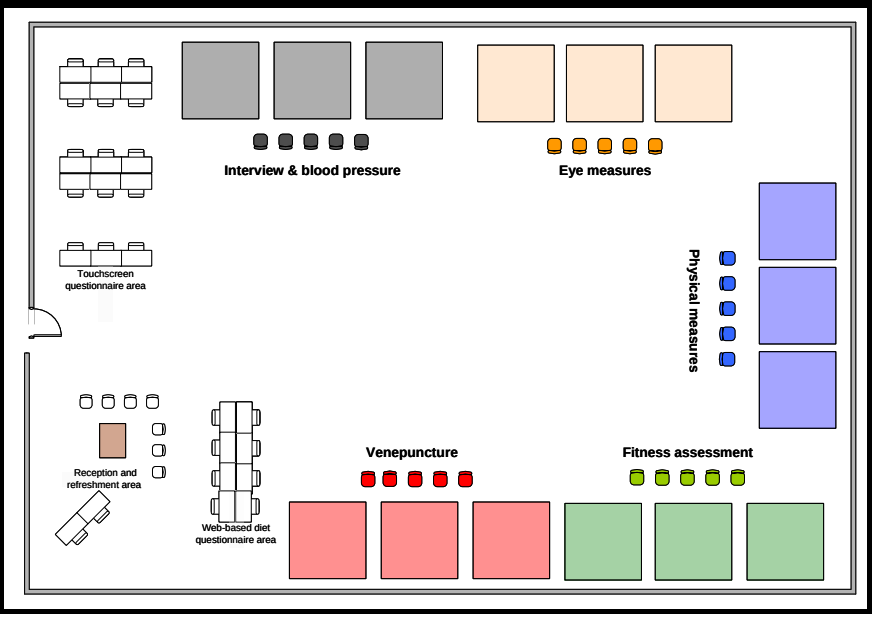
UK Biobank
Sheffield Assessment Centre
Ground Floor
2 Concourse Way
Sheffield Digital Campus
Sheffield
S1 2BJ

Tel 0800 0 276 276
Web www.ukbiobank.ac.uk

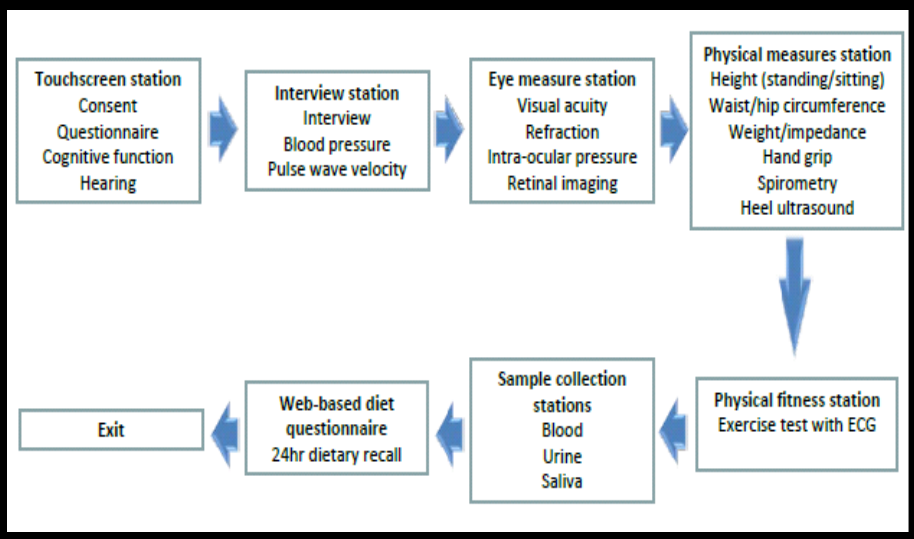




Baseline assessment visit: floor plan



Baseline assessment visit: questions, measures & samples



Strategy for including questions / measures in baseline assessment

- Public health importance of relevant condition
- Likely importance of factors assessed as determinants of subsequent health outcomes
- Reliability and validity of assessment methods
- Lower threshold for inclusion on touchscreen, with limited time for measurements (~30 mins)
- Availability of other sources of information about the factor (e.g. previous medical records; biological samples; internet diet/activity diaries)

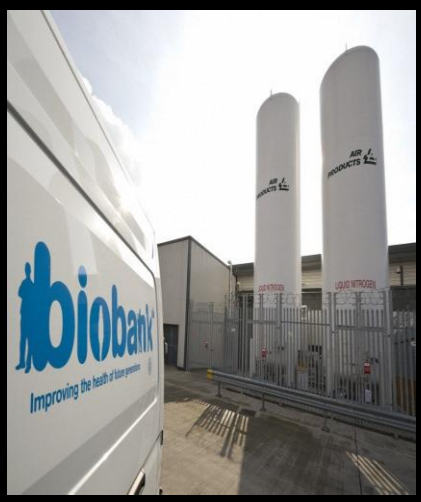
Strategy for sample collection and handling

- Blood, urine & saliva
  - wide range of possible assays
  - wide physiological coverage
- Careful choice of anticoagulants and preservatives
  - widest possible range of potential future uses
- Minimal processing in assessment centres
- Overnight transport to central laboratory for automated blood fractionation and processing
- Storage facilities
  - automated -80^{o} C and back-up liquid nitrogen
  - physical security and reliable tracking of individual samples

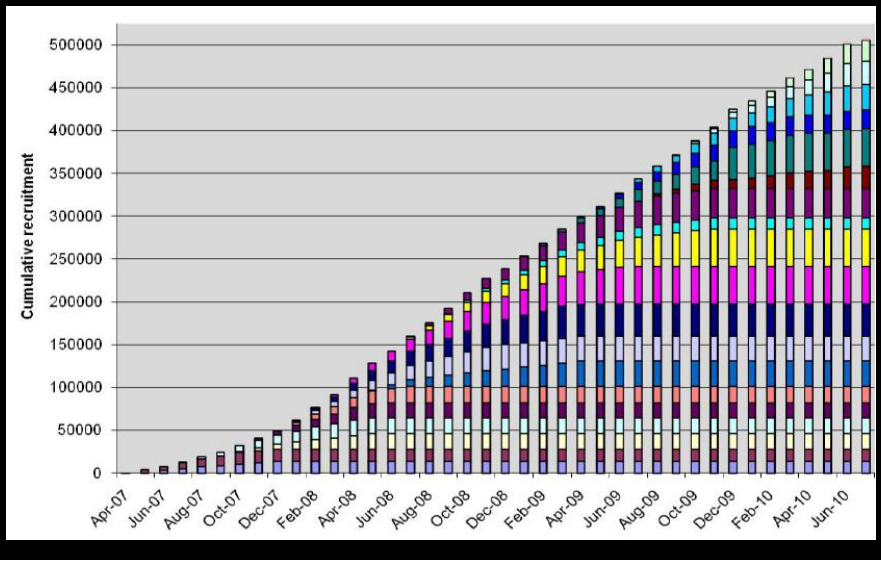
Biological samples

Sample collection tube	Fractions	Potential assays
	Plasma	Plasma proteome & metabonome
EDTA X2	Buffy coat	Genomic DNA
	Red cells	Membrane lipids & heavy metals
Lithium heparin (PST)	Plasma	Plasma proteome & metabonome (without haemolysis)
Silica clot accelerator (SST)	Serum	Serum proteome & metabonome (without haemolysis)
Acid citrate dextrose	Whole blood	DNA from EBV immortalised cell lines B cell transcriptome
EDTA (4ml)	Whole blood	Standard haematological parameters
Urine	Urine	Urine proteome and metabonome
Tempus tube (RNA)	Whole blood	Blood (+ other tissues) transcriptome
Saliva	Mixed saliva	Saliva + mucosal proteome, metabonome, & microbiome

Automated -80°C archive: rapid access to samples



Recruitment of 503,000 participants between April 2007 and June 2010



Participant characteristics

- 46% male
- 57% aged 40-59; 43% aged 60-69
- Less socioeconomically deprived than UK average but all strata represented
- 85% urban
- 94.5% white; 5.5% other (reflects ethnic mix for UK)
- 58% paid employment / self employed
- 89% recruited in England; 7% in Scotland; 4% in Wales

Plans for enhanced phenotyping

- Web-based questionnaires for exposures & outcomes – eg. diet, physical activity, psychological state, cognition
- Repeat assessments of subsample every 4-5 years
- Consent for re-contact studies
- Wrist-worn accelerometers to be mailed to 100,000 participants

……and with further funding….

- Imaging visit in 100,000 participants (to include whole body and brain MRI)
- Standard panel of assays on samples from all participants (e.g. lipids) and DNA extraction

Long term follow-up in UK Biobank

Value of resource depends on:

- Rich baseline data, samples and further planned enhancements
- Comprehensive and detailed follow-up of health of participants

Key advantages for UK Biobank:

- NHS provides majority of healthcare in the UK
- Cohort-wide linkage to a wide range of routine coded health records possible

Prevalent conditions* at recruitment

Condition	Cases (n)
Diabetes	26,000
MI	12,000
COPD	12,000
Stroke	7,000
Breast cancer	11,000
Colorectal cancer	3,000
Prostate cancer	3,000
Lung cancer	441
Rheumatoid arthritis	6,000

** by self report, confirmed by trained interviewer, rounded to nearest 1,000*

Incident outcomes* during follow-up

Condition	2012	2017	2022
Diabetes	10,000	25,000	40,000
MI/CHD death	7,000	17,000	28,000
Stroke	2,000	5,000	9,000
COPD	3,000	8,000	14,000
Breast cancer	2,500	6,000	10,000
Colorectal cancer	1,500	3,500	7,000
Prostate cancer	1,500	3,500	7,000
Lung cancer	1,000	2,000	4,000
Hip fracture	1,000	2,500	6,000
Alzheimer's	1,000	3,000	9,000

**Estimates based on UK age- and sex-specific rates, adjusted for potential healthy cohort effects and losses to follow-up , rounded to nearest 500*

Follow-up through linkage to medical and health-related records

• Written consent for linkage to medical & other health-related records

• Linkage using unique identifier (NHS#/CHI#) and/or name-address-DOB

• Both retrospectively and prospectively to wide range of records

• Detailed approach for record linkage different in Scotland, Wales & England

• Expert group with representatives from Scotland, Wales & England to advise on methods

Progress with follow-up

- Key cohort-wide data sources:
  - Death registrations
  - Cancer registrations
  - Hospital episode data
  - Primary care data
  - Web questionnaire

- Other linkages being considered

  imaging; histopathology reports / specimens; dental, occupational, disability / incapacity benefits; screening; private medical records

Adjudication of outcomes

- Outcomes working group
- Advising on methods for ascertainment, confirmation, and sub-classification of disease outcomes

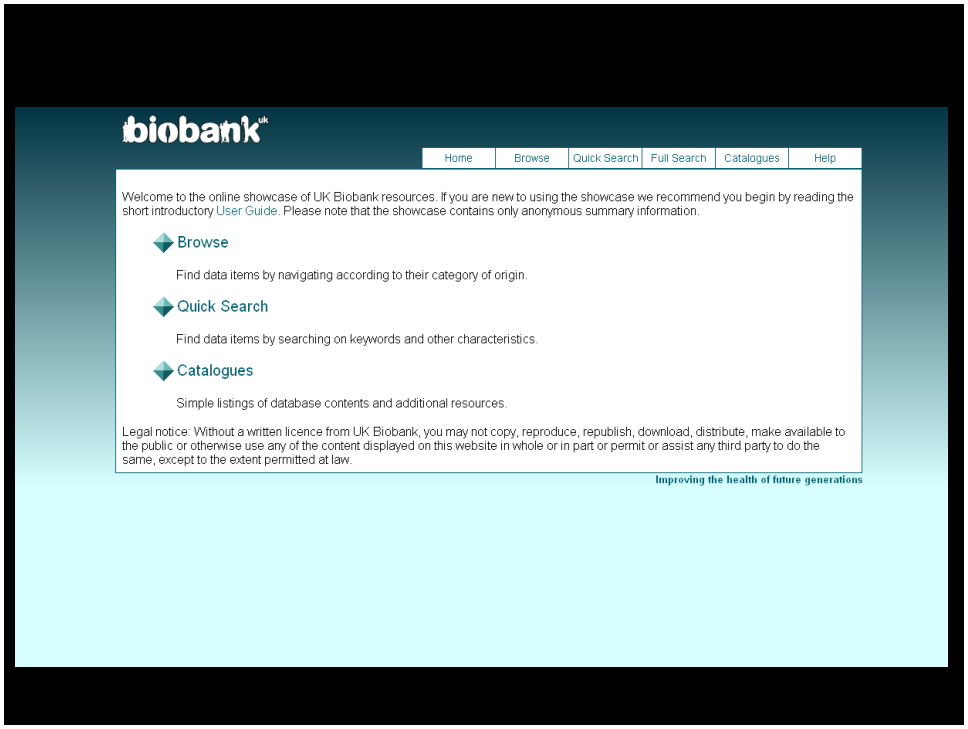
Cancer Diabetes Cardiac outcomes Stroke	Pilots underway
Neurodegenerative outcomes Mental health outcomes Musculoskeletal outcomes Respiratory outcomes	Proposals being developed

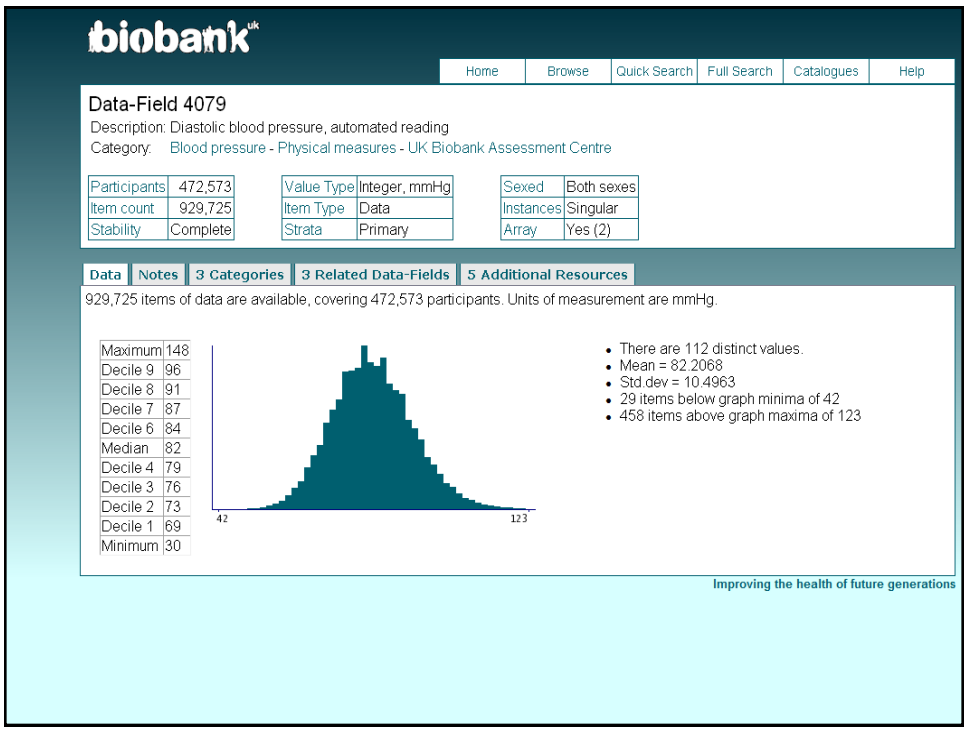
Strategy for access to UK Biobank

- Open access for health-related research in the public good
- UK Biobank scientists have no preferential access
- Review of applications by the co-ordinating centre and Access Sub-Committee
- Charged only for cost of application and provision of the data, assay results or samples the scientists require
- Results to be shared with UK Biobank so that advances can be built on by others

Please visit www.ukbiobank.ac.uk

Access to the resource launched 30 March 2012


biobank uk
Home
Browse
Quick Search
Full Search
Catalogues
Help
Welcome to the online showcase of UK Biobank resources. If you are new to using the showcase we recommend you begin by reading the short introductory User Guide. Please note that the showcase contains only anonymous summary information.
Browse
Find data items by navigating according to their category of origin.
Quick Search
Find data items by searching on keywords and other characteristics.
Catalogues
Simple listings of database contents and additional resources.
Legal notice: Without a written licence from UK Biobank, you may not copy, reproduce, republish, download, distribute, make available to the public or otherwise use any of the content displayed on this website in whole or in part or permit or assist any third party to do the same, except to the extent permitted at law.
Improving the health of future generations


biobank uk
Home
Browse
Quick Search
Full Search
Catalogues
Help
Data-Field 4079
Description: Diastolic blood pressure, automated reading
Category: Blood pressure - Physical measures - UK Biobank Assessment Centre
Participants 472,573
Item count 929,725
Stability Complete
Value Type Integer, mmHg
Item Type Data
Strata Primary
Sexed Both sexes
Instances Singular
Array Yes (2)
Data
Notes
3 Categories
3 Related Data-Fields
5 Additional Resources
929,725 items of data are available, covering 472,573 participants. Units of measurement are mmHg.
Maximum 148
Decile 9 96
Decile 8 91
Decile 7 87
Decile 6 84
Median 82
Decile 4 79
Decile 3 76
Decile 2 73
Decile 1 69
Minimum 30
42
123
There are 112 distinct values.
Mean = 82.2068
Std.dev = 10.4963
29 items below graph minima of 42
458 items above graph maxima of 123
Improving the health of future generations