DOI: 10.1111/1475-5890.12354

Fiscal Studies
The Journal of Applied Public Economics
IFS Institute for Fiscal Studies

SYMPOSIUM PAPER

Understanding Society: health, biomarker and genetic data

Michaela Benzeval | **Edith Aguirre** | **Meena Kumari**

Institute for Social and Economic Research, University of Essex

Correspondence
Michaela Benzeval, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK.
Email: mbenzeval@essex.ac.uk

Submitted: November 2023

Abstract
Understanding Society, the UK Household Longitudinal Study includes a wide range of health measures, and in particular biomarker and genetic data. This makes it a unique resource for research on the economics of health. We review the main features of the biomarker data, how they are collected, and evidence on data quality. We also discuss examples of how these data have been used in economic research to date.

KEYWORDS
health, household, longitudinal, panel survey, UKHLS

JEL CLASSIFICATION
C8, I1, I14

1 | INTRODUCTION

To formulate effective policies that enhance individuals' health and their economic and social opportunities, it is essential to understand the complex relationships between people's socioeconomic circumstances and their health across their life span. Unfortunately, too often the 'biomedical literature has generally treated socioeconomic position as a unitary construct. Likewise, the social science literature has tended to treat health as a unitary construct'.[1] However, for effective research, we need datasets that comprehensively capture different domains of individuals' social and economic lives and different dimensions of their health. One way many social science surveys have achieved this is by integrating health and biomarkers into well-established, high-quality longitudinal datasets, which already capture of a breadth of information on social and economic circumstances, such as Understanding Society.[2]

Understanding Society, the UK Household Longitudinal Study (UKHLS) is a large, multi-domain, household panel study that began in 2009. It builds on and incorporates the British Household

[1] Herd, Goesling and House (2007, p. 7).

[2] Benzeval, Kumari and Jones, 2016.

Panel Survey (BHPS), which began in 1991. For a fuller description of the study design and content, see Benzeval, Crossley and Aguirre (2023a). From the inception of the BHPS to now, a wide range of consistent measures of different health domains have been included in the study, alongside rich social and economic data. These include: measures of long-term health conditions; physical and mental health functioning; disabilities; symptoms; access to health care; information on conceptions, pregnancy, birth; early life health and development; and health behaviours. In wave 2/3 of Understanding Society (2010–12), an additional interview by a nurse was conducted to enhance the study with objectives/biological measures of health. The collection of these biological health measures will be repeated at wave 16 (2024–26) and wave 22 (2030–32).

The aim of this paper is to outline the availability of health and, in more detail, biological data in Understanding Society and to explore their potential applications in economic research. In Section 2, we introduce the rationale for including multiple health measures in a multi-domain social science study of probability-based samples of households over time, such as Understanding Society. In Section 3, we outline the different health and biological data collections conducted and/or planned. Section 4 illustrates the kinds of research opportunities possible in economics with biomarker and genetic data. Section 5 describes how to access and cite the data. Finally, we conclude in Section 6 by identifying how research using such data might contribute to current policy challenges.

2 | WHY IS IT IMPORTANT TO COLLECT HEALTH INFORMATION IN LONGITUDINAL MULTI-DOMAIN SURVEYS?

Understanding health, its causes and consequences for other domains of life, is a significant focus of economic research. Data for such research can come from cross-sectional or longitudinal surveys, social or medical studies, and administrative health records. Each source has its advantages for different kinds of research questions. For example, repeated cross-sectional surveys allow researchers to study time trends in the prevalence of health and disease in different population groups, which is important for planning health and social care resources but does not enable research into the development or consequences of those conditions. Many cohort studies provide rich information on health and social conditions at key stages in people's lives so that development can be tracked, but they are fixed in the specific context – period, occupation, life stage, etc. – in which they began, and hence cannot provide evidence on the full population over time. Medical studies of specific conditions, often based on volunteers, can have very high compliance rates among those who are affected by a condition to enable the detailed study of the progress of a disease, its biological underpinnings, and the impact of potential treatments, but do not provide insights into the extent to which such health conditions are experienced in the broad population. Administrative health records can provide very detailed information on medical conditions and their treatments to enable effective analysis of the costs of different kinds of health care, but they only have data on those people who seek help and are treated for their health conditions. Moreover, administrative data can say little about the social and economic causes of those conditions, or the impact or cost of them on people's wider lives or families.

Understanding Society, as a multi-domain, large probability-based sample of households, with annual data collected over a long time-span,[3] overcomes many of these limitations.

- As a multi-domain study, Understanding Society enables the association between a wide range of experiences and topics to be researched. For example, Arulsamy and Delaney (2022) investigated the impact of the introduction of auto-enrolment for pensions on health-related inequalities in pension uptake. Controlling for a wide range of covariates, they find that this policy reduced inequalities caused by men with mental health problems not enrolling in workplace pension schemes before the policy change.

[3] Benzeval et al., 2023a.

- As a large probability-based sample, research in Understanding Society can provide evidence on the health experiences of the whole population and subgroups within it. Based on knowing the probability of the initial selection into the sample from the general population, inverse probability weights are produced to adjust for designed oversampling, initial non-response and subsequent attrition[4] and have been shown to minimise bias based on observed characteristics.[5] For example, Banks, Karjalainen and Waters (2023) used Understanding Society data to investigate inequalities in disabilities over the life cycle and by education for the IFS Deaton Review of Inequalities. They found that rates of disability increased with age across life cycles and were highest for those without qualifications.
- As a study of households, Understanding Society enables researchers to study the inter-relations of health conditions among family members: for example, a child may inherit health conditions from their parents; or a child's health condition may affect their parents' mental health or employability. One partner in a couple losing their job may affect the shared financial resources of the family, and hence the health of the other family members. Davillas and Pudney (2017) investigated concordance of health among couples and found strong correlations for biological risk factors for heart disease as well as self-assessed general health and functional difficulties. These associations were not affected by marriage duration, suggesting shared environments/lifestyles and partner selection accounted for observed similarities.
- Having consistent longitudinal annual data over a long time-span enables researchers to investigate how aspects of health and development during childhood and even pre-conception may influence health and other outcomes, such as education and occupational opportunities in later life. For example, using data from the BHPS and Understanding Society, Martínez-Jiménez (2023) finds an association between parents' non-employment during childhood and health in adulthood. However, the health domains that are affected depend on when in childhood the situation occurred, which parent was affected and the family's social and economic circumstances. Having consistent longitudinal annual data also allows for the investigation of unexpected exogenous shocks. For example, using Understanding Society data from before and during the COVID-19 pandemic, Madia, Moscone and Nicodemo (2023) demonstrated that the health of carers who began caring during the pandemic was poorer than those who had been carers before it started.

In this context, including biomarkers – indicators of normal biological processes, pathogenic processes, or biological responses to an exposure or intervention[6] – in social and economic surveys brings further advantages for economic research. Biomarkers include physical measures such as height and weight, blood pressure and lung function. They also cover measures derived from samples collected from the body (e.g. blood, urine or saliva) to provide indicators of disease or risk of disease such as cholesterol, liver and kidney function. As such, they are objective measures of health (although still subject to measurement error) which often provide earlier and more precise measurements of health and functioning. For example, the biomarker HbA1c indicates the presence of diabetes. Comparing participants with high levels of this biomarker in Understanding Society with those who report that they have been diagnosed with diabetes identifies a small proportion of the population who have the condition but have not been diagnosed.[7] The characteristics of such participants can then be investigated to identify subgroups who could be targeted for screening. Another benefit of biomarker data is that they enable the underlying biological pathways that link economic circumstances and health to be explored.[8] This might involve studying genetic data, which are inherited and fixed;

[4] Lynn, Cabrera-Álvarez and Clarke, 2023.

[5] Cabrera-Álvarez, James and Lynn, 2023.

[6] FDA–NIH Biomarker Working Group, 2016.

[7] Kumari and Benzeval, 2021.

[8] Kumari and Benzeval, 2021.

epigenetic data, which are not inherited and are amenable to the environment and affect if and how genes are expressed; and proteomics data, which measure proteins produced by genes to cause differ bodily processes to take place. Investigating how different economic situations interact with or affect different stages of this biological pathway may help us better understand the influence of society on health.

In Understanding Society, there has been one main wave of biomarker data collection (wave 2/3, 2010–12) to date, which affects the ways in which these data can be used in research. Biomarkers can be used in economic research as outcomes for short- or long-term economic precursors, for example, in studies of the association between childhood circumstances and later-life risk of heart disease. They can also be used as explanatory factors, for example, investigating the role of testosterone on the propensity for entrepreneurism. Finally, they are sometimes also used as instruments to estimate causal relationships (for example, random genetic variants are potential instruments for endogenous health states). Looking forward, Understanding Society will begin collecting repeat measures of biomarker data in 2024 (see Section 3). Once these data are available, it will be possible to investigate changes in health over time and the time sequence of the development of disease through investigating biological pathways.

3 | HEALTH AND BIOMARKER DATA IN UNDERSTANDING SOCIETY

In this section, we outline the health data available in Understanding Society and, in particular, the biomarker data collected at wave 2/3, and the plans, and preparation already undertaken, for the data collection that starts in January 2024 on wave 16.

As noted above, Understanding Society, like the BHPS before it, has included self-reported health data across a range of dimensions of health relevant to different life stages (see Table 1). It includes annual measures of long-term health conditions, and frequently employed health scales such as the General Health Questionnaire (GHQ), a widely used and validated measure of psychological distress and the short form 12 (SF12), a validated measure of physical and mental health functioning. The study has rotating modules on health behaviours, sleep and well-being, and periodic measures of cognition and psychological traits. There are also health measures for children, with information on pregnancy and birth, specialist development scales at ages 1 and 3; then, the Strengths and Difficulties Questionnaire (SDQ) is administered to parents about their children up to age 12. Once aged 10, children have their own questionnaire, which includes SDQ and measures of well-being and health behaviours.

As well as self-reported health data, the study has asked participants for their consent to link to wide-ranging administrative data including NHS records. Linked NHS administrative data are available by application to the UK Longitudinal Linkage Collaboration.[9] In the first phase, permission for linkage from NHS Digital was based on analyses focused on COVID-19, and participants' consent was collected as part of our COVID-19 survey (see below). Recently, this approval has been widened, and we hope to add participants who have consented to NHS linkage as part of the main survey.

3.1 | Wave 2/3 biomarker data collection

After the waves 2 and 3 main data collection, participants were offered an interview by a nurse in their home, approximately five months later (full details in user guide[10]).[11] Data were collected at

[9] See https://ukllc.ac.uk/.

[10] Institute for Social and Economic Research, 2022a.

[11] Ethics approval was given by Oxfordshire A REC (Ref: 10/H0604/2) 'Waves 2 and 3: *Understanding Society* – UK Household Longitudinal Study: A Biosocial Component'.

TABLE 1 Health data in Understanding Society

Population group	Overall health	Disease and illness	Physical, mental and cognitive functioning	Positive well-being	Health care use	Health behaviours (various rotation patterns)
Adults, aged 16+	Self-assessed health Health satisfaction	Doctor diagnosis of major conditions	SF-12 – generic physical and mental functioning Limiting illness and specific functional limitations GHQ-12 (from w7, 2yrs) activities of daily living Measures of cognition (w3)	The Warwick-Edinburgh Mental Wellbeing Scales (WEMWBS) (3yrs) Life satisfaction	From w7, Use of GP, hospital and outpatientsFrom w7, every 2yrs, detailed use of health and social care	Smoking (ecigs from w7, quitting from w9) Alcohol use (changed to CAGE from w7) Diet Physical activity (changed to IPAQ from w7) (3yrs) Pittsburgh sleep questionnaire Gambling (w13)
Young adults, aged 15–21						Extra annual – smoking, alcohol, illegal drug
Youth, aged 10–15	Global health (2yrs)		(2yrs) Strength and difficulties questionnaire (SDQ) Limiting health condition (w9)	(2yrs) Self-esteem Life satisfaction		Smoking (ecigs w7) Alcohol consumption Diet Physical activity Illegal drugs Sleep (w12)
Children, aged <10 (completed by responsible adult)	Long-term health conditions		Age-appropriate SDQ at ages 5 and 8 Vineland scale at age 3 Broad development stages			
Pregnancy and birth (completed by mother)	Birthweight and gestation					Smoking and drinking in pregnancy Breast feeding Infant behaviours – crying, sleeping and eating behaviour

TABLE 2 Sample sizes and response rates for wave 2/3 nurse data collection

	GPS (% of eligible) Wave 2 (2010–11)	BHPS (% of eligible) Wave 3 (2011–12)	Total
Eligible adults	26,699	8,864	35,563
Nurse interview	15,591 (58)	5,053 (57)	20,644 (58)
Consent blood for analysis	10,760 (40)	3,673 (41)	14,433 (41)
Consent DNA	10,579 (40)	3,607 (41)	14,186 (40)
Blood sample provided	9,905 (38)	3,342 (38)	13,247 (38)

wave 2 (2010–12) from the General Population Sample (GPS; a stratified clustered sample of the UK households) and wave 3 (2011–12) for the BHPS. All adults who took part in the relevant main wave were selected for the nurse interview, provided they had conducted their main interview in English, and did not live in Northern Ireland. In the second year of wave 2, eligibility was restricted to a random 0.81 sample of the GPS households to accommodate the BHPS interviews, which took place simultaneously.

In total, across the GPS and BHPS sample, 35,563 respondents were eligible for the nurse interview of which 20,644 (58 per cent) took part. Of those, approximately 67 per cent consented to give blood and provide DNA. Blood samples were successfully taken for 13,247 participants. Details are presented in Table 2.

The biomarkers available from this data collection are listed in Table 3. A range of physical measures were taken by the nurse, including blood pressure, lung function, waist circumference, height, weight, body fat and hand grip strength. These measures were conducted according to protocols used by the Health Survey for England. A number of quality indicators are included in the dataset, recorded by the nurse at the time of measurement. In addition, nurses collected venous blood samples, which were used by the Newcastle upon Tyne Hospitals NHS Foundations Trust to produce a number of analytes. The laboratory conducted internal and external quality control analyses; both showed the test results were within acceptable limits. Mean results for each analyte by age groups were compared to data from the Health Survey for England and/or the English Longitudinal Study of Ageing. In general, means were similar across the age distribution.[12] The measures were chosen to capture key chronic diseases, such as heart disease, diabetes, chronic kidney disease and processes relating to stress and ageing, such as inflammatory markers, measures of 'wear and tear' on the immune system and body-building hormones related to growth and decline.

Blood samples were also used to extract DNA and measure genetics and epigenetics, and more recently proteomics (a range of proteins) were measured.

3.2 | Genetic data

In order to conduct genetics research, genotyping is carried out, which identifies differences in DNA sequencing among individuals. Each point of genetic variation on the genome is known as a single nucleotide polymorphism (SNP).[13] At wave 2/3, genotyping was conducted by the Sanger Institute using the Illumina core and exome chip for just over 10,000 samples. At the time, large-scale genotyping was limited to people of White European descent because the reliability of techniques to accurately genotype people varied in different ethnic groups. After data cleaning and other quality control steps, approximately 9,900 samples are available for analysis. An individual's genome contains

[12] Institute for Social and Economic Research, 2022b.

[13] National Human Genome Research Institute, 2023.

TABLE 3 Biomarkers available in Understanding Society

Biomarker	Purpose
Physical measures	
Adiposity Height and weight Waist circumference Percentage of body fat (bioelectrical impedance)	BMI and assessment of excess body fat: obesity and risk factor for range of major chronic conditions
Respiratory function (spirometry) (FVC, FEV_1, PF, FEV_1/FVC)	To detect both obstructive and restrictive respiratory diseases including COPD
Diastolic and systolic blood pressure	Risk factor for stroke and heart conditions, cardio-vascular disease
Resting pulse rate	Measure of general fitness
Grip strength	Indicator muscle strength, associated with functional limitations and disability in older ages
Blood biomarkers	
'Fat in the blood' cholesterol and triglycerides	Associated with heart disease (CVD)
Glucose intolerance – HbA1c	Undiagnosed or poorly managed diabetes
Inflammatory markers – c-reactive protein, fibrinogen	Measures of inflammation – acute due injury or infection, or chronic – response to stress
Anaemia – haemoglobin, ferritin	Marker for poor nutrition; increases with age, with significant health consequences
Liver function – ALP, ALT, AST, GGT, albumin	Associated with alcohol, drugs, obesity, as well as consequence of other diseases
Kidney function – creatinine	Kidney disease increases with age, and is associated other diseases
Hormones – testosterone, IGF1, DHEAS	Associated with stress processes, building muscles, ageing, specifically Testosterone - marker aggression IGF1 – associated diet, diabetes and cancer DHEAS – associated heart disease, muscle strength, cognition
CMV seropositivity	Immunosenescence – wear and tear on immune system, chronic stress, associated diabetes
Genotyping – using Illumina core and exome chip for people of White European descent	Genetic markers for differences in a person's DNA from other individuals, which enable the study of the role of genetics in the development of health and other characteristics
DNA methylation – Illumina EPIC array for people of White European descent	Epigenetics markers, which identify whether, if and how genes are expressed has been altered by the environment
Proteomics – Olink® Target 96 Neurology panel Proteomics – Olink® Target 96 Cardiometabolic panel	Proteins are produced by genes to cause different bodily processes to take place

about six billion genetic sequences; over 99.6 per cent are identical across humans. The Illumina core and exome chip measured approximately 500K SNPs across all genetic sites. These measured SNPs were chosen partly for their own importance and partly as they are informative predictors of other SNPs (both biologically and statistically). They are employed in statistical models to impute over 24M SNPs based on the UK10k reference dataset (a dataset where the whole genome has been measured

for 10,000 people).[14] Quality scores are produced for the imputed data. The measured and/or imputed data are available for research, as outlined below. Analysis of Understanding Society genetics data has replicated 54 previously known associations and identified a number more,[15] demonstrating the data's validity and usefulness for genetics research.

Polygenic scores or indices have been created for different conditions based on an estimation of an individual's genetic predisposition to a specific trait, disease or characteristic. A 'discovery' genome-wide association study is conducted to identify the strength of the association between individual SNPs and the outcome of interest; for example, the genome-wide association study of testosterone identified specific SNPs that explained 10 per cent of variation in testosterone. The polygenic score created from this has been used in studies of social and economic circumstances.[16] A range of different polygenic scores has been deposited with the main datasets.[17]

3.3 | Epigenetics

During the past decade, the study of how the environment and other factors can change the way that genes are expressed (epigenetics) has significantly gained recognition. One of many mechanisms that are involved in the control of gene expression, DNA methylation, has been measured across the genome for all BHPS participants and a random subsample of GPS participants of White European descent. In total, DNA methylation data are available for approximately 3,650 participants. DNA methylation was assessed using the Illumina Infinium HumanMethylationEPIC BeadChip by the University of Exeter. Following quality control and data cleaning,[18] there are some 860K methylation sites measured, across the genome, which is about 3 per cent of all possible sites.[19]

DNA methylation data have been used to create biomarkers of age and deposited with the main datasets.[20] Biomarkers of age are built with supervised machine learning methods, trained against chronological age or wider phenotypes (the label for observed characteristics in biological research) of age. These include: early ('first generation') biomarkers of age, such as 'Horvath'[21] and 'Hannum',[22] which were only trained against age;[23] 'second generation' biomarkers of age 'Lin' and 'Phenoage',[24] which were trained against age, biology and prediction of mortality;[25] and more recent ('third generation') biomarkers of age 'DunedinPoAm'[26] and 'DunedinPACE',[27] thought to reflect the speed of ageing.[28]

[14] Prins et al., 2017.

[15] Prins et al., 2017.

[16] Hughes and Kumari, 2019.

[17] Institute for Social and Economic Research, 2022c.

[18] Mansell et al., 2019.

[19] Pidsley et al., 2016.

[20] Institute for Social and Economic Research, 2021b.

[21] Horvath, 2013.

[22] Hannum et al., 2013.

[23] Hughes et al., 2018.

[24] Levine et al., 2018.

[25] Bao et al., 2022.

[26] Belsky et al., 2020.

[27] Belsky et al., 2022.

[28] Clair, Baker and Kumari, 2023.

3.4 | Proteomics

Proteomics, the analysis of a large set of protein molecules, allows researchers to study different stages on the biological path between genetics and health and other outcomes. Proteins were measured by Olink® for 46 per cent of the blood samples for two areas of disease risk using the Olink® Target 96 Cardiometabolic and Neurology panel, together comprising 184 proteins. Analyses of these data suggest that measurements were associated with the time delay between blood collection and receipt by the laboratory, but these associations were not sufficient to have an impact the relationship between proteins and factors such as educational attainment.[29]

3.5 | Future data collection

Until recently, it has not been possible to repeat the measurement of biological markers for health in Understanding Society. There were a number of reasons for this. The cost of fielding nurses on a study as large as Understanding Society was prohibitive. But also, it was practically difficult given the limited UK-wide capacity of the research nurses and the concurrent demands on them from other biomedical studies. Moreover, in the intervening period, Understanding Society has moved from a face-to-face to mixed-mode study. Participants value the flexibility and choice that mixed-mode surveys offer them. Given this, the study team needed to identify a new way of collecting biomarker data for a reasonable budget and that did not require face-to-face interviews.

Using the study's Innovation Panel at wave 12 (IP12), a mixed-mode biomarker data collection experiment was conducted. The study team investigated whether valid biomarker data could be collected that were consistent across modes and with previous measurements. Households were randomly assigned to three modes of interview: a face-to-face nurse interview, a face-to-face interview, or web-administered data collection.[30,31]

Briefly, in the nurse data collection, height, weight and blood pressure were measured by the same protocol as in wave 2/3, and participants were asked to provide both venous and dried blood samples, which were collected by the nurse. In the interviewer data collection, the interviewer took the same physical measures, and participants were asked to take a dried blood sample themselves. For both these groups, participants were also asked to self-report their height and weight. In the web data collection, participants were asked to provide a dried blood sample. Across the whole sample, participants were asked to measure their own blood pressure in the local pharmacy or at home. Collecting hair samples for adults and children was also attempted but this had very low uptake in children.

The overall response was consistent across nurse, interviewer and web data collections – approximately 60 per cent of households (1,408/2,401). This was a slight improvement on the 57.5 per cent response rate across wave 2/3. Uptake of specific measures (e.g. blood samples) was highest for nurse collection and lowest online.[32] Provision of blood samples significantly increased when selected results were fed back to participants, especially in the web and interview data collections.[33] The distribution of blood pressure measures was similar across modes (more so for systolic than diastolic measures), but the measurement error was bigger in the self-collected measures than in those measured by an interviewer or nurse. As is well known, self-measurement of height and weight was

[29] Dearman, Kumari and Bao, 2023.

[30] Al Baghal et al., 2021.

[31] Ethics approval was granted by East of England: Essex Research Ethics Committee (Ref 19/EE/0146) 'Understanding Society Health Innovation Panel: Biomeasure and health data collection from the Innovation Panel of the UK Household Longitudinal Study'.

[32] See 'A comparison of nurse-led and interviewer and web mixed-mode survey designs: impact on response rates and provision of biomarker samples' by T. Al Baghal et al., submitted to *Public Opinion Quarterly*.

[33] Benzeval et al., 2023b.

biased.[34] Given this, experiments on alternative ways of collecting adiposity data were included at IP15. Tape measures were sent to participants to measure their own waist and hips, and they were asked to download an app (Body Volume Image, BVI) that calculates body size based on a selfie photograph. Compared with interviewer-guided measurement of waist and hips, there was a 3 cm difference with self-measurement for waist and hips, but a 9 cm and 18 cm difference, respectively, with the BVI app output.[35] Self-measurement, but not the app, was therefore taken forward into subsequent data collections.

3.6 | Wave 16

Drawing on this evidence, wave 16 (2024–26) is designed as a health-focused wave with biomarker data collection. It will include the five UKHLS samples: the GPS and BHPS, to maximise the opportunity for investigating changes in health with these data (albeit over a long period); the Ethnic Minority Boost Sample (EMBS) and Immigrant and Ethnic Minority Boost Sample (IEMB), to create opportunities for new research to examining health biomarker distributions among these population groups; and the wave 14 GPS boost, which will increase the cross-section sample for wave 16 analyses and the potential longitudinal sample in future waves (Kumari, Al Baghal and Benzeval, 2022). In the interviewer data collection, the interviewer will measure height, weight, waist, hips and blood pressure. In the web mode, participants will be asked to measure their own waist, hips and blood pressure and self-report their height and weight. In both modes, participants will be asked to provide a capillary blood sample, which will be used to produce the main analytes measured at wave 2/3. A range of studies have validated capillary blood measures against venous blood samples.[36] Wave 16 will include a repeat of the cognition modules carried at wave 3 for face-to-face interviews, and new cognition modules (the Many Brains initiative[37])[38] on the web for all participants to create a new baseline for measuring cognition for the study going forward. The protocol was tested through a pilot study in 2023,[39] and fieldwork for the main wave begins in January 2024 and will end mid-2026,[40] with the data deposited at the UK Data Service at the end of 2026.

After wave 16, DNA will be extracted (with consent) from all participants on whom it has not previously been measured. This will include ethnic minority groups who were not included in the wave 2/3 data collection, participants who were children at the time and are now adults, as well as new sample members. Since the original genetic analysis was conducted, techniques have improved, enabling the whole population to be included in such genetic research. This will create a wide range of new research opportunities, as genetic analysis is often not available on non-White population groups, and there will also be a significant number of multi-generation families with genetic data. Because DNA methylation changes over time for a given individual, DNA methylation will be measured again on all those participants who provide a blood sample and agree to DNA analysis at wave 16. This will provide a unique longitudinal methylation dataset.

[34] Davillas and Jones, 2021.

[35] Serodio, Burton and Jäckle, 2023.

[36] See, for example, Crimmins et al. (2014).

[37] See https://www.manybrains.net/.

[38] The source code for the online tests is available at https://osf.io/gcjzu.

[39] Woods et al., 2023.

[40] Ethics approval was given by the East of England: Essex REC (Ref: 22/EE/0260) 'Understanding Society Wave 16: Biomarker data collection in the UK Household Longitudinal Study'

3.7 | Beyond wave 16

Going forward, biomarkers will be collected every six waves, to provide repeat data at a reasonable interval to capture change across the whole age range. Protocols for these data collections will follow wave 16 to ensure that it is possible to examine genuine change. Where feasible, participants will collect data in the same mode as used in wave 16. While detailed plans for wave 22 are yet to be made, two areas of new development are being investigated by the study team. First, biomarkers that provide insight into the pathways by which the social, biological and health spheres might be connected are crucial. A key example of this is cortisol, the collection of which has had varying success in large-scale population studies.[41] The second key area is to expand biomarker data collection to children in the study.

3.8 | COVID-19 survey

In 2020 and 2021, Understanding Society participants to the main survey were invited to participate in additional data collection focused on their experiences of the COVID-19 pandemic.[42] In March 2021, participants were invited to provide a capillary blood sample, which was tested for COVID-19 antibodies.[43] Of the 13,355 who took part in the relevant COVID-19 survey, 77 per cent requested a blood sample kit, 67 per cent returned it, and it was possible to produce a valid result for 91 per cent of these samples. One study employing these data showed that people who experienced lower social cohesion or loneliness had a lower antibody response to receiving the COVID-19 vaccine than others, with the authors suggesting that improving social cohesion may therefore improve vaccine efficacy.[44]

4 | ILLUSTRATIVE RESEARCH USES IN ECONOMICS

Biomarkers and genetic information have emerged as valuable tools in economic research in three different capacities. First, they have been widely used as outcome variables, allowing researchers to measure the impacts of economic policies or interventions on individuals' health and well-being.[45] Secondly, they have been employed as explanatory variables, shedding light on the relationship between biological factors and economic behaviours.[46] One specific example of this is research investigating the nature versus nurture debates in child development.[47] Finally, an emerging trend within the field is the implementation of biological information as instruments, addressing endogeneity concerns and enabling the establishment of causal links between economic factors and health outcomes.[48]

In a ground-breaking contribution to the field, a pioneering study by Adda and Cornaglia (2006) incorporated biomarkers as outcome variables, to estimate the impact of taxation on smoking behaviour. Specifically, smoking intensity was measured based on cotinine concentration relative to cigarette consumption. The findings revealed a notable trend where smokers increased their smoking intensity in response to higher state-level excise taxes. The results cast doubt on the efficacy of excise

[41] Adam and Kumari, 2009; Halpern et al., 2012; Abell et al., 2016.

[42] Institute for Social and Economic Research, 2021a.

[43] Ethics approval was obtained from London – City & East Research Ethics Committee (Ref: 21/HRA/0644) 'Understanding Society March 2021 COVID-19 survey'.

[44] Gallagher et al., 2022.

[45] Adda and Cornaglia, 2006; Davillas and Pudney, 2017; Carrieri, Davillas and Jones, 2020.

[46] Patel, Wolfe and Williams, 2019; Hand, 2020; Rustichini et al., 2023.

[47] See the forthcoming article by Houmark, Ronda and Rosholm (2023).

[48] von Hinke Kessler et al., 2014; Eibich et al., 2022.

taxes as a regulatory tool for reducing smoking rates. Additionally, the authors advocate for the wider adoption of biomarkers in economic research, highlighting the potential benefits of this innovative approach.[49] In another noteworthy paper that showcases the potential of biomarkers employing Understanding Society data, a comprehensive approach is adopted by incorporating several health indicators, both subjective self-assessments and objective biomarkers, to investigate the alignment of health conditions within marital and cohabiting partners in the UK. By investigating the correlations between health states of partners in relationships, insights are offered into the concordance patterns between various health measures. A distinction is made between assortative mating and the subsequent causal effects of shared lifestyles and environments during the marriage/cohabitation. Findings suggested that shared lifestyle choices and partner selection based on similar health profiles both contribute significantly to the observed health concordance among couples.[50]

Both biomarkers and genetic data have been used as explanatory or mediator variables in a range of economic models. In a very recently published paper by Rustichini et al. (2023), existing theories of the role of parental investment for intergenerational mobility are extended. Their paper departs from traditional models that assume a constant heritability coefficient, demonstrating that heritability is endogenously determined by genotype distribution and mating preferences. The study empirically demonstrates that genetic factors, measured by polygenic score, are associated with educational achievement, particularly college attainment. The research also investigates the role of cognitive skills and personality in mediating genetic effects. Overall, these findings highlight the potential role of genetics, particularly via cognitive skills, in shaping educational outcomes and intergenerational mobility.[51]

Understanding Society's biomarkers have been also used in economic research to explore self-employment from different angles. To gain deeper insights in the relationship between self-employment and physiological outcomes, the concept of allostatic load was employed as a mediator between self-employment and both physical and mental health. Allostatic load is a measure of physiological wear and tear on the body[52] based on eight specific biomarkers, including: albumin, C-reactive protein, mean systolic blood pressure, mean diastolic blood pressure, mean pulse rate, cholesterol total, glycated haemoglobin, and high-density lipoprotein cholesterol. Results indicated a marginal positive association between self-employment and allostatic load, that is, self-employment leads to higher allostatic load (wear and tear). Moreover, self-employment exhibited a marginal negative association on physical health through allostatic load, but there was no noticeable effect of self-employment on mental health through allostatic load.[53] Similarly, research has examined whether testosterone levels are associated with an increased propensity for self-employment.

Finally, although less common, economic research has successfully employed biological information as instrumental variables. For example, the impact of prenatal alcohol exposure on child academic achievement was analysed using a genetic variant in the alcohol metabolism gene ADH1B as an instrument for alcohol exposure. Initial correlations between measured alcohol intake and academic performance yielded mixed results, showing a positive association with wine consumption and a negative association with beer consumption. The study uncovered the endogeneity of alcohol intake, where maternal alcohol consumption/beverage choice is influenced by socio-economic status, as is child academic achievement. However, the genetic instrument employed is unrelated to potential confounders. This showed that low to moderate alcohol exposure during pregnancy may negatively affect childhood and adolescent outcomes.[54] In a more recent study,

[49] Adda and Cornaglia, 2006.

[50] Davillas and Pudney, 2017.

[51] Rustichini et al., 2023.

[52] McEwen and Seeman, 1999.

[53] Patel et al., 2019.

[54] von Hinke Kessler et al., 2014.

Eibich et al. (2022) address concerns about endogeneity when studying the relationship between testosterone levels and unemployment dynamics, by using genetic variants predicting testosterone levels as instruments. Using Understanding Society data, polygenic scores were derived from three genetic markers. Findings confirm the negative impact of testosterone on unemployment risk, for both unemployed and employed men. Additionally, the association between testosterone levels and cognitive and non-cognitive skills, as well as job search behaviour, is explored. Empirical evidence shows that testosterone may influence these traits, which, in turn, affect labour market outcomes. The results have significant implications for labour market policies, highlighting the importance of considering biological factors in understanding job search behaviour.[55]

The use of biomarkers in economics is a developing and evolving field. Biomarkers offer valuable insights into various health-related events, allowing researchers to measure and quantify aspects of human biology that would otherwise remain hidden. These objective measures can enhance the precision and reliability of research findings, potentially reducing measurement error and mitigating the bias associated with self-reported data. Biomarkers also hold promise in uncovering causal relationships, particularly when employed in instrumental variable analyses. However, recurring challenges are limited sample sizes and cross-sectional data, which can constrain the robustness of results. Biomarkers are also subject to variability over time, in terms of both short-term fluctuations and long-term stability, which can complicate the interpretation of findings. Furthermore, the reliance on specific biomarker instruments, the inability to account for self-selection bias and the data being collected from specific groups with particular genetic, demographic and environmental characteristics, represent additional weaknesses, potentially limiting the generalisability of results and their applicability to other populations.

Understanding Society, particularly when wave 16 data become available, overcomes many of these issues, making it a unique and valuable resource for economic research. It provides biomarker data on a large sample of the whole population, and with wave 16 data, a significant subsample including measures at two points in time.

5 | DATA ACCESS AND DATA CITATION

Understanding Society data are freely available to researchers at the UK Data Service (UKDS) (specific DOI links are listed below). Genetics and epigenetics data combined with survey data are available by application from Institute for Social and Economic Research, University of Essex.[56] The study team provides rich metadata and wide-ranging resources on the study website.[57]

It is crucial to the viability of Understanding Society that researchers who use it in their research cite the data by including its bibliographic reference in their reference list. This enables funders to assess the value of the study in terms of the research being produced. It is also beneficial to the researcher.[58] The citations for the wave 2/3 nurse data are:

> University of Essex, Institute for Social and Economic Research and National Centre for Social Research (2023). Understanding Society: Waves 2 and 3 Nurse Health Assessment, 2010– 2012 [data collection]. 5th Edition. UK Data Service. SN:7251, DOI: http://doi.org/10.5255/UKDA-SN-7251-5.

[55] Eibich et al., 2022.

[56] See https://www.understandingsociety.ac.uk/documentation/health-assessment/accessing-data/genetics-application.

[57] See https://www.understandingsociety.ac.uk/documentation/health-assessment.

[58] Benzeval et al., 2023a.

This (EUL) version of the data file includes the biomarkers, proteomics and epigenetic clocks, as well as aggregated information on medications. The special licence version of this dataset includes more detailed information on medications as well as the polygenetic risk scores:

> University of Essex, Institute for Social and Economic Research (2023). *Understanding Society: Waves 2–3 Nurse Health Assessment, 2010–2012: Special Licence Access*. [data collection]. 3rd Edition. UK Data Service. SN:7587, DOI: http://doi.org/10.5255/UKDA-SN-7587-3.

The citations for the main datasets are:

> University of Essex, Institute for Social and Economic Research (2022). Understanding Society: Waves 1–12, 2009–2021 and Harmonised BHPS: Waves 1–18, 1991–2009. [data collection]. 17th Edition. UK Data Service. SN: 6614, DOI: https://doi.org/10.5255/UKDA-SN-6614-18.

> University of Essex, Institute for Social and Economic Research (2021). Understanding Society: Innovation Panel, Waves 1–13, 2008–2020. [data collection]. 11th Edition. UK Data Service. SN: 6849, http://doi.org/10.5255/UKDA-SN-6849-14.

> University of Essex, Institute for Social and Economic Research (2021). Understanding Society: COVID-19 Study, 2020–2021. [data collection]. 11th Edition. UK Data Service. SN: 8644, DOI: https://doi.org/10.5255/UKDA-SN-8644-11.

The study team regularly search for research citing the data and include relevant publications on the website https://www.understandingsociety.ac.uk/research/publications, and in materials about the study, which are shared with participants (to encourage them to continue taking part) and with policy makers through summaries in briefings, invitations to knowledge exchange events etc.

6 | CONCLUSION

The UK, and other developed nations face wide-ranging inter-related economic and public health challenges. For example, England's Chief Medical Officer recently identified the importance of improving older people's quality of life,[59] and the new Office for Health Improvement and Disparities is charged with tackling obesity, improving mental health, promoting physical activity and addressing inequalities.[60] All of these public health priorities have their roots in people's economic circumstances and have significant consequences for the economy and health services. Having high-quality longitudinal data is crucial to identify the causes of health problems for different subgroups of the population and to develop appropriate policy responses. Understanding Society has rich economic and health data collected annually on people of all ages since 2009 (and 1991 for the BHPS sample), with funding available to continue this data collection until wave 22 (2030–32). In addition, it currently has one round of biomarker data together with genetic information and, by the end of this period, there will be a further two waves of biomarker data. Given all of the other benefits for conducting health research in Understanding Society, this makes it a unique resource for such research.

[59] Whitty, 2023.

[60] Hunter, Littlejohns and Weale, 2022.

ACKNOWLEDGEMENTS

Understanding Society is an initiative funded by the Economic and Social Research Council and various government departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service. We are grateful to all of the participants who give up their time, year upon year, to make this study possible. We thank James Banks, Monica Costa Dias and Thomas Crossley for comments on an earlier draft of this paper.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the UK Data Service at http://doi.org/10.5255/UKDA-SN-7251-5, reference number SN: 7251.

ORCID

Michaela Benzeval https://orcid.org/0000-0002-0372-2098
Edith Aguirre https://orcid.org/0000-0002-4912-5483
Meena Kumari https://orcid.org/0000-0001-9716-1035

REFERENCES

Adda, J., & Cornaglia, F. (2006), Taxes, cigarette consumption, and smoking intensity. *American Economic Review*, 96 (4), 1013–28.

Abell, J. G., Stalder, T., Ferrie, J. E., Shipley, M. J., Kirschbaum, C., Kivimäki, M., & Kumari, M. (2016), Assessing cortisol from hair samples in a large observational cohort: the Whitehall II study. *Psychoneuroendocrinology*, 73, 148–56, https://doi.org/10.1016/j.psyneuen.2016.07.214.

Adam, E. K., & Kumari, M. (2009), Assessing salivary cortisol in large-scale, epidemiological research. *Psychoneuroendocrinology*, 34, 1423–36, https://doi.org/10.1016/j.psyneuen.2009.06.011.

Al Baghal, T., Benzeval, M., Burton, J., Crossley, T. F., Kumari, M., & Rajatileka, S. (2021), Collection of biomarkers using nurses, interviewers, and participants: the design of IP12. Understanding Society Working Paper 2021-06, https://www.understandingsociety.ac.uk/research/publications/546970.

Arulsamy, K., & Delaney, L. (2022), The impact of automatic enrolment on the mental health gap in pension participation: evidence from the UK. *Journal of Health Economics*, 86, 102673, https://doi.org/10.1016/j.jhealeco.2022.102673.

Banks, J., Karjalainen, H., & Waters, T. (2023), Inequalities in disability. IFS Deaton Review of Inequalities, https://ifs.org.uk/inequality/inequalities-in-disability/.

Bao, Y. et al. (2022), Social mobility across the lifecourse and DNA methylation age acceleration in adults in the UK. *Scientific Reports*, 12, 22284, https://doi.org/10.1038/s41598-022-26433-2.

Belsky, D. W. et al. (2020), Quantification of the pace of biological aging in humans through a blood test, the DunedinPoAm DNA methylation algorithm. *ELife*, 9, e54870, https://doi.org/10.7554/eLife.54870.

Belsky, D. W. et al. (2022), DunedinPACE, a DNA methylation biomarker of the pace of aging. *ELife*, 11, e73420, https://doi.org/10.7554/eLife.73420.

Benzeval, M., Andrayas, A., Mazza, J., Al Baghal, T., Burton, J., Crossley, T. F., & Kumari, M. (2023b), Does the feedback of blood results in observational studies influence response and consent? A randomised study of the Understanding Society Innovation Panel. *BMC Medical Research Methodology*, 23, 134, https://doi.org/10.1186/s12874-023-01948-y.

Benzeval, M., Crossley, T. F., & Aguirre, E. (2023a), A symposium on Understanding Society, the UK Household Longitudinal Study: introduction. *Fiscal Studies*, 44, (this issue), https://doi.org/10.1111/1475-5890.12355

Benzeval, M., Kumari, M., & Jones, A. M. (2016), How do biomarkers and genetics contribute to Understanding Society? *Health Economics*, 25, 1219–22, https://doi.org/10.1002/hec.3400.

Cabrera-Álvarez, P., James, N., & Lynn, P. (2023), Panel attrition in the General Population Sample and the Immigrant and Ethnic Minority Boost of Understanding Society. Understanding Society Working Paper 2023-03, https://www.understandingsociety.ac.uk/research/publications/547668.

Carrieri, V., Davillas, A., & Jones, A. M. (2020), A latent class approach to inequity in health using biomarker data. *Health Economics*, 29, 808–26, https://doi.org/10.1002/hec.4022.

Clair, A., Baker, E., & Kumari, M. (2023), Are housing circumstances associated with faster epigenetic ageing? *Journal of Epidemiology and Community Health*, forthcoming, https://doi.org/10.1136/jech-2023-220523.

Crimmins, E., Kim, J. K., McCreath, H., Faul, J., Weir, D., & Seeman, T. (2014), Validation of blood-based assays using dried blood spots for use in large population studies. *Biodemography and Social Biology*, 60, 38–48, https://doi.org/10.1080/19485565.2014.901885.

Davillas, A., & Jones, A. M. (2021), The implications of self-reported body weight and height for measurement error in BMI. *Economics Letters*, 209, 110101, https://doi.org/10.1016/j.econlet.2021.110101.

Davillas, A., & Pudney, S. (2017), Concordance of health states in couples: analysis of self-reported, nurse administered and blood-based biomarker data in the UK Understanding Society panel. *Journal of Health Economics*, 56, 87–102, https://doi.org/10.1016/j.jhealeco.2017.09.010.

Dearman, A., Kumari, M., & Bao, Y. (2023), Proteomics in Understanding Society: pre-analytic condition impacts on measurement. Understanding Society Working Paper 2023-05, https://www.understandingsociety.ac.uk/research/publications/547688.

Eibich, P., Kanabar, R., Plum, A., & Schmied, J. (2022), In and out of unemployment—labour market transitions and the role of testosterone. *Economics & Human Biology*, 46, 101123, https://doi.org/10.1016/j.ehb.2022.101123.

FDA–NIH Biomarker Working Group. (2016), BEST (Biomarkers, EndpointS, and other Tools) Resource [Internet]. Glossary, available from https://www.ncbi.nlm.nih.gov/books/NBK338448/.

Gallagher, S., Howard, S., Muldoon, O. T., & Whittaker, A. C. (2022), Social cohesion and loneliness are associated with the antibody response to COVID-19 vaccination. *Brain, Behavior, and Immunity*, 103, 179–85, https://doi.org/10.1016/j.bbi.2022.04.017.

Halpern, C. T., Whitsel, E. A., Wagner, B., & Harris, K. M. (2012), Challenges of measuring diurnal cortisol concentrations in a large population-based field study. *Psychoneuroendocrinology*, 37, 499–508, https://doi.org/10.1016/j.psyneuen.2011.07.019.

Hand, C. (2020), Biology and being green: the effect of prenatal testosterone exposure on pro-environmental consumption behaviour. *Journal of Business Research*, 120, 619–26, https://doi.org/10.1016/j.jbusres.2019.02.034.

Hannum, G. et al. (2013), Genome-wide methylation profiles reveal quantitative views of human aging rates, *Molecular Cell*, 49, 359–67, https://doi.org/10.1016/j.molcel.2012.10.016.

Herd, P., Goesling, B., & House, J. S. (2007), Socioeconomic position and health: the differential effects of education versus income on the onset versus progression of health problems. *Journal of Health and Social Behavior*, 48, 223–38, https://doi.org/10.1177/002214650704800302.

Horvath, S. (2013), DNA methylation age of human tissues and cell types. *Genome Biology*, 14, R115, https://doi.org/10.1186/gb-2013-14-10-r115.

Houmark, M. A., Ronda, V., & Rosholm, M. (2023), The nurture of nature and the nature of nurture: how genes and investments interact in the formation of skills. *American Economic Review*, forthcoming.

Hughes, A., & Kumari, M. (2019), Testosterone, risk, and socioeconomic position in British men: exploring causal directionality. *Social Science & Medicine*, 220, 129–40, https://doi.org/10.1016/j.socscimed.2018.11.004.

Hughes, A., Smart, M., Gorrie-Stone, T., Hannon, E., Mill, J., Bao, Y., Burrage, J., Schalkwyk, L., & Kumari, M. (2018), Socioeconomic position and DNA methylation age acceleration across the life course. *American Journal of Epidemiology*, 187, 2346–54, https://doi.org/10.1093/aje/kwy155.

Hunter, D. J., Littlejohns, P., & Weale, A. (2022), Reforming the public health system in England. *Lancet Public Health*, 7, e797–e800, https://doi.org/10.1016/S2468-2667(22)00199-2.

Institute for Social and Economic Research. (2021a), Understanding Society COVID-19 User Guide, Version 10.0, https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/covid-19/user-guides/covid-19-user-guide.pdf.

Institute for Social and Economic Research. (2021b), Understanding Society: waves 2–3 nurse health, 'epigenetic clocks' derived from DNA methylation, 2010–2012, User Guide, Version 1, https://doc.ukdataservice.ac.uk/doc/7251/mrdoc/pdf/7251_epigenetic_clocks_user_guide.pdf.

Institute for Social and Economic Research. (2022a), Understanding Society: waves 2 and 3 nurse health assessment, 2010–2012, User Guide, Version 3, https://doc.ukdataservice.ac.uk/doc/7251/mrdoc/pdf/7251_nurse_health_assessment_user_guide.pdf.

Institute for Social and Economic Research. (2022b), Understanding Society: Biomarker User Guide and Glossary, Version 2, https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/health/user-guides/7251-UnderstandingSociety-Biomarker-UserGuide-2014.pdf.

Institute for Social and Economic Research. (2022c), Understanding Society: polygenic scores, 2019, User Guide, Version 1, https://doc.ukdataservice.ac.uk/doc/7587/mrdoc/pdf/7587_polygenic_scores_user_guide.pdf.

Kumari, M., & Benzeval, M. (2021), Collecting biomarker data in longitudinal surveys. In P. Lynn (ed.), *Advances in Longitudinal Survey Methodology*, Chichester: Wiley, 26–46.

Kumari, M., Al Baghal, T., & Benzeval, M. (2022), A framework for the re-collection of biomarkers in Understanding Society at wave 16. Understanding Society Working Paper 2022-08, https://www.understandingsociety.ac.uk/research/publications/547469.

Levine, M. E. et al. (2018), An epigenetic biomarker of aging for lifespan and healthspan. *Aging (Albany NY)*, 10, 573–91, https://doi.org/10.18632/aging.101414.

Lynn, P., Cabrera-Álvarez, P., & Clarke, P. (2023), Sample composition and representativeness on Understanding Society. *Fiscal Studies*, 44, (this issue), https://doi.org/10.1111/1475-5890.12357

McEwen, B. S., & Seeman, T. (1999), Protective and damaging effects of mediators of stress: elaborating and testing the concepts of allostasis and allostatic load. *Annals of the New York Academy of Sciences*, 896, 30–47, https://doi.org/10.1111/j.1749-6632.1999.tb08103.x.

Madia, J. E., Moscone, F., & Nicodemo, C. (2023), Studying informal care during the pandemic: mental health, gender and job status. *Economics & Human Biology*, 50, 101245, https://doi.org/10.1016/j.ehb.2023.101245.

Mansell, G., Gorrie-Stone, T. J., Bao, Y., Kumari, M., Schalkwyk, L. S., Mill, J., & Hannon, E. (2019), Guidance for DNA methylation studies: statistical insights from the Illumina EPIC array. *BMC Genomics*, 20, 366, https://doi.org/10.1186/s12864-019-5761-7.

Martínez-Jiménez, M. (2023), Parental nonemployment in childhood and children's health later in life. *Economics & Human Biology*, 49, 101241, https://doi.org/10.1016/j.ehb.2023.101241.

National Human Genome Research Institute. (2023), Single nucleotide polymorphisms (SNPs), https://www.genome.gov/genetics-glossary/Single-Nucleotide-Polymorphisms.

Patel, P. C., Wolfe, M. T., & Williams, T. A. (2019), Self-employment and allostatic load. *Journal of Business Venturing*, 34, 731–51, https://doi.org/10.1016/j.jbusvent.2018.05.004.

Pidsley, R. et al. (2016), Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology*, 17, 208, https://doi.org/10.1186/s13059-016-1066-1.

Prins, B. P. et al. (2017), Genome-wide analysis of health-related biomarkers in the UK Household Longitudinal Study reveals novel associations. *Scientific Reports*, 7, 11008, https://doi.org/10.1038/s41598-017-10812-1.

Rustichini, A., Iacono, W. G., Lee, J. J., & McGue, M. (2023), Educational attainment and intergenerational mobility: a polygenic score analysis. *Journal of Political Economy*, 131, 2724–79, https://doi.org/10.1086/724860.

Serodio, P., Al Baghal, T., Jäckle, A., Burton, J. and Kumari, M. (2023), Can we rely on self-reported biomarker data? Benchmarking the accuracy and feasibility of self-reported hip and waist measurements using analog and digital methods. Section 4.8 in J. Vine et al. (eds), Understanding Society Innovation Panel Wave 15: results from methodological experiments. Understanding Society Working Paper 2023-10. Colchester: University of Essex.

von Hinke Kessler, S., Wehby, G. L., Lewis, S., & Zuccolo, L. (2014), Alcohol exposure in utero and child academic achievement. *Economic Journal*, 124, 634–67, https://doi.org/10.1111/ecoj.12144.

Whitty, C. (2023), Chief Medical Officer's annual report 2023: health in an ageing society, Department for Health and Social Care, London: HMSO, https://www.gov.uk/government/publications/chief-medical-officers-annual-report-2023-health-in-an-ageing-society.

Woods, C., Burton, J., Al Baghal, T., Benzeval, M., Parutis, V., & Kumari, M. (2023), Understanding Society Wave 16 pilot: biomarker data collection in the UK Household Longitudinal Study, Understanding Society Working Paper 2023-04, https://www.understandingsociety.ac.uk/research/publications/547674.

How to cite this article: Benzeval, M., Aguirre, E., & Kumari, M. (2023), Understanding Society: health, biomarker and genetic data. *Fiscal Studies*, 1–17. https://doi.org/10.1111/1475-5890.12354