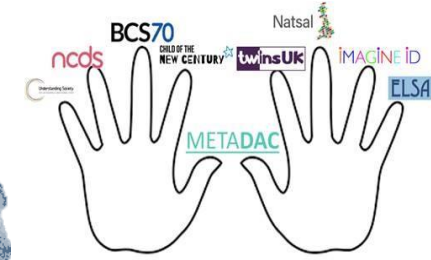


# METADAC



# Data Access Made Easy

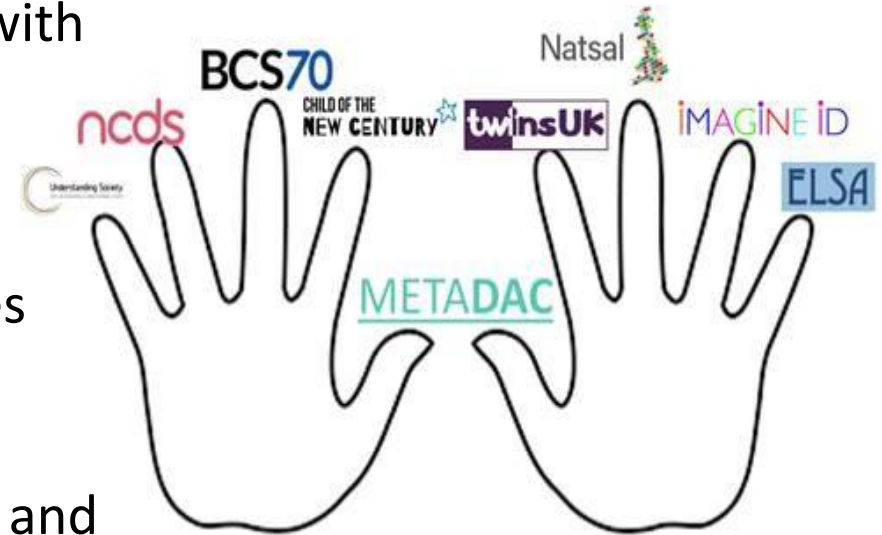
How to Access Genotypic and  
Phenotypic Data from UK  
Longitudinal Studies

Madeleine Murtagh, Neil Walker,  
Meena Kumari, Alissa Goodman,  
Samantha Aceto

# METADAC

# METADAC: Accessing biosocial and genotypic data

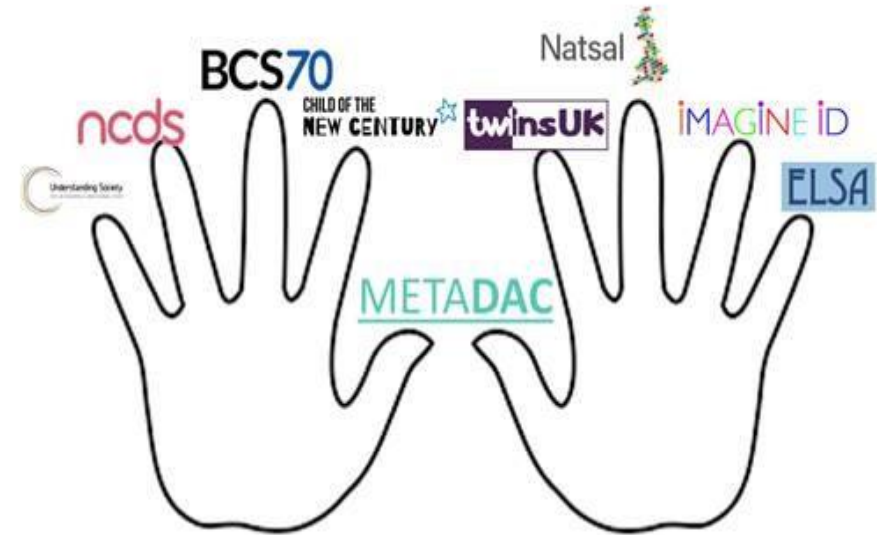
- METADAC manages access to data and samples from eight UK longitudinal population studies
  - Decades of biosocial and genomic data and samples with consent for (re)use
- METADAC Access Committee
  - Assesses applications for sensitive data, combined genotype and phenotype data, and samples
  - Genotype-only data available from EGA
  - Phenotype-only data available from UKDA
  - Committee comprises expertise from across domains and disciplines, including that of study participants



# METADAC: Aims and Principles

---

- METADAC Aims
  - Protecting participants
  - Protecting the study
  - Protecting researchers
- METADAC Access Principles
  - Independence and transparency
  - Interdisciplinary expertise including that of study participants
  - Collective governance & participant-centric decision making





**Understanding Society**

THE UK HOUSEHOLD LONGITUDINAL STUDY

# Genomic data in *Understanding Society*: the UK Household Longitudinal study (UKHLS)

Meena Kumari  
Professor of Biological and Social Epidemiology

**Festival of Genomics**

30<sup>th</sup> January 2020

An initiative by the Economic and Social Research Council, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by the National Centre for Social Research.

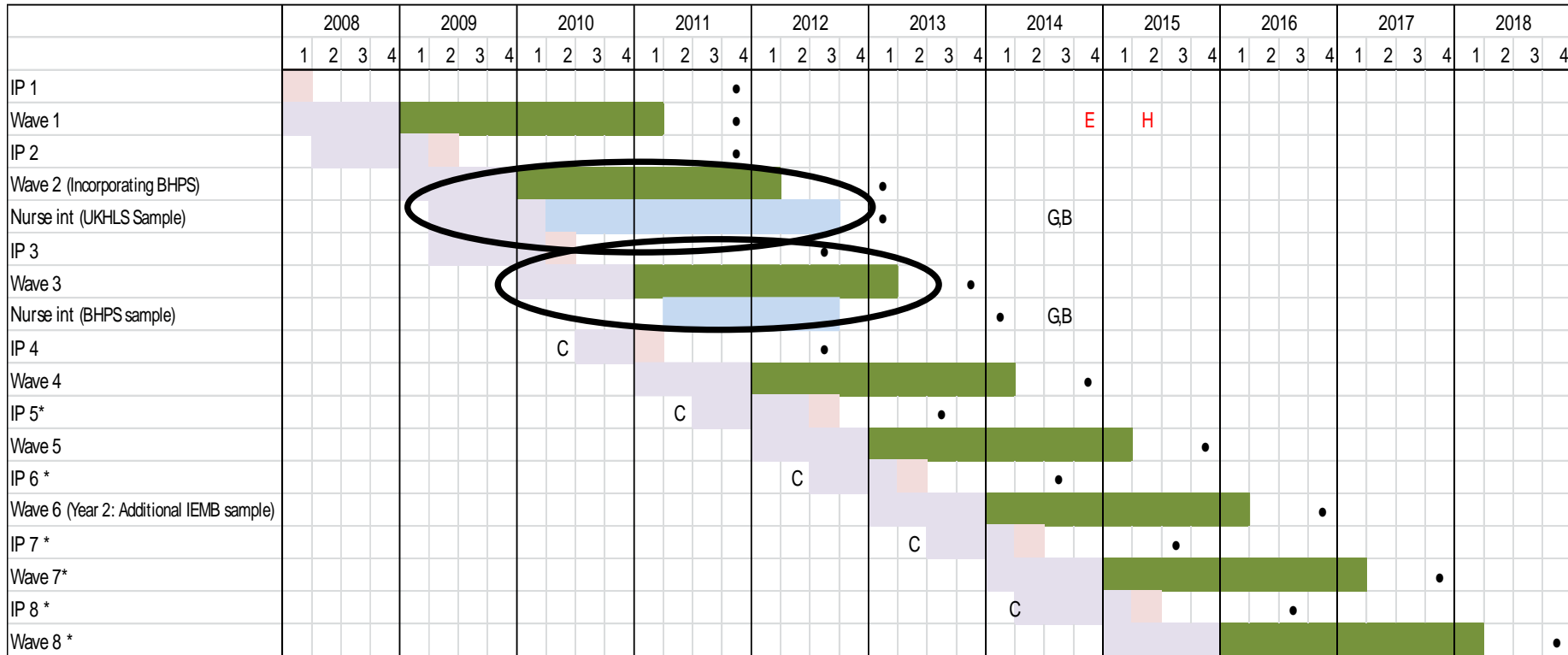
# Understanding Society

- Large sample (40,000 households, 100,000 people) gives an opportunity to explore rarer events, behaviours and conditions
- Allows examination of small sub-populations, such as teenage parents or disabled people
- All age panel study but the size allows pseudo cohorts to be constructed and hence examination of age, cohort and period effects
- Includes geographical identifiers (GOR) allowing examination of the regional variations

# Initial sample consists of:

- **General Population Sample:** 30,000 UK households in 2009 (58% household response; 82% adult response = 41,047 interviews)
- **Ethnic Minority Boost:** 1,000 adult individuals across five main ethnic groups (Indian, Pakistani, Bangladeshi, Caribbean, African) (52% household response; 72% adult response = 6,685 interviews)
- **British Household Panel Survey (1991 - ):** approximately 8,000 households, 13,454 adults consenting to continue
- **An Innovation Panel:** 1500 households, 2568 adults for methodological research
- **TOTAL: 39,805 households: 101,087 individuals: 63,755 adults, 5,788 youths**
- **NEW: Migrant Boost** sample in planning = 4500 households

# Understanding Society: the UK Household Longitudinal Study: Survey and data release timeplan waves 1-8



**KEY**

Preparation:

Fieldwork:

- IP
- Main
- Nurse

\* Multi-mode data collection

IEMB Immigrant and Ethnic Minority Boost

C Competition for experiments

Planned data release for:

- Questionnaire
- B Blood analytes
- G Genotype data
- E Linked education data
- H Linked health data



# Nurse assessments

Measure	Applications
Height and weight Waist circumference Percent body-fat (bioelectrical impedance)	BMI and assessment of excess body fat: obesity and risk factor for range of major chronic conditions and social outcomes
Respiratory function (Spirometry) (FVC, FEV <sub>1</sub> , PF, FEV <sub>1</sub> /FVC)	To detect both obstructive and restrictive respiratory diseases including COPD.
Diastolic and systolic blood pressure, resting pulse rate	Risk factor for stroke and heart conditions Risk cardio-vascular disease
Grip strength	Indicator muscle strength. Functional limitations and disability in older ages.
Blood samples (non-fasting), 19.8ml)	For the extraction of analytes and DNA
Short questionnaire on health on day of measurement, medications etc	Factors that may need to be considered in analysing physical measures and bloods

# Understanding Society

## Genetic Data

- **9,944** Individuals who also have biomarker data.
- **Population:** England, Scotland and Wales.
- Genotyped at the **Wellcome Trust Sanger Institute** (Genome Research)
- **Illumina** Human Core Exome BeadChip
- **>500,000 SNPs**
- **>8,000,000 Imputed SNPs**



56%



44%

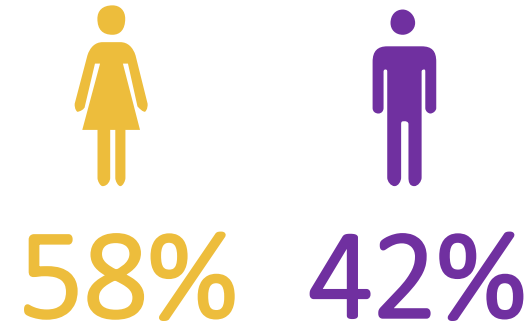
**wellcome**trust

Eleftheria Zeggini  
Karoline Kuckenbäcker  
Bram Prins

# Understanding Society

## Epigenetic Data

- Selected from those who had genetic Data
- Blood processed within 1-2 days
- British Household Panel Survey (BHPS)
- At least 10 years of BHPS data (2/3rds have 20 years)
- **Data Normalised and cleaned by Tyler Gorrie-Stone, Professor Leo Schalkwyk**



- **1175 samples**
- **Infinium MethylationEPIC BeadChip**
- **>850,000 methylation sites across the genome**
- **857,071 sites after QC**



**Professor Jonathon Mill**  
**Dr Eilis Hannon**  
**Dr Joe Burrage**

# Understanding Society

phenotypic data

- Health
  - Long term conditions
  - Health functioning/mental and physical
  - Personality/cognition
  - Health behaviours
  - Biomarker data
- Social
  - Demographic variables
  - Measures of social position (education, employment, income, occupation)
  - Social integration and participation

# Genomic data in *the 1958 Birth Cohort* (National Child Development Study)

## Centre for Longitudinal Studies (CLS)

Prof Alissa Goodman  
Director, Centre for Longitudinal Studies

CENTRE FOR  
LONGITUDINAL  
STUDIES

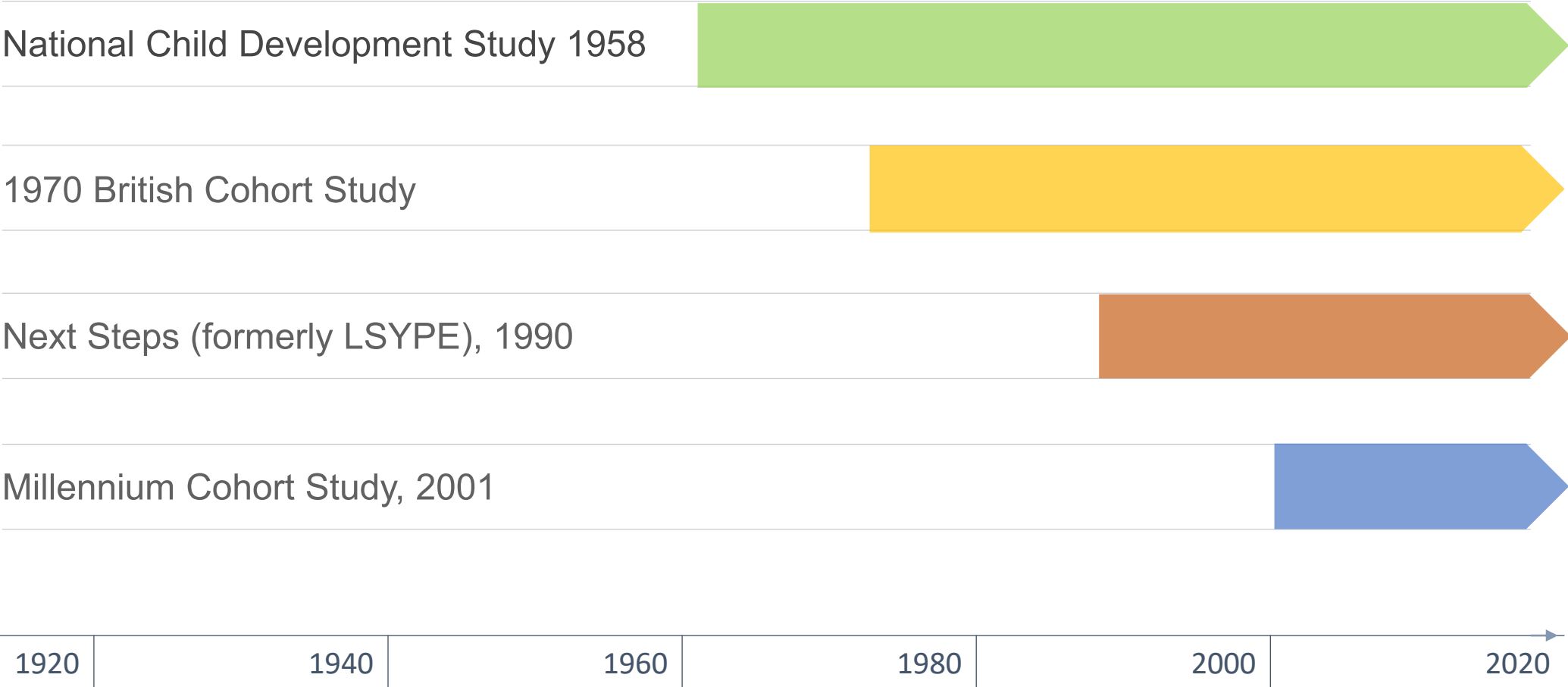
The logo for the Centre for Longitudinal Studies, consisting of three horizontal red lines of varying lengths below the text.

Festival of Genomics






30<sup>th</sup> January 2020






### Cohort Studies Timeline



# NCDS 58 A study of everyone born in one week in 1958

	1958	1965	1969	1974	1981	1991	2000	2003	2004	2008	2013
	<b>Birth</b>	<b>7</b>	<b>11</b>	<b>16</b>	<b>23</b>	<b>33</b>	<b>42</b>	<b>44/5</b>	<b>46</b>	<b>50</b>	<b>55</b>
 <b>main respondent</b>	<b>mother</b>	<b>parents</b>	<b>parents</b>	<b>cohort member / parents</b>	<b>subject</b>	<b>subject</b>	<b>subject</b>	<b>subject</b>	<b>subject</b>	<b>subject</b>	<b>subject</b>
 <b>secondary respondent</b>	<b>medical</b>	<b>medical / school</b>	<b>medical / school</b>	<b>medical / school</b>		<b>partner mother children</b>					
 <b>survey instruments</b>		<b>cognitive assessments</b>	<b>cognitive assessments</b>	<b>cognitive assessments</b>						<b>cognitive assessments</b>	
 <b>linked data</b>				<b>area of residence (census)</b>	<b>area of residence (census)</b>						
 <b>response rate</b>	<b>17,415</b>	<b>15,425</b>	<b>15,337</b>	<b>14,654</b>	<b>12,537</b>	<b>11,469</b>	<b>11,419</b>	<b>9,377</b>	<b>9,534</b>	<b>9,790</b>	<b>9,137</b>

# Whole study phenotypic information

 Birth	 School years	 Adult
<p>Household composition</p> <p>Parental social class</p> <p>Obstetric history</p> <p>Smoking in pregnancy</p> <p>Pregnancy (problems, antenatal care)</p> <p>Labour (length, pain relief, problems)</p> <p>Birthweight, length</p>	<p>Household composition</p> <p>Parental social class</p> <p>Parental employment</p> <p>Financial circumstances</p> <p>Housing</p> <p>Health</p> <p>Cognitive tests</p> <p>Emotions and behaviour</p> <p>School</p> <p>Views and expectations</p> <p>Attainment</p>	<p>Household composition</p> <p>Employment</p> <p>Social class</p> <p>Income</p> <p>Housing</p> <p>Health (including biomarkers)</p> <p>Well-being and mental health</p> <p>Health-related behaviour</p> <p>Training and qualifications</p> <p>Basic skills</p> <p>Cognitive tests</p> <p>Views and expectations</p>



# NCDS58 biomedical data



- Approximately 9,000 study members took part at age 44/5 (2002/3)
- Biosamples: blood, saliva
- Blood pressure, pulse
- Standing and sitting height
- Weight, waist and hip circumferences
- Respiratory symptoms, ventilatory function (FEV1 and FVC)
- Visual acuity (near and distant), refractive error
- Hearing thresholds
- Depression and anxiety disorder (CIS-R)
- Chronic widespread pain
- Use of medications
- Alcohol use (AUDIT)
- Food frequency questionnaire, exercise habits
  
- Extensively assayed
- Extensively genotyped

# CLS cohorts genetic data

- 1958 cohort, c. 8,000 study members, immortalised cell lines, and genotyped
- Epigenetic data for c. 500 at age 44/5, c. 1000 more will be available soon
- Epigenetic data for c. 200 at age 61/2 will be available soon
  
- 1970 cohort, blood samples taken, c. 6,000 study members at age 46
- Epigenetic data for c. 240 samples so far, c. 200 more will be available soon
  
- MCS cohort – genetic trios. Illumina Global Screening Array
- 4,500 – study members and both natural parents, in total 24,000 samples

# METADAC: Access Criteria

---

## METADAC application assessment criteria: Accountability, respect and communication

- Project falls within study remit
- Aligns with ethical conditions
- Unlikely to upset or alienate participants
- Unlikely to bring disrepute to the study
- Non-disclosive
- Applicants must be bone fide researchers
- PI must be a senior researcher
- Use of depletable samples requires external review
  
- Data minimisation
- Does the application run the risk of upsetting or alienating study members or of reducing their willingness to continue as participants?

# The fast track to successful METADAC applications

## 1) Potential for social controversy

- Reputational damage/ alienation of participants that has not been addressed by the applicants.

## 2) Too many, or poorly justified variables list

- The list has not been refined to include only those variables required to answer the research question – risk of disclosure/breach of privacy
- The list includes variables that are unexplained in relationship to the work - data minimisation is not evidenced

## 3) Plain language summary

- The summary lacks content, does not clearly describe the project, uses technical words, or is otherwise too complex for the 12-year-old reader to understand.

# The fast track to successful METADAC applications

Learn more about using data and samples from UK longitudinal studies

Prepare better data and samples access applications

Interactive Workshop

April 1<sup>st</sup> @ Wellcome Trust, London

Register your interest: [metadac@Newcastle.ac.uk](mailto:metadac@Newcastle.ac.uk)

# Thank you!

## Questions?

Email: [metadac@newcastle.ac.uk](mailto:metadac@newcastle.ac.uk)

Twitter: @metadac

www: [metadac@ac.uk](http://metadac@ac.uk)

