Introduction to the Millennium Cohort Study

29 June 2022
1pm-2.30pm
Plan for today

- About CLS
- Overview of MCS
- Examples of research
- Data, by theme
- Data enhancements and innovations
- Sample design and weights
- Available resources and data access
- Plan for future sweeps
- Q&A to the panel*
Introductions

- Vanessa Moulton, Senior Researcher
- Morag Henderson, Associate Professor
- Nicolas Libuy Rios, Research Fellow
- Rachel Rosenberg, Research Data Manager
- Lucy Haselden, Survey Manager
About CLS

• The Centre for Longitudinal Studies is home to four* national longitudinal cohort studies, which follow the lives of tens of thousands of people

• Each of our four studies follows large, nationally representative groups of people born in a given time period (week, year etc)

• By collecting information from the same people over time, as they live their lives, our studies are powerful resources for answering important research questions
Timeline of the studies

- 1958 National Child Development Study
- 1970 British Cohort Study
- Next Steps
- Millennium Cohort Study
Timeline of the studies

1958 National Child Development Study

1970 British Cohort Study

Millennium Cohort Study follows around 19,000 children born in the UK in 2000-02

Next Steps

2000, 2020
## COVID-19 surveys

<table>
<thead>
<tr>
<th>Wave</th>
<th>Dates</th>
<th>Context</th>
<th>Mode</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 1</td>
<td>May 2020</td>
<td>First national lockdown</td>
<td>Web</td>
</tr>
<tr>
<td>Wave 2</td>
<td>Sept/Oct 2020</td>
<td>Eased restrictions</td>
<td>Web</td>
</tr>
<tr>
<td>Wave 3</td>
<td>Feb/Mar 2021</td>
<td>Third national lockdown</td>
<td>Web → Telephone</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study</th>
<th>Age in 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCS</td>
<td>19</td>
</tr>
<tr>
<td>Next Steps</td>
<td>30</td>
</tr>
<tr>
<td>BCS70</td>
<td>50</td>
</tr>
<tr>
<td>NCDS</td>
<td>62</td>
</tr>
</tbody>
</table>

[https://cls.ucl.ac.uk/covid-19-survey/](https://cls.ucl.ac.uk/covid-19-survey/)
Other CLS-Affiliated Studies

Early Life Cohort Feasibility Study
- A two-year ESRC funded project which started in April 2021 that will test the feasibility of a new UK-wide birth cohort study

Children of the 2020s Study
- A new nationally representative birth cohort study of babies in England which has been commissioned by the Department for Education (DfE)
- It will include babies born September - November 2021, and seeks to recruit over 8,500 families in mid 2022

COVID Social Mobility and Opportunities (COSMO) study
- The study began in 2021 with a representative sample of young people in Year 11 across England
- UKRI funded, led by researchers from the UCL Centre for Education Policy and Equalising Opportunities and the Sutton Trust, in collaboration with CLS
About the Millennium Cohort Study (MCS)
About the UK Millennium Cohort Study

- Millennium Cohort Study (MCS) is a birth cohort study of around 19,000 individuals born in 398 areas of the UK in 2000-2002

- Highly multidisciplinary survey, designed to capture the influence of early family context on child development and outcomes throughout childhood, into adolescence and through adulthood

- Collects information about cohort member’s physical, socio-emotional, cognitive and behavioural development, as well as parental health, parents’ employment and education, income and poverty; housing; neighbourhood among other topics


The MCS population

All children born between
1/09/2000 and 31/8/2001 (England and Wales)
24/11/2000 and 11/1/2002 (Scotland and Northern Ireland)

AND

alive and living in the UK at age 9 months, and eligible to receive Child Benefit at that age

Oversamples for:
- Wales, Scotland and Northern Ireland
- areas with high child poverty and,
- in England, areas with higher ethnic minority populations
### Overview of MCS timeline and main survey

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Resident Parents</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household questionnaire</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Main resident parent - Interview and questionnaire self-completion</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>(x)</td>
</tr>
<tr>
<td>Main parents’ partner - Interview and questionnaire self-completion</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>(x)</td>
</tr>
<tr>
<td>Questionnaire self-completion (in interview)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Main interview - interview and online questionnaire</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older siblings</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teachers</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Overview of MCS cohort members (CM) data

<table>
<thead>
<tr>
<th>Time</th>
<th>9m</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>11</th>
<th>14</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><strong>Physical measurements</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><strong>Cognitive assessments</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td><strong>Activity monitor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Time use record</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Saliva for DNA &amp; genotyping</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Consent for linking admin data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Physical measurements
- Cognitive assessments
- Activity monitor
- Time use record
- Saliva for DNA & genotyping
- Consent for linking admin data

**Data collection periods:**
- 2001
- 2004
- 2006
- 2008
- 2012
- 2015
- 2018

**Health records:**
- Education records
- Parents’ and CM’s economic records
- Police and criminal justice
MCS achieved sample (Number of families)

Longitudinal participation

- 9 months: 18,552
- 3 years: 15,590
- 5 years: 15,246
- 7 years: 13,857
- 11 years: 13,827
- 14 years: 11,726
- 17 years: 10,625
For each family, there are also response rates for:
- Cohort member (singleton, twin and triplets)
- Main parent
- Main parents’ partner
Examples of research using MCS
After-school clubs 'boost poorer pupils' results'

By Katherine Seligren
Education reporter

03 20 April 2016

< image of children participating in an after-school club with a teacher holding a whiteboard.

HIGHWAYSTARZ PHOTOGRAPHY
Children over seven suffer most from divorce fallout

Fathers leaving the family home after a child is older than three caused a higher rate of emotional problems.

GETTY IMAGES
The Guardian

Children living in more costly homes have fewer mental health problems - study

UCL findings highlight key role that value of family home plays in child development
More than a third of UK teenagers overweight or obese

One in five young people obese at 17 and one in seven (14%) were overweight, with those from poorest families at greatest risk

By Patrick Sawyer, SENIOR NEWS REPORTER
17 December 2020 - 12:01am
Data, by theme
Themes

- Family and relationships
- Education and cognition
- Physical health
- Mental health and well-being
- Behaviours, attitudes and identity

...but there are many more
Family and Relationships
Family and relationships

- Who is in the household
- Relationship to cohort member
- Age/number of siblings
- Biological, step, adoptive parents
- Age of parents when the cohort member was born
- Pregnancy, labour and delivery
- Family context, e.g. marital status and partnership formation / dissolution, language spoken, ethnicity
- Socio-economic context of family e.g. parents’ employment, social class, income, housing
## Family and relationships

<table>
<thead>
<tr>
<th>Early years</th>
<th>Primary school years</th>
<th>Adolescence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parents relationship (9m)</td>
<td>Parents relationship (5)</td>
<td>YP Social Provisions Scale (14,17)</td>
</tr>
<tr>
<td>Social support (P, 9m, 3)</td>
<td>Parenting activities (5,7,11)</td>
<td>Parenting activities (14)</td>
</tr>
<tr>
<td>Father involvement (9m, 3)</td>
<td>Discipline (M, 5,7)</td>
<td>Relationship with parents (14,17)</td>
</tr>
<tr>
<td>Grandparents (9m, 3)</td>
<td>Partners relationship with CM (11)</td>
<td>Friends (14)</td>
</tr>
<tr>
<td>Maternal attachment (9m)</td>
<td>Relationship with siblings (5,7)</td>
<td>Romantic relationships (14.17)</td>
</tr>
<tr>
<td>Child-parent relationship – Pianta SF (3)</td>
<td>Friends (7,11)</td>
<td>Sexual experiences (14,17)</td>
</tr>
<tr>
<td>Parenting activities (3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parenting values (3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discipline (M, 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child environment – Home_Sf (Int., 3)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Education and cognitive measures
Educational attainment

- Parental education
- Early education provision (5,7)
- Foundation Stage Profile (and equiv.) (5)
- Teacher assessment (7,11)
  - Abilities, groupings/sets, behaviour, language, SEN etc.
- Key stage 1, 2, and GCSE (National Pupil Database)
  - GCSE or equiv. self-report
- Study intentions (11, 14)
- Higher education
  - Degree subject (17)
  - University type (17)
# MCS Cognitive Assessments/Skills

<table>
<thead>
<tr>
<th>Developmental milestones</th>
<th>9 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>School readiness (BSRA-R)</td>
<td>3</td>
</tr>
<tr>
<td>Colours, letters, numbers, sizes, comparisons, shapes</td>
<td></td>
</tr>
<tr>
<td>Verbal skills</td>
<td></td>
</tr>
<tr>
<td>BAS II Naming Vocabulary</td>
<td>3, 5</td>
</tr>
<tr>
<td>BAS II Word Reading</td>
<td>7</td>
</tr>
<tr>
<td>APU Vocabulary Test</td>
<td>14 (and parents)</td>
</tr>
<tr>
<td>Verbal reasoning</td>
<td>11</td>
</tr>
<tr>
<td>BAS II Verbal Similarities</td>
<td></td>
</tr>
<tr>
<td>Non-verbal reasoning</td>
<td>5</td>
</tr>
<tr>
<td>BAS II Picture Similarities</td>
<td></td>
</tr>
<tr>
<td>Mathematics and numeracy</td>
<td>7</td>
</tr>
<tr>
<td>NFER Progress in Maths</td>
<td>17</td>
</tr>
<tr>
<td>Number Analogies (GL)</td>
<td></td>
</tr>
<tr>
<td>Visual/spatial processing</td>
<td>5, 7</td>
</tr>
<tr>
<td>BAS II Pattern Construction</td>
<td></td>
</tr>
<tr>
<td>Decision making (CANTAB)</td>
<td>11, 14</td>
</tr>
<tr>
<td>Cambridge Gambling Task</td>
<td></td>
</tr>
<tr>
<td>Memory (spatial working) (CANTAB)</td>
<td>11</td>
</tr>
</tbody>
</table>

**BSRA-R** (Bracken School Readiness Assessment – Revised)

**BAS** (British Ability Scales)

**APU** (Applied Psychology Unit)

**NFER** (National Foundation for Education Research)

**GL Assessment**
Physical Health
Physical Health in MCS – measurements and samples

- **Anthropometrics:**
  - Height, Weight and BMI (3,5,7,11,14,17)
  - Waste circumference (5,7)
  - Body fat (11,14,17)

- **Samples:**
  - Saliva sample – DNA extraction and genotyping (14)
  - Oral fluid sample - assays used to detect variety of infections (3)
  - Milk teeth - test for exposure to lead in the environment. (7)

- **Activity monitor** (7,14)
  - Actigraph – every day for 7 days during waking hours (7)
  - Geneactiv accelerometer – 1 weekday & 1 weekend day measured physical and sedentary activity (14)
    - Activity monitor was paired with time use diary: sample and day selection the same
Assessment of Physical Health

- Parents health
- Parent/self-assessed of CM’s general health (3, 5, 7, 11, 14, 17)
- Medical conditions/Long standing illness (9m, 3, 5, 7, 11, 14, 17)
- Adolescence:
  - Pubertal Development Scale (adapted) (14)
  - Periods (14, 17)
  - Sex, contraception, pregnancy, STIs (17)
- Linked administrative health records:
  - Hospital Episodes Statistics (in Scotland, Wales)
Health behaviours

- Physical activity (7,11,14,17)
- Diet
  - YP Eating Choices Index (4 item index to measure healthiness of diet) (14,17)
  - Dieting (14,17)
  - Body image (14,17)
- Sleep (14,17)
- Drugs (14,17)
- Smoking (11,14,17)
- Drinking (11, 14, 17)
Mental health and well-being
## Mental health and well-being I

<table>
<thead>
<tr>
<th>Parent</th>
<th>Cohort member</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Strengths and Difficulties Questionnaire (SDQ)</strong> Emotional symptoms, conduct, and peer problems, and hyperactivity, prosocial behaviour - parent-report about CM (3,5,7,11,14,17) - teacher report (7,11) - CM report (17)</td>
</tr>
<tr>
<td></td>
<td><strong>Kessler 6 (K6) (3,5,7,11,14,17)</strong> Depressive and anxiety symptoms</td>
</tr>
<tr>
<td></td>
<td><strong>YP Kessler 6 (K6) (17)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Mood &amp; Feelings Questionnaire (Shortened) (14)</strong></td>
</tr>
<tr>
<td></td>
<td>Diagnosis of depression or anxiety (17)</td>
</tr>
<tr>
<td></td>
<td><strong>Self-harm: 6 questions from the Edinburgh Study of Youth and Transitions incl. attempted suicide (17)</strong></td>
</tr>
</tbody>
</table>
# Mental health and well-being II

<table>
<thead>
<tr>
<th>Parent</th>
<th>Cohort member</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shortened Rosenberg Self-Esteem Scale (9m)</td>
<td>YP Shortened Rosenberg Self-Esteem Scale, (11,14,17)</td>
</tr>
<tr>
<td>Life satisfaction (9m,3, 5, 7,11)</td>
<td>YP Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) (17)</td>
</tr>
<tr>
<td>Happiness (9m, 3, 5, 7,11)</td>
<td>Life satisfaction (11)</td>
</tr>
<tr>
<td></td>
<td>Happiness (11,14)</td>
</tr>
</tbody>
</table>
Behaviours, attitudes and identity
Behaviours

- Temperament (14 items – Carey Infant Temperament Scale) (9m)
- Personality - Big Five personality traits (OCEAN) (M&P 7,14) (CM, 17)
- Self-regulation (3,5,7)
- YP Brief Self-Control Scale (17)
- Risky and antisocial behaviour (11,14,17)
  - Stealing, gambling, criminality, gangs, contact with police
- Risk and time preferences (17)
- Bullying, victimisation (11,14,17)
Attitudes and identity

- Activities and interests (7, 11, 14, 17)
- Aspirations (7, 11, 14, 17)
- Attitudes: Politics, Employment, Couples with children, Abortion, Race, Religion, The environment
- Sexual, gender, ethnic, religious identity, sexual attraction
Data enhancements and innovations
Access to different types of data

Access to data held by the UK Data Service varies depending on how the data is classified:

**Tier 1:** [End User Licence (EUL)](https://cls.ucl.ac.uk/data-access-training/data-access/) for access to data with a low level of sensitivity and disclosivity.
- Most of our data are available under this licence.
- Your application is authorised directly by the UK Data Service, and you can download the data directly from there.

**Tier 2a:** [Special Licence (SL)](https://cls.ucl.ac.uk/data-access-training/data-access/) for access to moderately sensitive or disclosive data. Access through the UK Data Service and application approved by CLS before you can download the data.

**Tier 2b:** [Secure Access Licence (SA)](https://cls.ucl.ac.uk/data-access-training/data-access/) for access to the most sensitive and/or potentially disclosive data. Access through the UK Data Service and attend a specialised training course.

[https://cls.ucl.ac.uk/data-access-training/data-access/](https://cls.ucl.ac.uk/data-access-training/data-access/)
## CLS record linkage programme – available data in MCS

<table>
<thead>
<tr>
<th>Domain</th>
<th>Country</th>
<th>Data set / information</th>
<th>Access</th>
</tr>
</thead>
</table>
| Education | England | • National Pupil Database (NPD) KS1,KS2  
• National Pupil Database (NPD) (KS1-KS4 including GCSE results, pupil level school census data and absence data) | via the UKDS (SA) via the UKDS (SA) |
| | Scotland | • NPD KS1 | via the UKDS (SA) |
| | Wales | • Welsh National Pupil Database (KS1)  
• Welsh National Pupil Database (up to KS4) in SAIL Databank | Via the UKDS (SA) via SAIL |
| Health | Scotland | • Inpatient, Outpatient, Birth Records, Immunisation, Prescribing Information, Child Health Review, | via the UKDS (SA) |
| | Wales | • Up to age 11- Health, Emergency Hospital inpatient (PEDW), Primary Care General Practice, Welsh Demographic Service.  
• Cohort members up to age 14 and parents in SAIL Databank | via SAIL via SAIL |

SA (UKDS Secure Access); SAIL (Secured Anonymised Information Linkage) Databank
## Geographical data enhancements

Available via the UKDS

<table>
<thead>
<tr>
<th>Domain</th>
<th>Country</th>
<th>MCS sweep</th>
<th>Data set / information</th>
<th>Access</th>
</tr>
</thead>
</table>
| Neighbourhood       | UK       | 1 to 6    | • Geographic identifiers (SA)  
  • For example, electoral wards, Output Areas, Lower and Middle Super Output Areas, Parliamentary Constituency | via the UKDS (SA)                     |
|                     |          |           |                                                                                       |                                      |
|                     | Great Britain | 4,5,6     | • Ordnance Survey Point of Interest (Pol)  
  • 3 levels (3rd level 600+ classes e.g. fast food outlets)  
  • Ordnance Survey Integrated Transport Network (ITN)  
  • Topological snapshot of the entire road network to create 400, 800 and 1600 metres network-based buffers. | via the UKDS (EUL)                     |
|                     |          |           |                                                                                       |                                      |
| Great Britain       | 3,4,5,6  | 1 to 6    | • Banded distances:  
  • Home and school  
  • Home moves between sweeps | via the UKDS (EUL)                     |
Harmonised data sets
(in MCS and across other British cohorts)
Available via the UKDS (EUL)

<table>
<thead>
<tr>
<th>Domain</th>
<th>MCS sweep</th>
<th>Data set / information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Socio-Economic</td>
<td>5 (age 11)</td>
<td>• Highest parental social class (RG 1990 version)</td>
</tr>
<tr>
<td>Body Mass Index (BMI)</td>
<td>1 to 5</td>
<td>• Weight</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Height</td>
</tr>
<tr>
<td>Mental health</td>
<td>5 (age 11)</td>
<td>• Four domains: emotional, peer problems, behavioural and attention / hyperactivity problems</td>
</tr>
</tbody>
</table>
MCS COVID-19 and serology surveys
Available via the UKDS (EUL)

<table>
<thead>
<tr>
<th>COVID-19 surveys response</th>
<th>MCS cohort members</th>
<th>MCS parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wave 1</td>
<td>2,645</td>
<td>2,831</td>
</tr>
<tr>
<td>Wave 2</td>
<td>3,274</td>
<td>5,707</td>
</tr>
<tr>
<td>Wave 3</td>
<td>4,474</td>
<td>5,251</td>
</tr>
</tbody>
</table>

Serology survey response

<table>
<thead>
<tr>
<th></th>
<th>MCS cohort members</th>
<th>MCS Parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invited</td>
<td>5,266</td>
<td>7,143</td>
</tr>
<tr>
<td>Consented</td>
<td>1,397</td>
<td>3,214</td>
</tr>
<tr>
<td>Blood sample returned</td>
<td>1,140</td>
<td>2,266</td>
</tr>
</tbody>
</table>

https://cls.ucl.ac.uk/covid-19-survey/

Serology Survey:
- Participants who took part in one of three COVID-19 Surveys were invited to provide a finger-prick blood sample
- Two antibody tests conducted - N-assay and S-assay
  - N-assay more likely to identify naturally occurring antibodies through exposure to virus
  - S-assay more likely to identify antibodies occurring following vaccination

https://cls.ucl.ac.uk/covid-19-survey/covid-19-antibody-testing/
MCS6 saliva samples were collected from cohort members and resident biological parents for DNA extraction:
- Approx. 200 twins or triplets
- Triads (cohort, mother, father) @47%, cohort and one parent @44%

Now available, genotyped data - single chip *(Infinium Global Screening Array)*

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort member</td>
<td>8,031</td>
</tr>
<tr>
<td>Mother</td>
<td>8,181</td>
</tr>
<tr>
<td>Father</td>
<td>4,782</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>20,994</strong></td>
</tr>
</tbody>
</table>

Working on: polygenic scores for multiple phenotypes

Genetic data access

- Data access system via CLS Data Access Committee
- Data sharing principles and approval criteria set out on the CLS Data Access Framework
- Complete CLS Data Access Request - which includes:
  i. Research project description (up to 500 words)
  ii. Brief methodology description (up to 500 words)
  iii. Ethico-legal issues including sensitive or controversial social topics (up to 500 words)

- And attaching MCS_Data_Dictionary.xlsx with requested variables
- Bespoke survey dataset identified by a specific project ID

https://cls.ucl.ac.uk/data-access-training/data-access/accessing-data-directly-from-cls/
Time use diary (MCS6 - age 14)

- Time use diary was paired with activity monitor: sample and day selection the same
- Full record of activities collected for 2 selected days; also collected information on where they were, who they were with, & how much they liked the activity
  - 4,851 CMs completed day 1 of the time-use record
  - 4,095 CMs completed day 2 of the time-use record
  - 8,946 TOTAL days of information

- Sleep & personal care
- School, homework & education,
- Paid or unpaid work...
- …Hobbies and other free time activities
- Any other activity
Sample design and weights
MCS Sample Design

- Stage 1: Selection of electoral wards (398)
  - Ethnic
  - Disadvantaged
  - Advantage

- Stage 2: Selection of families (with children alive and living in the UK at age 9 months (27201))

- Sample of children drawn from child benefit records held by DWP
Population stratified by country - England, Wales, Scotland, Northern Ireland.

Further stratification within country:

1) (England only) ‘Ethnic minority’ stratum: children living in wards which, in the 1991 Census, had an ethnic minority indicator of at least 30%, i.e. at least 30% of their total population fell into the two categories 'Black' (Black Caribbean, Black African and Black Other) or 'Asian’ (Indian, Pakistani and Bangladeshi)

2) ‘Disadvantaged' stratum: children living in wards, other than those in stratum (1), which fell into the upper quartile (poorest 25% of wards) of the ward-based Child Poverty Index (CPI) for England and Wales

3) ‘Advantaged' stratum: children living in wards, other than those falling into stratum (1), which were not in the top quartile of the CPI
Clustering

Primary sampling unit: electoral ward
Within wards, aim to recruit 100% of children born in the eligible period

<table>
<thead>
<tr>
<th>COUNTRY</th>
<th>Number of Sampled Wards</th>
<th>MCS1</th>
<th>MCS2</th>
<th>MCS3</th>
<th>MCS4</th>
<th>MCS5</th>
<th>MCS6</th>
<th>MCS7</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENGLAND</td>
<td>200</td>
<td>11532</td>
<td>10050</td>
<td>9717</td>
<td>8839</td>
<td>8618</td>
<td>7678</td>
<td>7124</td>
</tr>
<tr>
<td>WALES</td>
<td>73</td>
<td>2761</td>
<td>2261</td>
<td>2181</td>
<td>2018</td>
<td>1881</td>
<td>1669</td>
<td>1415</td>
</tr>
<tr>
<td>SCOTLAND</td>
<td>62</td>
<td>2336</td>
<td>1814</td>
<td>1814</td>
<td>1628</td>
<td>1480</td>
<td>1263</td>
<td>1105</td>
</tr>
<tr>
<td>N IRELAND</td>
<td>62</td>
<td>1923</td>
<td>1465</td>
<td>1534</td>
<td>1372</td>
<td>1308</td>
<td>1116</td>
<td>981</td>
</tr>
<tr>
<td>TOTAL UK</td>
<td>398</td>
<td>18552</td>
<td>15590</td>
<td>15246</td>
<td>13857</td>
<td>13287</td>
<td>11726</td>
<td>10625</td>
</tr>
</tbody>
</table>
Weighting

Analysis should take into account:

i. stratification and clustering sample design features – otherwise standard errors will likely be underestimated and resulting significance tests invalid
  ▪ **Stratification variable**
    9 different strata - stratum variable is *pttype2*
  ▪ **Clustering variable**
    398 wards were the primary sampling unit - ward variable is *sptn00*

ii. attrition and non-response
Attrition and non-response

- Attrition is the discontinued participation of some individuals in a longitudinal survey for reasons that are unknown and/or beyond the control of the researcher.

- Unit non-response (attrition) and item non-response (missing data).

- Types of unit non-response:
  - Non-contact
  - Refusal
  - Inability

- Non-response on the increase in all surveys.

- Non-response may not be permanent.

- Non-response/attrition can have some important implications.
Sampling and attrition weights

- Sampling weights: adjust the sample composition to take account of over-sampling in the first wave.
- Attrition weights: adjust the sample composition to take account of the loss of particular type of respondents.
- Adjustment means giving more importance (weight) to a particular group.
- Overall weights = Sampling Wgt $\times$ Attrition Wgt
Applying survey design in Stata: `svy`

Example using variables from MCS7:

```
svyset sptn00 [pweight=govwt2], strata(pttype2) fpc(nh2)
```

- `sptn00`: Electoral ward ID.
- `govwt2`: Overall weight in MCS7 for whole of UK analysis (you need to choose the correct one).
- `pttype2`: Stratum ID.
- `nh2`: Finite population coefficient.
Choosing the correct weight

- The choice of the weight depends on the analysis you are doing. The weight should correspond to the most recent sweep of data you are using and depending on whether it is a whole of UK or single country analysis.

- If you are using a Stata regression command which does not support ‘svy’ then you should use the weight option (specifying the correct weight e.g. `govwt2` for MCS7) and use the stratum (`pttype2`) as another control variable in your regression.
Resources and data access
The Millennium Cohort Study (MCS), known as ‘Child of the New Century’ to cohort members and their families, is following the lives of around 19,000 young people born across England, Scotland, Wales and Northern Ireland in 2000-02. The study began with an original sample of 18,918 cohort members.
Millennium Cohort Study: Age 17 Sweep (MCS7)
User guide

MCS7
Technical Report

Derived variables user guide
1st edition, August 2020
Available resources

- User guides
  - Overview of measures; Response and weights
- Questionnaires
  - Exact question wording; Questionnaire routing; Variable names
- Data documentation
  - Data notes; Coding frames; Variables lists, including derived variables
- Technical reports
  - Sample and questionnaire design, development
  - Fieldwork, response, ethics
  - Coding, editing
- Data dictionaries
- Previous journal publications [https://cls.ucl.ac.uk/publications-and-resources/](https://cls.ucl.ac.uk/publications-and-resources/)
SEARCH Closer Discovery (discovery.closer.ac.uk/)
Data freely available to researchers, government analysts and third sector workers: [https://ukdataservice.ac.uk/](https://ukdataservice.ac.uk/)
Available Resources: UK Data Service

Millennium Cohort Study: Seventh Survey, 2018

Documentation

- Title: Information about the MCS Longitudinal Data Dictionary
  File name: mcs_longitudinal_data_dictionary_20210203.xlsx
  Size: 1.35 MB

- Title: MCS Data Handling Guide with syntax in R, STATA and SPSS, August 2020
  File name: mcs_data_handling_guide_ed1_2020-08-12.pdf
  Size: 1.2 MB

  File name: mcs7_derived_variables_user_guide.pdf
  Size: 0.4 MB

- Title: MCS7 Parent Online (CAWI) Questionnaire
  File name: mcs7_parent_online-cawi-questionnaire.pdf
  Size: 1.11 MB

- Title: MCS7 Technical Report
  File name: mcs7_technical_report.pdf
  Size: 1.65 MB

  File name: mcs7_user_guide_age17_2ed_2020_12_08.pdf
  Size: 1.03 MB

- Title: MCS7 Young Person Interview
  File name: mcs7-young-person-interview.pdf
  Size: 1.18 MB

- Title: MCS7 Young Person Online (CAWI) Questionnaire
  File name: mcs7-young-person-online-cawi-questionnaire.pdf
  Size: 0.93 MB

- Title: MCS7 Young Person Self-completion Questionnaire
  File name: mcs7-young-person-self-completion-questionnaire.pdf
  Size: 0.95 MB

- Title: UK Data Archive Citation File for Study 8682
  File name: UKDA_Study_8682_information.htm
  Size: 0 MB

- Title: UK Data Archive Data Dictionaries
  File name: ukda_data_dictionaries.zip
  Size: 0.1 MB

- Title: UK Data Archive ReadMe File for Study 8682
  File name: README8682.htm
  Size: 0 MB

Millennium Cohort Study: Longitudinal Family File, 2001-2018

Documentation

- Title: MCS Data Handling Guide with Syntax in R, STATA and SPSS, August 2020
  File name: mcs_data_handling_guide_ed1_2020-08-12.pdf
  Size: 1.22 MB

- Title: MCS Longitudinal Family File, August 2020, Edition 2
  File name: mcs_longitudinal_family_file_2ed_2020_08_02.pdf
  Size: 0.17 MB

- Title: UK Data Archive Citation File for Study 8172
  File name: UKDA_Study_8172_information.htm
  Size: 0 MB

- Title: UK Data Archive Data Dictionaries
  File name: ukda_data_dictionaries.zip
  Size: 0 MB

- Title: UK Data Archive ReadMe File for Study 8172
  File name: README8172.htm
  Size: 0 MB
MCS ‘Survey and Biomeasures Data’

---

**Millennium Cohort Study**

<table>
<thead>
<tr>
<th>SN</th>
<th>Study description</th>
<th>Explore online</th>
<th>Select</th>
</tr>
</thead>
<tbody>
<tr>
<td>0602</td>
<td>Millennium Cohort Study, Seventh Survey, 2016</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8172</td>
<td>Millennium Cohort Study, Longitudinal Family File, 2001-2018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0156</td>
<td>Millennium Cohort Study, Sixth Survey, 2015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7464</td>
<td>Millennium Cohort Study, Fifth Survey, 2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7201</td>
<td>Millennium Cohort Study, First Survey, Health Visitor Survey, 2002-2002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7238</td>
<td>Millennium Cohort Study, Fourth Survey, Physical Activity Data, 2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5411</td>
<td>Millennium Cohort Study, Fourth Survey, 2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5795</td>
<td>Millennium Cohort Study, Third Survey, 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5559</td>
<td>Millennium Cohort Study, Survey of Mothers who Received Assisted Fertility Treatment, 2003</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5380</td>
<td>Millennium Cohort Study, Second Survey, 2002-2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4683</td>
<td>Millennium Cohort Study, First Survey, 2001-2003</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Centre for Longitudinal Studies**

---

**GN 33359**

Millennium Cohort Study - Survey and Biomeasures Data

---

**GN 33445**

Millennium Cohort Study - Linked Administrative Data
Accessing the data

- Register and set up an account at the UK Data Service
- Search for the data using the ‘find data’ tab
  - MCS
- Before downloading the data
  - Click on ‘Request Access’
  - Click on ‘Complete Actions’
  - Agree to standard ‘End User Licence’
  - Read and agree extra conditions
  - Choose data format and download zip file
    - SPSS
    - STATA
    - TAB (tab-delimited file)
Merging data within and across sweeps

Identify appropriate files
- Establish number of cases in target population

1. Check file structure: flat v hierarchical
- Transform if necessary

2. Identify merging variables:
   a. Unique ‘key’ cohort ID (member or family)
   b. Other ID’s depending on merge
       • Check the same variable name (case sensitive, changed across sweep etc)
       • Create identical variable name if necessary

3. Check merged correctly
Looking ahead
Looking ahead

- MCS8, Age 22 survey
  - Consultation complete
  - Piloting and development
  - Data deposit at UKDS ~ 2025

- Linked administrative data
  - Linked Hospital Episodes Statistics ~ Summer 2022
  - Health data beyond 14 (Wales) ~ Autumn 2022
  - University application (England; UCAS) ~ end 2022
  - Police National Computer (PNC) England ~ end of 2022
Proposed MCS8 content

**Education:**
- Academic and vocational learning and apprenticeships
- Educational experiences
- Qualifications

**Employment:**
- Occupation
- Income
- Labour market transitions
- Asset accumulation

**Physical health and health behaviours**
- Longstanding
- COVID-19
- Height/Weight/Body fat
- Substance use
- Sleep
- Diet and exercise

**Cognitive skills and processes:**
- Short-term memory
- Working memory and attention

**Mental health and well-being**
- Psychological distress
- Mental well-being
- Life satisfaction
- Loneliness
- Coping mechanisms
- Self-esteem
- Positive childhood experiences
- Body dysmorphia

**Family and relationships**
- Support from parents and family members
- Social roles and interactions
- Partnership formation and fertility

**Risky behaviours**

**Activities, attitudes, aspirations**

**Also:**
- Interviewing resident partners
- Collecting information on offspring
We’ve covered

- A brief overview of CLS and the MCS
- Examples of the types of research
- Data in the MCS by themes
- Enhancements and innovations
- Sample designs and weights
- Resources available and data access
- Q&A to the panel
Q&A to the panel

Please complete the feedback form
Questions to the panel

MCS team
- Lucy Haselden, Survey Manager
- Rachel Rosenberg, Research Data Manager
- Nicolas Libuy Rios, Research Fellow
- Vanessa Moulton, Senior Researcher
- and Morag Henderson, Associate Professor
IOE.CLS Data User Support

clsfeedback@ucl.ac.uk

Thanks!
Additional slides
# Key identifiers (ID’s) - Use these ID’s to link datasets

<table>
<thead>
<tr>
<th>Key identifier</th>
<th>Key cohort member/family identifier format</th>
</tr>
</thead>
</table>
| **MCSID**      | Family identifier  
|                | - every family has the same ID across sweeps  
|                | - everyone in the family has the same family ID (Cohort Members, parents, relatives and other people living in the household)  
|                | 7 characters:  
|                | M followed by 5 digits, and a single character  
|                | e.g. M10029A |
| **CNUM**       | Cohort members, 1, 2 (twins) or 3 (triplets) |
| **PNUM**       | Person number, for everyone else in the family apart from cohort members: parents, siblings, grandparents, etc |
| **ELIG**       | Eligibility for parent interview: a person selected for **Main** respondent and a person selected for **Partner** respondent |
| **RESP**       | Response outcome for the person eligible for parent interview (ELIG): **Main**, **Partner** respondent or No interview (in cases of refusal). |
# File structures: Flat v hierarchical

<table>
<thead>
<tr>
<th>Type of file structure</th>
<th>Format</th>
<th>Examples in the cohorts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat</td>
<td>1 record per case</td>
<td>MCS – family files</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>1 or more records per case</td>
<td>Household files</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Activities e.g. employment histories</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relationship histories</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Time use diaries</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MCS - Person within family</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BCSID</th>
<th>Sex</th>
<th>Country</th>
<th>Emp1</th>
<th>Emp2</th>
<th>Emp3</th>
</tr>
</thead>
<tbody>
<tr>
<td>M56768A</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>M46791B</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M87955C</td>
<td>2</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>M27614D</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>M34972E</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Examples in the cohorts**
- MCS: – family files
- Household files
- Activities e.g. employment histories
- Relationship histories
- Time use diaries
- MCS - Person within family

**BCSID**:
- M56768A
- M46791B
- M87955C
- M27614D
- M34972E

**Sex**:
- 1: Male
- 2: Female

**Country**:
- 1: Country A
- 2: Country B

**Emp**:
- 1: One record
- 3: Three records
## Linked health administrative data in the cohorts

<table>
<thead>
<tr>
<th>Country</th>
<th>Study</th>
<th>Data set / information</th>
<th>Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>England</td>
<td>NCDS, BCS, Next Steps</td>
<td><strong>Hospital Episodes Statistics (HES)</strong>&lt;br&gt;• Admitted Patient Care (APC)&lt;br&gt;• Critical Care (CC) – linked to APC&lt;br&gt;• Accident &amp; Emergency (A&amp;E)&lt;br&gt;• Outpatient Care (OP)</td>
<td>Available at UKDS <a href="https://cls.ucl.ac.uk">Link</a> via Secure Server</td>
</tr>
<tr>
<td>Scotland</td>
<td>BCS, NCDS, MCS</td>
<td><strong>Scottish Medical Records (SMR)</strong>&lt;br&gt;• Inpatient&lt;br&gt;• Outpatient&lt;br&gt;• Birth and Neonatal Records&lt;br&gt;• Prescribing Information&lt;br&gt;• Maternity&lt;br&gt;• Immunisations (SRS), Child Health Review</td>
<td>Available at UKDS via Secure Server</td>
</tr>
<tr>
<td>Scotland</td>
<td>BCS, NCDS only&lt;br&gt;MCS only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wales</td>
<td>MCS</td>
<td>• Health data assets from SAIL Databank (e.g. emergency department, outpatient, COVID data) up to age 14 and for cohort members’ parents&lt;br&gt;• Hospitalisations &amp; no. of diagnoses from ICD 10 chapter codes up to age 11</td>
<td>Available at Secure Anonymised Information Linkage (SAIL) Databank Available at UKDS via Secure Server</td>
</tr>
</tbody>
</table>

### Coming soon:
- HES data from NHS Digital linked to MCS
- HES refresh: new A&E, APC, CC, OP including COVID-19 data
- CLS will refresh the Welsh health dataset linked to MCS (2001-2012) currently available at the UKDS

More information available [cls.ucl.ac.uk](https://cls.ucl.ac.uk), including HES and SMR user guides and introductory training webinar video