UK Biobank: lessons in data collection

Naomi Allen
Senior Epidemiologist, UK Biobank
University of Oxford
Naomi.allen@ndph.ox.ac.uk
UK Biobank is a major national health resource designed to improve the prevention, diagnosis and treatment of a wide range of illnesses – including cancer, cardiovascular disease, diabetes, arthritis, osteoporosis, eye disorders, depression and dementia.
UK Biobank in a nutshell

- A large prospective cohort study
- 500,000 UK adults age 40-69 at recruitment, 2006-2010
- Baseline data on a wide range of lifestyle factors, environment, medical history, physical measures & biological samples
- Consent for follow-up through health records for all types of health research
- Open-access to researchers worldwide (academia & industry)
Recruitment into UK Biobank

- Using individual GP practices for recruitment purposes impractical
- Direct mailing of invitations using contact details held by the NHS
- Invited 9.2 million; 503,000 joined
Rented office space as an assessment centre

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
Production line approach to assessment visit

improved throughput; usb stick transfer; qualified staff at each stage
Recruitment of 500,000 participants: April 2007 - July 2010
Strategy for inclusion of questions

- Importance determinants of subsequent health outcomes
- Reliability and validity of assessment methods
- Time taken to obtain the information
- Availability of alternative sources of information about the factor (e.g. previous medical records; biological samples)
# Topics covered by touchscreen and interview

## Touchscreen

<table>
<thead>
<tr>
<th>Topic</th>
<th>Median time (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socio-demographics</td>
<td>1.7</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>0.1</td>
</tr>
<tr>
<td>Employment</td>
<td>1.4</td>
</tr>
<tr>
<td>Physical activity</td>
<td>4.4</td>
</tr>
<tr>
<td>Smoking</td>
<td>1.5</td>
</tr>
<tr>
<td>Diet</td>
<td>4.5</td>
</tr>
<tr>
<td>Alcohol</td>
<td>1.1</td>
</tr>
<tr>
<td>Sleep</td>
<td>1.2</td>
</tr>
<tr>
<td>Sun exposure</td>
<td>1.3</td>
</tr>
<tr>
<td>Environmental exposures</td>
<td>1.0</td>
</tr>
<tr>
<td>Early life factors</td>
<td>0.8</td>
</tr>
<tr>
<td>Family history of common diseases</td>
<td>1.6</td>
</tr>
<tr>
<td>Reproductive history &amp; screening (women)</td>
<td>2.4</td>
</tr>
<tr>
<td>Reproductive history &amp; screening (men)</td>
<td>0.8</td>
</tr>
<tr>
<td>Sexual history</td>
<td>0.4</td>
</tr>
<tr>
<td>Medical history</td>
<td>3.7</td>
</tr>
<tr>
<td>Noise exposure</td>
<td>1.0</td>
</tr>
<tr>
<td>Psychological status</td>
<td>4.5</td>
</tr>
<tr>
<td>Cognitive function tests</td>
<td>10.0</td>
</tr>
<tr>
<td>Hearing speech-in-noise test</td>
<td>8.0</td>
</tr>
</tbody>
</table>

**Total time** 52.5

## Nurse Interview

<table>
<thead>
<tr>
<th>Topic</th>
<th>Median time (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medical history</td>
<td>3.1</td>
</tr>
<tr>
<td>Occupation</td>
<td>0.4</td>
</tr>
<tr>
<td>Other</td>
<td>0.6</td>
</tr>
</tbody>
</table>

**Total time** 4.5

---

*Electronic signature pad*
Example of checks on touchscreen questions

How many years have you lived at your current address?

80 years

7  8  9  Clear
4  5  6  Less than a year
1  2  3  Do not know
0  Prefer not to answer
How many years have you lived at your current address?

Please alter your answer
The value (80) you entered is too high.
Please re-enter.

Continue
Interview to confirm responses and reduce error

- Response to previous touchscreen questions shown
- Validation by nurse
- Immediate coding using drop-down menus for:
  - disease history
  - medications
All 500,000 participants
- Blood pressure & heart rate
- Height (standing/seated)
- Waist/hip circumference
- Weight/impedance
- Spirometry
- Heel ultrasound

Subset: 175,000 participants
- Hearing test
- Vascular reactivity

Subset: 120,000 participants
- Visual acuity, refractive index & intraocular pressure

Subset: 85,000 participants
- Retinal images & optical coherence tomograms
- Fitness test & ECG limb leads

Measures based on relevance, reliability and feasibility

All data directly downloaded into central database
Direct automated data entry of all raw data, results and meta-data

**Spirometry data**

Device ID = 01823
- Result set 0
  - FVC = 5.12 litres
  - FEV1 = 4.4 litres
  - PEF = 667 litres/min
  - (ranked 0)
- Result set 1
  - FVC = 5 litres
  - FEV1 = 4.34 litres
  - PEF = 658 litres/min
  - (ranked 1)

![Spirometry Results Graph](chart.png)
Direct data entry of physical measure

Direct automated data entry of all raw data, results and meta-data

Visual acuity

Optical coherence tomography scan
### Statistical monitoring tools

- Daily reports on results per device, assessor, centre
- Enables rapid changes to be made (e.g., recalibration of device; retraining of staff) to ensure no bias in data capture methods
• Blood, urine and saliva collected
• Choice of anticoagulants and preservatives to allow widest possible range of potential future uses
• Pilot studies of sample processing procedures
• Overnight transport to central lab for automated blood fractionation and processing
• Storage facilities (automated -80°C and back-up liquid nitrogen) to provide physical security
• Reliable tracking of individual samples through LIMS
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
<table>
<thead>
<tr>
<th>Data type</th>
<th>Diagnostic coding systems</th>
<th>Country</th>
<th>Data providers</th>
<th>Data available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Certified deaths by cause</td>
<td>ICD-10</td>
<td>England</td>
<td>NHS Digital &amp; ONS</td>
<td>2006-2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wales</td>
<td>Central Register &amp; ISD</td>
<td>2006-2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scotland</td>
<td>Central Register &amp; ISD</td>
<td>2006-2016</td>
</tr>
</tbody>
</table>
## National cohort-wide linked healthcare datasets

<table>
<thead>
<tr>
<th>Data type</th>
<th>Diagnostic coding systems</th>
<th>Country</th>
<th>Data providers</th>
<th>Data available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Certified deaths by cause</td>
<td></td>
<td>England</td>
<td>NHS Digital &amp; ONS</td>
<td>2006-2016</td>
</tr>
<tr>
<td></td>
<td>ICD-10</td>
<td>Wales</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scotland</td>
<td>Central Register &amp; ISD</td>
<td>2006-2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wales</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scotland</td>
<td>Central Register &amp; ISD</td>
<td>1957-2016</td>
</tr>
<tr>
<td>Data type</td>
<td>Diagnostic coding systems</td>
<td>Country</td>
<td>Data providers</td>
<td>Data available</td>
</tr>
<tr>
<td>----------------------------</td>
<td>---------------------------</td>
<td>-----------</td>
<td>-----------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Certified deaths by cause</td>
<td>ICD-10</td>
<td>England</td>
<td>NHS Digital &amp; ONS</td>
<td>2006-2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wales</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scotland</td>
<td>Central Register &amp; ISD</td>
<td>2006-2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wales</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scotland</td>
<td>Central Register &amp; ISD</td>
<td>1957-2016</td>
</tr>
<tr>
<td>Hospital admissions</td>
<td>ICD-9 &amp; 10 and OPCS4</td>
<td>England</td>
<td>NHS Digital (HES)</td>
<td>1996-2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wales</td>
<td>SAIL (PEDW)</td>
<td>1998-2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Scotland</td>
<td>ISD, NSS (SMR)</td>
<td>1981-2016</td>
</tr>
</tbody>
</table>
Follow-up of health outcomes

- Cancer registrations (79,000)
- Hospital admissions (400,000)
- Death registrations (14,000)
- Online Questionnaires (200,000)

- Primary care data (pending)
- Screening programmes Potentially feasible
- Pathology records Potentially feasible
UK Biobank is a national and international health resource with unparalleled research opportunities, open to all bona fide health researchers. UK Biobank aims to improve the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses – including cancer, heart diseases, stroke, diabetes, arthritis, osteoporosis, eye disorders, depression and forms of dementia. It is following the health and well-being of 500,000 volunteer participants and provides health information, which does not identify them, to approved researchers in the UK and overseas, from academia and industry. Scientists, please ensure you read the background materials before registering. To our participants, we say thank you for supporting this important resource to improve health. Without you, none of the research featured on this website would be possible.

Read more about Biobank UK