UK Biobank: a large prospective cohort study into the causes of common complex diseases.

Presentation to participants, 22\textsuperscript{nd} April 2015.
Overall strategy for UK Biobank resource

- 0.5M UK men and women aged 40-69 years
- Extensive baseline questions and physical measures, as well as stored blood, saliva and urine samples that allow many different types of analysis in the future.
- General consent from participants for follow-up through all health-related records.
Power of UK Biobank

- **Prospective**: Assess the full effects of a particular exposure (e.g. smoking) on all health outcomes (e.g. cancer, vascular disease, lung disease)

- **Detail**: Wide range of questions, measures and samples allows assessment of many exposures

- **Size**: Sufficiently large number of people develop a condition to allow reliable assessment of causes, and of interactions between different exposures
Expected numbers of incident disease outcomes during 5, 10 and 15 years of follow-up

<table>
<thead>
<tr>
<th>Condition</th>
<th>2012</th>
<th>2017</th>
<th>2022</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diabetes</td>
<td>10,000</td>
<td>25,000</td>
<td>40,000</td>
</tr>
<tr>
<td>MI/CHD death</td>
<td>7,000</td>
<td>17,000</td>
<td>28,000</td>
</tr>
<tr>
<td>Stroke</td>
<td>2,000</td>
<td>5,000</td>
<td>9,000</td>
</tr>
<tr>
<td>COPD</td>
<td>3,000</td>
<td>8,000</td>
<td>14,000</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>2,500</td>
<td>6,000</td>
<td>10,000</td>
</tr>
<tr>
<td>Colorectal cancer</td>
<td>1,500</td>
<td>3,500</td>
<td>7,000</td>
</tr>
<tr>
<td>Prostate cancer</td>
<td>1,500</td>
<td>3,500</td>
<td>7,000</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>500</td>
<td>2,000</td>
<td>4,000</td>
</tr>
<tr>
<td>Hip fracture</td>
<td>1,000</td>
<td>2,500</td>
<td>6,000</td>
</tr>
<tr>
<td>Alzheimer’s</td>
<td>500</td>
<td>3,000</td>
<td>9,000</td>
</tr>
</tbody>
</table>
Value of size: CHD versus SBP for 5K vs 50K vs 500K people in the Prospective Studies Collaboration (PSC)
Recruitment of the cohort

Scottish cluster
Glasgow, Edinburgh

North/North West cluster
Bury, Manchester, Liverpool, Sheffield

South West and Wales cluster
Cardiff, Bristol, Swansea

Midlands cluster
Nottingham, Birmingham, Oxford

London cluster
Bart’s, East London, Hounslow, Reading

North Eastern cluster
Newcastle, Middlesbrough, Leeds

20% urban, 80% rural

- 46% male, 54% female
- 57% aged 40-59, 43% aged 60-69
- All strata for UK population represented
- 85% urban
- 94.5% white, 5.5% other ethnicities
Baseline assessment visit (with enhancements)

- **Touchscreen station**
  - Consent, Touchscreen questionnaire,
  - Cognitive function, hearing test

- **Interview station**
  - Interview
  - Interview and blood pressure

- **Interview and blood pressure station**
  - Vascular reactivity

- **Eye measure station**
  - Visual acuity and refraction,
  - Intra-ocular pressure,
  - Retinal photography,
  - Optical Coherence Tomography

- **Physical measures station**
  - Hand grip,
  - Height (standing, sitting),
  - Waist/hip circumference,
  - Weight/impedence,
  - Spirometry,
  - Heel ultrasound

- **Sample collection station**
  - Blood, urine
  - Saliva, blood for RNA

- **Physical fitness station**
  - Exercise test with ECG

- **Physical function test**

- **Exit**
We collected 50mls blood and a urine and saliva sample....

<table>
<thead>
<tr>
<th>K$_2$EDTA</th>
<th>Silica</th>
<th>Li/Hep</th>
<th>Acid citrate dextrose</th>
<th>K$_2$EDTA</th>
<th>Tempus tube</th>
</tr>
</thead>
</table>

- Plasma, Buffy coat, Red cells
- Serum
- Plasma
- PBLs
- Urine analytes
- RNA
- Saliva
...that were processed at our site in Cheadle.

- 700 Participants per day
- 4900 Vacutainers per day
- 21,000 1ml Samples per day
- 14 million 1ml aliquots
UK Biobank – main site and archive
UK Biobank – back-up site
Current activities

- **Provision of access to resource**
- Follow up of the health of our participants
- Enhancing the study
  - Web-based diet questionnaires
  - Assessment of seven day physical activity
  - Measuring markers in blood and urine
  - **Genotyping of all 500,000 participants**
  - Imaging of 100,000 participants
UK Biobank is supporting a range of research questions
Follow up of half a million people

- All participants are registered with a GP
- NHS provides the great majority of healthcare in the UK.
- National datasets about healthcare and people’s outcomes exist.

<table>
<thead>
<tr>
<th>Comprehensive, scalable and affordable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Register of deaths (and causes)</td>
</tr>
<tr>
<td>Cancer registrations</td>
</tr>
<tr>
<td>Hospital data</td>
</tr>
<tr>
<td>Primary care data</td>
</tr>
</tbody>
</table>
Follow up of half a million people – web-based tools

**Affordable, repeat administrations, efficiency of data collection and analysis**

- Dietary questionnaire.
- Can assess conditions that are very hard to find out about through linking to health records:
  - Cognitive function
  - Occupation
  - Mental health
  - Quality of life
• The human genome is comprised of almost 3 billion bases (or base pairs)
• Each human chromosome carries a unique set of genes which produce specific proteins
• We each have about 30,000 genes (2% of the DNA)
• Genetic diversity, together with environmental influence, is what makes each human being different from every other
• Individuals are 99.9% identical at the DNA sequence level
• It is the variation that causes the differences between us and drives evolutionary change
• The most common form of variation is a SNP
SNP – single nucleotide polymorphism

Person A
GGATCGATTTCCGATTTAGCGATCGATTTACGTTCGATCG

Person b
GGATCGATTTTGGATTTAGCGATCGATTTACGTTCGATCG
Genotyping at UK Biobank

- High quality genomic DNA is extracted from the white cells on automated systems at Cheadle.
- DNA is genotyped at our partners at Affymetrix.
- The genotypes are “called” and then checked in Oxford.
- The data are enormously enhanced through imputation.
Wellcome Trust Case Control Consortium: 2,000 cases of 7 different conditions and 3,000 controls (Nature 2007)
Overall value of large scale imaging

- Well characterised group of people with linkage to health outcomes
- Increased power through size and measurement precision
- Unique value of multi-modal imaging in large numbers of individuals that can be linked with the wealth of data already in UK Biobank.
- This will be the largest study of its kind in the world.
Imaging at UK Biobank
Abdominal MRI

- Accurate measure of fat distribution across abdominal compartments – link to risk/health outcome
  - More informative than BMI, WHR
  - Volume and distribution of fat and its link to metabolic disease.
- Liver and muscle fractions as predictors of outcomes
- Interactions between body phenotype and genetic/lifestyle factors
  - No large in depth studies to date
  - Impact of physical activity, calorific reduction

Leads – Tony Goldstone and Jimmy Bell
BMI = 23.50 kg/m²; TAT = 13.19 l
BMI = 23.52 kg/m²; TAT = 21.79 l
BMI = 23.53 kg/m²; TAT = 17.33 l
BMI = 23.57 kg/m²; TAT = 21.43 l
BMI = 23.88 kg/m²; TAT = 16.84 l
BMI = 24.27 kg/m²; TAT = 24.11 l
BMI = 24.29 kg/m²; TAT = 14.33 l
BMI = 24.07 kg/m²; TAT = 12.40 l
BMI = 23.62 kg/m²; TAT = 26.17 l

Data from Prof. Jimmy Bell and Tony Goldstone – Imperial College
Appearances can be deceptive

Phenotype: “thin on the outside, fat on the inside” = TOFI

Data from Prof. Jimmy Bell and Tony Goldstone – Imperial College
Influence of Dieting

Following a weight loss of 33 kg

Data from Prof. Jimmy Bell and Tony Goldstone – Imperial College
Summary

• UK Biobank is a globally-unique resource of data and samples linked to health outcomes
• It will be used to understand the causes of common diseases and of health
• As data are added it becomes increasingly valuable
• New technologies are offering the potential for ground-breaking, affordable discoveries that could not have been envisaged when the study was first proposed
• The way that UK Biobank has been delivered has been recognised nationally and internationally
• It is something of which the North West and the UK can be proud