The UK Biobank Resource

Tom Littlejohns
Nuffield Department of Population Health
University of Oxford
What is UK Biobank and why is it important?

Common exposures (i.e. smoking) with big effects (i.e. cancer) are fairly well known, but….

The causes of most disease and ill health are combination of things – lifestyle, environment, physical condition, genetics etc. (we need a lot of data)

Small effects from common exposures or large effects from rare exposures difficult to detect (we need a lot of people)

It’s complicated! (we need to collaborate)

UK Biobank is a unique, ground-breaking resource

• A lot of data – including data collected at the start and ongoing collection
• Large
• Accessible to scientists worldwide
Recruitment

Between 2006-2010, mailed invite to those aged 40-69 years old, registered with the NHS and lived within ~25 miles of 1 of 22 assessment centres.

- Invitees: 9,238,453
- Accepted: 576,926
- Consented: 503,317

Newcastle

~37,000
Jan 2008 – March 2009
Data collection at recruitment

Touchscreen questionnaire

- Demographics
- Lifestyle
- Environment
- Early life
- Family history
- Psychosocial
- Health
- Hearing test
- Cognitive test
Data collection at recruitment

- Demographics
- Lifestyle
- Environment
- Early life
- Family history
- Psychosocial
- Health
- Hearing test
- Cognitive test

More detail on
- Early life
- Employment
- Medical conditions
- Medications
- Operations
Data collection at recruitment

- Touchscreen questionnaire
  - Demographics
  - Lifestyle
  - Environment
  - Early life
  - Family history
  - Psychosocial
  - Health
  - Hearing test
  - Cognitive test

- Verbal medical interview
  - More detail on
    - Early life
    - Employment
    - Medical conditions
    - Medications
    - Operations

- Physical measures
  - Blood pressure
  - Arterial stiffness
  - Hand grip
  - BMI/weight
  - Spirometry
  - Heel US
  - Eye
  - Exercise ECG
Data collection at recruitment

- Touchscreen questionnaire
- Verbal medical interview
- Physical measures
- Sample Collection

**Touchscreen questionnaire:**
- Demographics
- Lifestyle
- Environment
- Early life
- Family history
- Psychosocial
- Health
- Hearing test
- Cognitive test

**Verbal medical interview:**
- More detail on
  - Early life
  - Employment
  - Medical conditions
  - Medications
  - Operations

**Physical measures:**
- Blood pressure
- Arterial stiffness
- Hand grip
- BMI/weight
- Spirometry
- Heel US
- Eye
- Exercise ECG

**Sample Collection:**
- Blood (45 ml)
- Urine (9ml)
- Saliva (2.5ml)
Genetics and biomarkers

More than 15 million aliquots

‘Single nucleotide polymorphisms’ or SNPs – ~10 million in human genome

SNPs are variation in DNA

A A T C G
A A T G G

PS. I might have massively oversimplified the genetics of eye colour with this example.
Genetics and biomarkers

820,000 SNPs directly measured

Exome and whole genome sequencing underway

Samples also used to measure biomarkers – a biological indicator of a certain process

- Diagnostic markers
- Risk factors for disease
### Infectious agents

<table>
<thead>
<tr>
<th>Pathogen (acute disease)</th>
<th>% sero-positive in 10,000 (pilot)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSV-1 (oral herpes)</td>
<td>69.8%</td>
</tr>
<tr>
<td>HSV-2 (genital herpes)</td>
<td>16.2%</td>
</tr>
<tr>
<td>VZV (chickenpox/shingles)</td>
<td>92.5%</td>
</tr>
<tr>
<td>EBV (glandular fever)</td>
<td>94.7%</td>
</tr>
<tr>
<td>CMV</td>
<td>58.2%</td>
</tr>
<tr>
<td>HHV-6 (roseola)</td>
<td>90.8%</td>
</tr>
<tr>
<td>HHV-7 (roseola)</td>
<td>94.7%</td>
</tr>
<tr>
<td>KSHV (Kaposi’s sarcoma)</td>
<td>8.1%</td>
</tr>
<tr>
<td>HPV-16</td>
<td>4.4%</td>
</tr>
<tr>
<td>HPV-18</td>
<td>2.7%</td>
</tr>
<tr>
<td>JC polyomavirus</td>
<td>57.5%</td>
</tr>
<tr>
<td>BK polyomavirus</td>
<td>95.4%</td>
</tr>
<tr>
<td>MCV polyomavirus</td>
<td>66.7%</td>
</tr>
<tr>
<td>HBV</td>
<td>2.5%</td>
</tr>
<tr>
<td>HCV</td>
<td>0.3%</td>
</tr>
<tr>
<td>HIV (AIDS)</td>
<td>0.2%</td>
</tr>
<tr>
<td>HTLV-1</td>
<td>1.6%</td>
</tr>
<tr>
<td><em>H. pylori</em></td>
<td>31.5%</td>
</tr>
<tr>
<td><em>C. trachomatis</em></td>
<td>21.4%</td>
</tr>
<tr>
<td><em>T. Gondii</em> (toxoplasmosis)</td>
<td>28.0%</td>
</tr>
</tbody>
</table>
Online Questionnaires

2011/12 - 24hr diet recall x4

2014/15 - Cognitive function

2015 - Occupational history

2016 - Mental health

2017 - Digestive health

Email for 2/3rds
Every 6-12 months
Useful for in-depth detail on outcomes

Next few years...

Pain

Food preferences

Sleep

Quality of life
Accelerometer

Objective measure of physical activity

Worn by 100,000 participants

Feb 2013 – Dec 2015

Wear for 7 days

Repeat in 2,500 4 times over 1 year for seasonal variability
Sleep

- Evening person
- Morning person

Bicycling

- Cycles to work
- Doesn't cycle to work

Plus multimodal imaging assessment of 100,000 participants...more later
Medical records

All participants:
registered with a GP in the NHS
consented to linkage to health-related records

NHS provides majority of healthcare in UK

National datasets about healthcare & health outcomes exist

….so link to these datasets…..

Currently death registry, cancer registry and hospital inpatient records…soon primary care
## Some examples of expected number of outcomes

<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>Heart attack</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>Chronic obstructive lung disease</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>Hip fracture</td>
<td>1,000</td>
<td>2,500</td>
<td>6,000</td>
</tr>
<tr>
<td>Alzheimer’s disease</td>
<td>1,000</td>
<td>3,000</td>
<td>9,000</td>
</tr>
</tbody>
</table>
Good news for UKB ppts however…

Mortality rates ~50% lower in UK Biobank compared to the general population

![Graph showing mortality rates for men and women in UK Biobank and general population.](attachment:image)
Number of all Applications Received by Year (excluding draft)
2012 to 11th November 2018 (Total = 1,519)
Research output

Number of publications

Year

2013  2014  2015  2016  2017  2018
Some examples

Frequency and phenotype of type 1 diabetes in the first six decades of life: a cross-sectional, genetically stratified survival analysis from UK Biobank

Nicholas J Thomas, Samuel E Jones, Michael N Weedon, Beverley M Shields, Richard A Oram*, Andrew T Hattersley*

Genetic risk, incident stroke, and the benefits of adhering to a healthy lifestyle: cohort study of 306,473 UK Biobank participants

Loes CA Rutten-Jacobs,1,2 Susanna C Larsson,3 Rainer Malik,4 Kristiina Rannikmäe,5,6 Cathie L Sudlow,5,6,7 Martin Dichgans,4,8,9 Hugh S Markus,2 Matthew Traylor2
Thank you for listening!

www.ukbiobank.ac.uk