Understanding Society

THE UK HOUSEHOLD LONGITUDINAL STUDY

# Recent developments in *Understanding Society*

*Meena Kumari, Professor of Biological and Social Epidemiology*
*mkumari@essex.ac.uk*
*Health Study Users Conference*
*29th July 2023*

University of Essex

ISER

UKRI Economic and Social Research Council

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 NatCen Social Research and Kantar Public

# General overview

- Overview of *Understanding Society*
- Recent data releases
  - PEACH
  - -omics
    - Epigenetic biomarkers of age
    - proteomics
- Next Steps

# What is Understanding Society (UK Household Longitudinal Study)

Understanding Society surveys a sample of individuals representing the UK population, who are interviewed within a household context:

- starts with randomly selected sample of households
- collects information about all residents of these households
- These residents and their off-springs form the core sample
- These core sample members are followed over their life course at 1 year intervals and within UK

Basic design is similar to household panel surveys in other countries e.g., PSID (US), SOEP (Germany), HILDA (Australia), SoFIE (New Zealand)

# Sample consists of:

- **General Population Sample (GPS):** 26,000 UK households from 2009 population (from wave 1)
- **Ethnic Minority Boost (EMB):** 4,000 households with at least one individual with an ethnic minority background across five main ethnic groups (Indian, Pakistani, Bangladeshi, Caribbean, African - from wave 1)
- **British Household Panel Survey (1991 - ):** approximately 8,000 households were added to continue in UKHLS (from wave 2)
- **Innovation Panel:** 1500 households for pioneering data collection and methodological research
- **Total:** 39,805 households (from waves 1 and 2)
- **A new Immigrant and Ethnic Minority Boost (IEMB)** sample added in 2015/6 = 2500 households with at least one person born outside the UK or from an ethnic minority background (from wave 6)

# Who provides the data?

- Most of the data is collected during interviews with adult household members (aged 16 years or above)
- Data is also collected from 10-15 year olds in separate short self-completion interviews
- Data about children 0-9 year olds is collected from their parents and guardians

- Prospective survey
- … with retrospective elements – changes since last interview, initial histories
- Indefinite life (not fixed life)

# Topics areas

- Biomarker, Genetic and Epigenetics
- COVID-19
- Education
- Employment
- Ethnicity and Immigration
- Family and Household
- Money and Finances
- Politics and Social attitudes
- Health and Wellbeing
- Young people
- Transport and Environment

# Health & wellbeing

Health data from two sources:

**1. Questionnaire i.e. individual reports it, e.g.:**

- Current prescriptions
- Currently has chronic condition
- GHQ measures mental health

**2. Objective data measured by a nurse** (in Wave 2 & 3):

- Height/weight
- Lung function, grip strength
- Blood analytes – for cholesterol, raised glucose, kidney & liver function, anaemia
- Genetic data

# Adult and children mental well-being

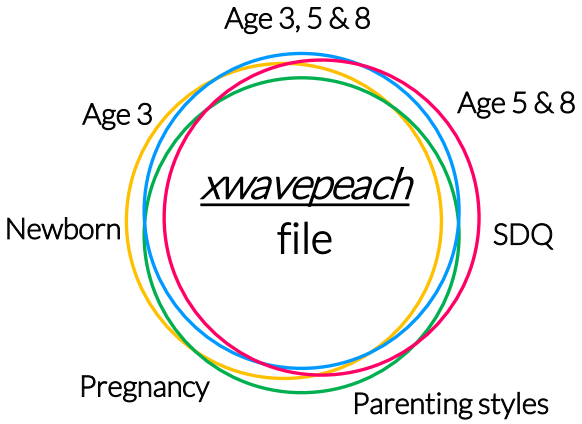
- Adults: Annual
  - General Health Questionnaire – 12 (continuous, categorical or caseness scales; individual items
  - Satisfaction
    - Domains: Overall life, job, leisure time, income, health
- Waves 1, 4, 7, 10
  - Edinburgh-Warwick Positive well being (short form)
- Children: Annual
  - Happiness with different aspects of life (school work, appearance, family, friends, school, life as a whole - can be combined to create an overall happiness score)
- Biennially (start at wave 1)
  - Strengths and Difficulties Questionnaire (SDQ) - can create a total difficulties score or 5 subscales

# Pregnancy and Early Childhood file (PEACH)

- Our primary goal is to facilitate researchers interested in child development topics to engage with Understanding Society data seamlessly.
- The file brings together most of the already available information on pregnancy and early childhood through mainstage, into a single location, enhancing its quality and value.
- It has been constructed by leveraging data from all children reported in the child file based on their eligibility for the child development questions.
- The information provided is at the child level, using the child's identifier (pidp) to ensure each row is uniquely identifiable.
- All variables within each category are time-specific and remain constant across all waves. *For instance, information such as the child's birth weight, the frequency of fussing or crying during infancy, and the frequency of reading to the child at ages 3, 5, and 8 are age-specific and remain constant over time.*
- A total of 303 variables are included, created from 164 questions.

# Pregnancy and Early Childhood file

- Its structure is designed to facilitate easy tracking of various aspects of a child's development, including whether they were born prematurely or past their due date, their ability to speak in full sentences at age 3, and their health condition at ages 3, 5, and 8, amongst other variables.
- In addition, the file includes the cross-wave identifier of the person who provided the information for the child and the wave when it was collected.
- We have classified the information available about children into 7 categories, based on content and when these were asked.




- Plans are to release an updated version of the file every year, after a new wave is released, as PEACH is not part of mainstage releases.

# -OMIC Data
## DNA methylation

- Selected from those who had genetic Data
- Measures in two batches:
  - Batch 1: BHPS Blood samples processed within 1-2 days
  - Batch 2: remaining BHPS samples and random sample from GPS
- **Data Normalised and cleaned by Tyler Gorrie-Stone, Professor Leo Schalkwyk**


58% 42%

- **3105 samples**
- **Infinium MethylationEPIC BeadChip**
- **>850,000 methylation sites across the genome**
- **857,071 sites after QC**


UNIVERSITY OF EXETER

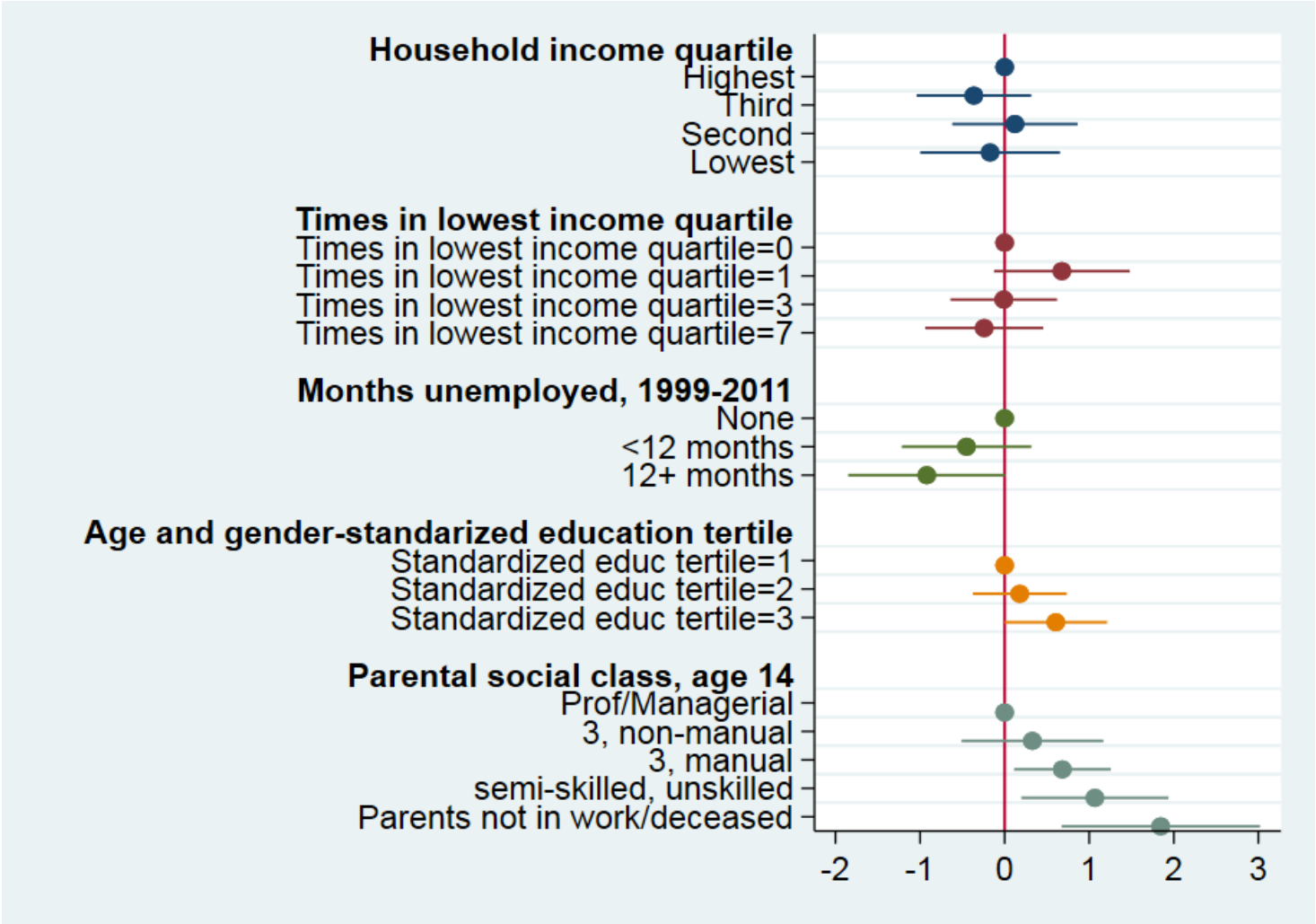
**Professor Jonathon Mill**
**Dr Eilis Hannon**
**Dr Joe Burrage**

# -OMIC data: biomarkers of age

- DNA Methylation ‘clocks’:
- built with **supervised machine learning methods**, such as penalized regression (e.g., lasso or elastic net) trained against chronological age or other phenotypes of age to identify an informative and sparse predictive set of methylation sites

- Understanding Society has deposited 5 such biomarkers of age
  - ‘Horvath’, ‘Hannum’ – first generation measure trained against age
  - ‘Lin’, ‘phenoage’ – second generation measure trained against biomarkers of age/prediction of mortality
  - ‘DunedinPoAm’ – third generation measure trained against longitudinal measures of age and considered a measure of speed of ageing

# Education and father's social class and 'accelerated age'




Analyses adjusted for age, age2, batch, cell type, smoking status, BMI



From: Hughes et al., AJE 2018

# -OMIC Data

## Proteomics

Cardiometabolic

Complementing and expanding on our previous portfolio of disease area-focused biomarker panels, our exploratory panels offers scientists the chance to cast a wider net in the quest to identify new biomarkers and relevant protein signatures, offering many new assays focused on important biological processes with wide-ranging clinical relevance. assays in this panel include proteins involved in key biological processes such as *cellular metabolic process, cell adhesion, immune response* and *complement activation.*

https://www.olink.com

Neurology

The panel offers a mix of established markers related to neurobiological processes and neurological diseases (e.g. neural development, axon guidance, synaptic function, or specific conditions such as Alzheimer's disease), as well as some more exploratory proteins with broader roles in processes such as cellular regulation, immunology, development and metabolism. This balanced selection provides an ideal basis for protein biomarker discovery in the neurology area.

- 184 biomolecules that are NOT established measures of health

- Includes proteins <u>related to</u>
  - cell metabolism
  - cell adhesion
  - immune response
  - complement (an immune pathway)
  - neuron formation and function
  - neurodegenerative diseases

# -OMIC Data
# Proteomics

- 184 biomolecules that are NOT established measures of health

- Includes proteins related to
  - cell metabolism
  - cell adhesion
  - immune response
  - complement (an immune pathway)
  - neuron formation and function
  - neurodegenerative diseases

# Exemplar analyses

## Social differences in proteins: association with educational attainment

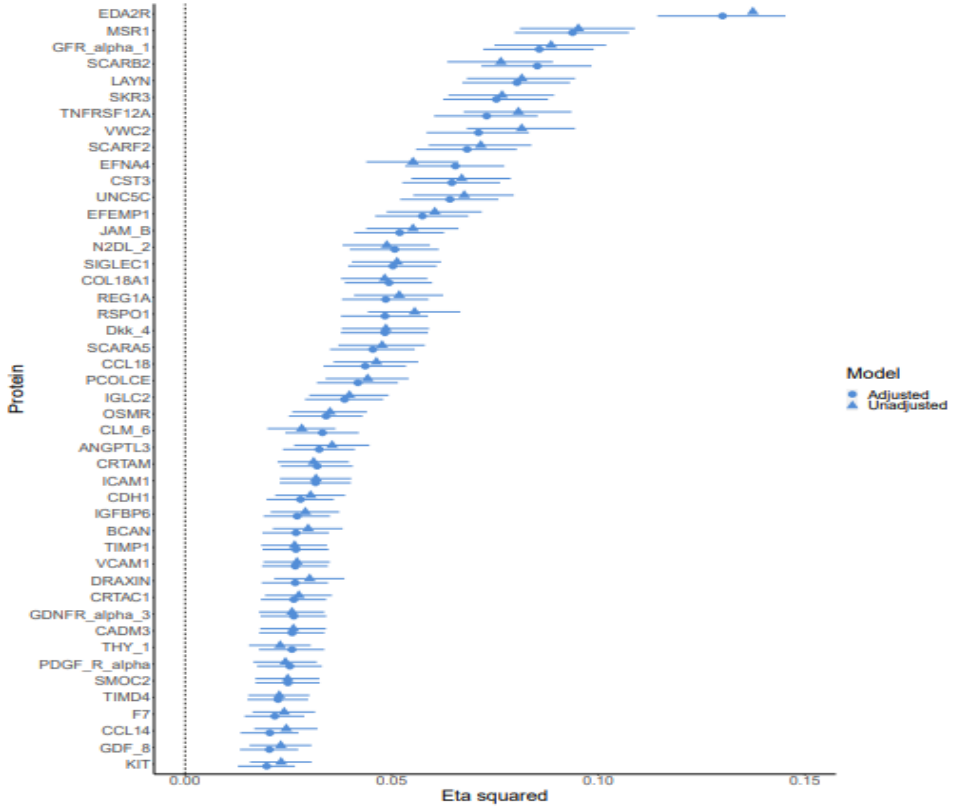



Figure 4: Associations between proteins and educational attainment

# Websites

- www.data-archive.ac.uk
- Understandingsociety.ac.uk

- genetics@understandingsociety.ac.uk
- mkumari@essex.ac.uk