



**Understanding Society
Working Paper Series**

No. 2022 – 08

October 2022

**A framework for the re-collection of biomarkers in Understanding
Society at Wave 16**

Meena Kumari, Tarek Al Baghal and Michaela Benzeval

University of Essex



Non-technical summary

Understanding the causes of change in society and health, and their consequences for different population groups is fundamental to research and good government. Understanding Society is a unique study for enabling such research as it collects wide ranging information about people's lives annually for people of all ages and from everyone in the household. Adding objective measures of health (biomarkers) to the Study, on a regular basis, will significantly add to opportunities for 'biosocial' research to better understand the two-way relationship between society and health.

Specifically, the following features of Understanding Society, and the interactions between them, make it particularly valuable for biosocial research:

- **Household panel** - many health issues strongly relate to a person's home circumstances and intra-household relationships, and thus there are significant potential benefits to studying health within household units.
- **Large ethnic minority sample** - there is significant and growing interest in better understanding health issues within ethnic minorities, and the study has advantages in terms of sample size and/or representation across age groups and geographic distribution over comparable data resources.
- **Rich longitudinal social data** - there is substantial research potential in linking the social trajectories captured within the study's data over time to the physical health of participants at different points in time. The rich and continuous data collection also supports analyses of natural policy experiments (e.g. different policies or interventions implemented across devolved administrations or at different times) or unanticipated societal events (e.g. the pandemic).

Understanding Society was funded by ESRC to collect biomarker data at waves 2/3 and has recently been funded to re-collect biomarker data at wave 16. This working paper outlines a framework of four possible research themes that such a biomarker data collection could support:

- Understanding the biological pathways that connect society and health.
- Prevalence of undiagnosed/sub-clinical measures in different social groups.
- Measuring impacts associated with macro-change in society (including unforeseen events).
- National representativeness/benchmarking.

It explains why Understanding Society is an effective study through which to support research under each theme and identifies which of the proposed biomarkers to be included in wave 16 will contribute to each theme.

Abstract

Adding the collection of objective measures of health (biomarkers) to Understanding Society, on a regular basis, will significantly add to opportunities for 'biosocial' research to better understand the two-way relationship between society and health. Understanding Society was funded by ESRC to collect biomarker data at waves 2/3 and has recently been funded to re-collect biomarker data at wave 16. This working paper outlines a framework of four possible research themes that such a biomarker data collection could support:

- Understanding the biological pathways that connect society and health.
- Prevalence of undiagnosed/sub-clinical measures in different social groups.
- Measuring impacts associated with macro-change in society (including unforeseen events).
- National representativeness/benchmarking.

It explains why Understanding Society is an effective study through which to support research under each theme and identifies which of the proposed biomarkers to be included in wave 16 will contribute to each theme.

JEL classification: C80 and I1

Keywords: longitudinal, biosocial, health, biomarkers, genetics,

Acknowledgements: Understanding Society is an initiative funded by the Economic and Social Research Council (ESRC) 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. Wave 16 is funded by ESRC grant ES/T002611/1, fieldwork will take place from 2024 to 2026 and data will be deposited in the UK Data Service for distribution at the end of 2026.

Contact: Meena Kumari, (mkumari@essex.ac.uk) Associate Director for Health, Biomarkers and Genetics, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK.

Foreword

Catherine Bromley, Deputy Director of Data Strategy and Infrastructure, ESRC

Understanding Society is a key part of ESRC's commitment to enable biosocial research. The 2019 ESRC Delivery Plan included a long-term ambition for its infrastructure investments to enable "ground-breaking research to provide new insights into how behavioural, environmental and biosocial factors interact to produce different outcomes for individuals and society". ESRC funded biomarker collection in Waves 2 and 3 of *Understanding Society* and invited the team to include proposals for biomarker collection in their proposals for Wave 16. The commissioning panel that reviewed the Wave 16 proposal (in October 2021) agreed in principle with biomarker collection but recommended an extended review period for the study team to develop a broader vision for biomarker collection in *Understanding Society*.

ESRC established a Task and Finish group to advise the *Understanding Society* team. The group included individuals with expertise from across the spectrum of biosocial research, with individuals from both predominantly social and medical backgrounds. There were significant users of *Understanding Society*, and individuals with substantial experience running other data infrastructures which have collected biological data.

The review process highlighted how the following features of *Understanding Society*, and the interactions between them, make it particularly valuable for biosocial research:

- **Household panel** - many health issues strongly relate to a person's home circumstances and intra-household relationships, and thus there are significant potential benefits to studying health within household units.
- **Large ethnic minority sample** - there is significant and growing interest in better understanding health issues within ethnic minorities, and the study has advantages in terms of sample size and/or representation across age groups and geographic distribution over comparable data resources.
- **Rich longitudinal social data** - there is substantial research potential in linking the social trajectories captured within the study's data to the physical health of participants at different points in time. The rich and continuous data collection also supports analyses of natural policy experiments (e.g. different policies or interventions implemented across devolved administrations or at different times) or unanticipated societal events (e.g. the pandemic)

From this a framework four high-level research themes were identified that a biomarker data collection in *Understanding Society* could support:

- Understanding the biological pathways that connect society and health.
- Prevalence of undiagnosed/sub-clinical measures in different social groups.
- Measuring impacts associated with macro-change in society (including unforeseen events).
- National representativeness/benchmarking.

This working paper outlines the framework, approved by ESRC's Management Board in July 2022, and explains why *Understanding Society* is an effective study through which to support each theme. It identifies which themes each of the proposed biomarkers might contribute to and the expected sample sizes for both the overall population and ethnic minority groups.

Introduction

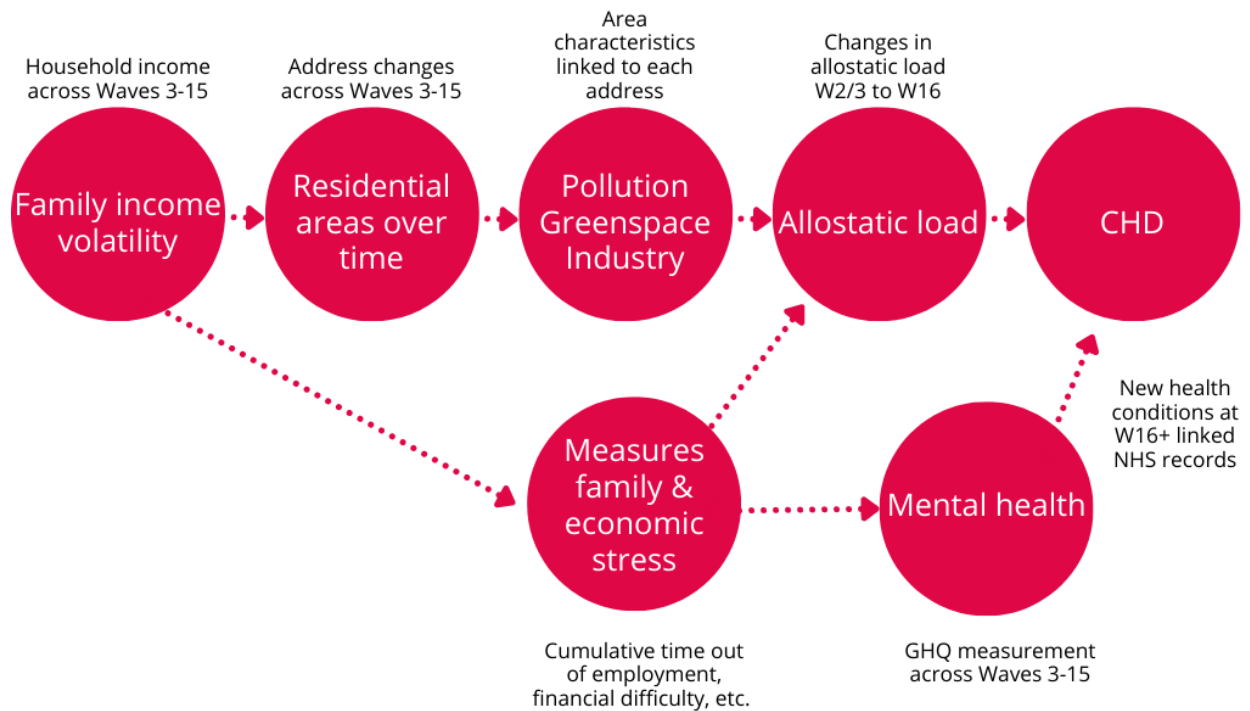
Understanding Society is a nationally representative sample, taking a range of measurements across the entire adult age range and from everyone in the household over time. This makes it unique as a resource for the academic community. In particular, the addition of biological information has afforded a range of interdisciplinary research. Biomarkers were first collected in the study during Waves 2/3 and the repeat collection of biological data will enable the research community to capitalise on the features of *Understanding Society* and its rich, detailed and repeat measurement of the social environment in the UK in a number of ways.

There are four main objectives to collecting biomarkers as part of *Understanding Society*. These are briefly outlined next. The complete list of biomarkers proposed to be collected in Wave 16, including to which objective each biomarker addresses, a brief rationale, whether it was previously collected at Waves 2/3, the longitudinal relevance, and whether it is an innovative or established measure, is presented immediately after in Table 1. This is then followed by a more in-depth narrative for each objective.

Objective A: Understanding the biological pathways that connect society and health

- **What it is:** The social environment and health interact in a bi-directional manner over time such that both can be explanatory factors for outcomes in the other. Social science and health researchers need extensive, high-quality data on both the environment and biology to best understand these associations. Such evidence can inform broad strategies to improve public health and reduce health inequalities as well as promote and maintain a productive workforce.
- **How *Understanding Society* contributes:** The study captures detailed information about the social environment and health, including biomarkers, across the lifespan and intergenerationally within families. The longitudinal core includes annual collection of extensive data on multiple facets of life with additional data provided through a number of administrative record linkages. Biomarker data was also previously collected at Waves 2/3. The biomarkers collected at Wave 16 will provide indicators of a variety of important health outcomes that may be influenced by people's preceding social environment and/or influence their subsequent social trajectories and health. Biomarker and social data will be useful for both cross-sectional and longitudinal analyses.
- *Relevant biomarkers: All biomarkers suggested for collection at wave 16 in Understanding Society are collected to enable researchers to address this purpose. For example, we are proposing cardiovascular factors. Much work has been conducted to understand the minimum biological components needed to assess the 'allostatic load', which includes metabolic markers, inflammatory markers, neuroendocrine markers and blood pressure. Measurement of DNA methylation would enable a recalculation of biological age to calculate within person 'accelerated age'.*

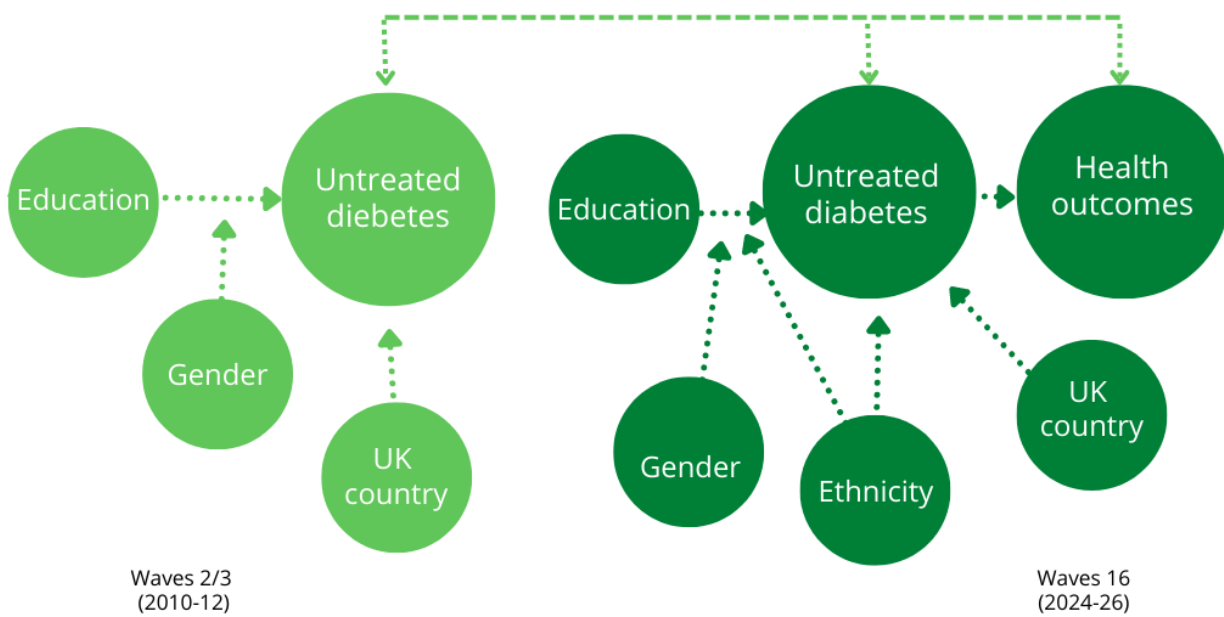
Example: Dynamics of neighbourhood environment and allostatic load



Objective B: Prevalence of undiagnosed condition/sub-clinical measures in different social groups

- **What it is:** The extent of health conditions in the population is not fully identified because in many instances conditions are undiagnosed and/or people are unaware of their symptoms. More research is needed about the extent of these health conditions (at any stage of diagnosis) for different parts of the population and across the entire lifespan in order to design effective prevention and early detection strategies. Understanding the groups that are more likely to have undiagnosed conditions, and factors associated with this, will enable the targeting of prevention policies.
- **How *Understanding Society* contributes:** Including biomarkers that indicate health conditions can identify the prevalence of diseases, including undiagnosed/untreated or poorly controlled cases, in a large representative sample. *Understanding Society* includes participants from sizable numbers of a variety of subgroups across all life stages with a longitudinal design capturing ageing. The nature of this sample allows for studies of prevalence of disease in sub groups of the population not frequently studied, how prevalence changes as people age and over time and track and compare longer term social and health outcomes by diagnosis or treatment status.
- **Relevant biomarkers:** Markers that are used in clinical practice such as total cholesterol, Blood pressure and HbA1c, for the categorisation of diabetes, markers of kidney function and liver function tests are relevant here. Genetic and DNA Methylation data.

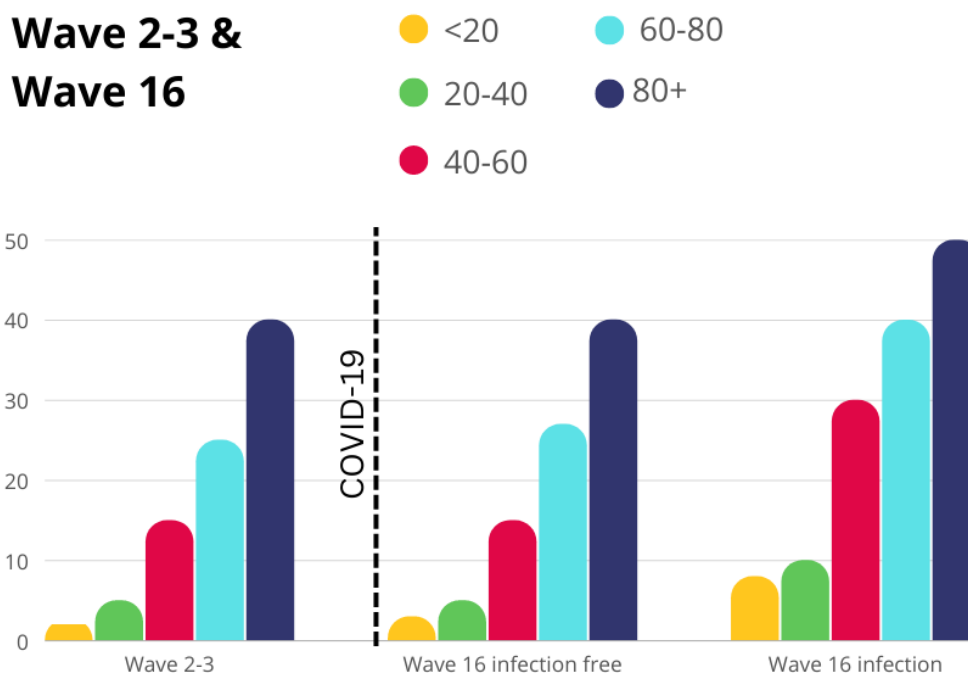
Example: Diabetes in 2010-12 and 2024-6 how do they compare?



Objective C: Measuring impacts associated with macro-change in society

- **What it is:** Society is constantly changing, including unforeseen events, which can have a direct impact on changes in living conditions and health. By capturing the social environment and health of individuals before, during and after these changes emerge, researchers can establish the pathways that societal developments can have on biology.
- **How *Understanding Society* contributes:** The longitudinal design and the large representative sample of the UK makes the study best placed to capture the immediate and long term impacts these societal changes on health through inclusion of biomarkers to the extensive social survey data. Capturing a multitude of biomarkers longitudinally, *Understanding Society* tracks and can make population level inferences of the impact of societal changes on health outcomes. Having before and after indicators through the longitudinal design allow for stronger claims of causality.
- **Relevant biomarkers:** Cardiovascular risk factors such as adiposity, blood pressure, inflammatory markers, kidney markers to examine change in inequalities (for example) and microbiome to see if it related to COVID exposure.

Example: Potential increase in proportion of participants with stage 3 or worse kidney function, measured by eGFR.



Objective D: National representativeness/benchmarking

- **What it is:** The provision of data that is used as a benchmark for similar biomarkers collected in other studies or clinical settings helps contextualise smaller specialised studies and promotes the uptake in usage of biomarkers. The benefits of benchmarking are only possible because *Understanding Society* is representative of the whole population.
- **How *Understanding Society* contributes:** The national and representative nature of the study enables it to be positioned as a benchmark for biomedical studies or routinely held clinical data, and there is precedence for it being used in this way. The extensive nature of the data includes measurements of risk factors for disease which provides population-level understanding of these biomarkers. *Understanding Society* also is at the forefront of setting these benchmarks for emergent measures as it incorporates new biomarkers as the study progresses. This includes newer measures such as polygenic scores or from the microbiome.
- **Relevant biomarkers: clinical risk factors: cardiovascular, kidney markers, liver function tests, genetic markers and microbiome data.**

Examples of use of *Understanding Society* for benchmarking or as a control

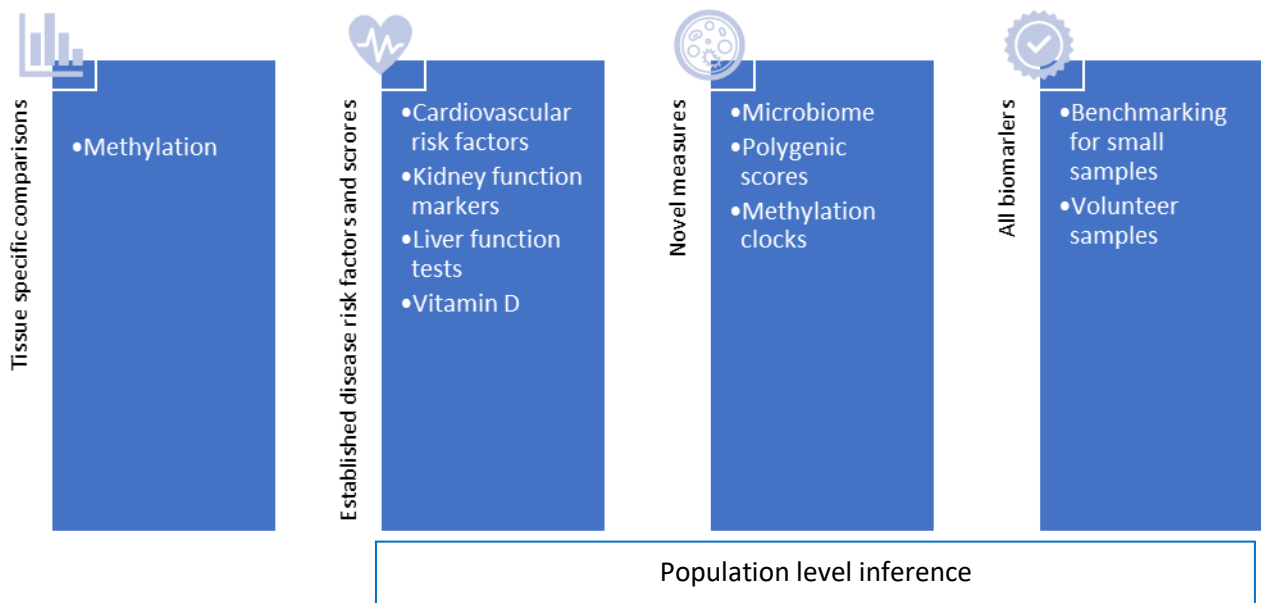


Table 1: Proposed Wave 16 biomarkers

Biomarker	Objectives	Rationale	Measured in Waves 2/3	Relevance of change	Established/ Innovative ^a
Blood pressure	A,B,C,D	Associated with a variety of social exposures: e.g. Socio economic position (SEP) (1). Component of metabolic syndrome (2) and allostatic load (3). Associated with cardiovascular disease and a variety of morbidities (4)	Y	Blood pressure trajectories are patterned by social position and are associated with mortality (5)	Established
Adiposity (Body Mass Index derived from height and weight, waist circumference)	A,C,D	Associated with a variety of social exposures: e.g. SEP (6), adverse childhood events (7); unemployment (8) occupational stress (9) Component of metabolic syndrome (2) and allostatic load (3). Associated with cardiovascular disease and a variety of morbidities (10)	Y	Trajectories in adult adiposity is patterned by social factors particularly in childhood (11).	Established
Blood samples (listed in order of priority)					
Total and HDL cholesterol, Triglycerides	A,B,C,D	Measures of fat in the blood, associated with a variety of social exposures: e.g. SEP (12), occupational stress (13). Component of metabolic	Y	Change in these and the cardiovascular risk factors above inform risk prediction over and above single measures (15)	Established

		syndrome (2) and allostatic load (3). Associated with cardiovascular disease and a variety of morbidities (14)			
HbA1c (glycated haemoglobin)	A,B,C,D	Used to categorise Diabetes. Type 2 diabetes is socially patterned (16) Component of allostatic load (3) Associated with mortality and a variety of morbidities (17)	Y	Demographic, disease risk factors, environmental and psychosocial factors are associated with change in HbA1c (18)	Established
C-Reactive Protein	A,C,D	Inflammatory marker. Socially patterned (19), associated with a number of outcomes including mental health, cardiovascular disease and cancer (20).	Y	Differences in CRP levels raised over time by social position (21), associated with frailty (22)	Established
Dehydroepiandrosterone (DHEAS)	A,C,D	Hormones associated with ageing, components of allostatic load, associated with frailty and cardiovascular disease (23)	Y	Trajectories associated with functioning (24)	Established
Testosterone, Steroid Hormone Binding Globulin	A,C,D	Hormone associated with frailty, type 2 diabetes and cardiovascular disease (25)	Y	Trajectories in total and bioavailable hormone levels associated with age and with various morbidity outcomes and mortality (26)	Established

Creatinine	A,C,D	Measurement of kidney function, which is patterned by SEP (27), and associated with cardiovascular disease and a variety of morbidities (28)	Y	Progression of kidney disease associated with mortality (29)	Established
Ferritin, Haemoglobin	A,D	Reflect iron stores and anaemia, which are associated with diet and with a variety of health outcomes (30)	Y	Changes will track changes in diet and other behaviours and health conditions.	Established
Liver function tests (alkaline phosphatase, alanine transferase, gamma glutamyl transpeptidase, bilirubin)	A,C,D	Reflects exposures such as alcohol intake and poor diet.	Y	Changes will track changes in behaviours and liver function more broadly	Established
Vitamin D (25-hydroxyVit-D)	A,D	Nutritional biomarker associated with a variety of health outcomes.	N		Innovative for UKHLS
DNA for Illumina Global Screening array *	A,B,D	Supplement earlier measures taken from DNA collected from White European groups	Proposed supplement for new sample members (which will increase trio sets) and ethnic groups		Innovative (for some disciplines and population groups and having trios)
Illumina EPIC array (methylation)*	A,C,D	Developed into biomarkers of age (31) that are socially patterned (32) and associated with mortality (as reviewed in (31))	Y	Longitudinal data leads to improvement in biomarker development (33)	longitudinal data is innovative for all, methylation data is innovative in social sciences
Stool sample	A,C,D	Microbiome a new frontier in then the	N		Innovative

		understanding the biology of human health, associated with the environment to a greater extent than with genetics (34).			
--	--	---	--	--	--

*we will keep abreast of developments in technology and price and use a newer array should this one be superseded.

Innovation in *Understanding Society*

DNA extraction for genome wide work and methylation measurements in a wider range of participants: The addition of genetic data will serve to contribute to on-going efforts that are incorporating different ethnicity data to their studies. The collection and extraction of DNA from new members such as the Wave 14 boost, and those that have reached 16+ between Wave 3 and Wave 16 will serve to significantly increase the family groups with genetic data, which is particularly relevant to genetic studies that seek to examine social phenotypes and outcomes in a causal framework as they are susceptible to bias due to familial confounding (35)

Vitamin D (25-hydroxyVit-D): Measurement was supported by the research community as it captures dietary, outdoor and other lifestyle behaviours (36) although there is diverse evidence on its utility (37). Measurement is being included but if blood samples are less than anticipated it is a low priority measure.

Microbiome: It is plausible that gut microbiome may play a role in the pathways that mediate the association of the social environment as health as evidence suggests associations with the psychosocial environment (38) and social relationships (39). We propose the collection, measurement of the composition of the gut microbiota in a subset of participants. Collection of the microbiome will help the community establish whether this is a measure with strong potential to provide insights into novel pathways that may play a role in social inequalities in health.

Objective A) Understanding the Biological Pathways that connect society and health

Of particular interest to social researchers are the mechanisms by which the environment ‘gets under the skin’; social scientists also view the ways in which health influences people’s social and economic circumstances vital for research. There is a continuous and adaptive two-way interaction between the environment and biology throughout the lifespan. *Understanding Society* captures detailed information about the social and physical environment from participants across the adult lifespan and from other sources, such as administrative records. There are several broad pathways that are of interest. The physical environment, for example pollution or occupational toxin exposure and stressful exposures due to social and economic experiences, manifests itself in different biological processes, which with repeated experiences, can lead to long term biological damage, so-called ‘allostatic load’. In turn health may influence people’s ability to work, and having objective measures such as biomarkers enables researchers to eliminate the possibility of subjectivity bias. In the body of work created from data collected at Wave 2/3 (2010-2013), biomarker data have been used individually or in combination and association with the environment. Examples of individual biomarker analyses include a use of

methylation algorithms that represent biological age describing 'accelerated age' in mid-life in those exposed to early life disadvantage (40) and disadvantaged social position across the lifespan (19) and poor housing and higher inflammatory markers (41). In combination, outputs have used the longitudinal data available in the survey to examine changes in labour market status to examine whether staying out of the labour market is better for allostatic load than getting a poor quality job (42). Further living in a disadvantaged neighbourhood (43) was examined in relation to allostatic load, an index comprising markers of multiple physiological systems.

This work serves to pinpoint the association of the social and physical environment with biomarkers that range from genomic to wider physiological measurements. Repeat assessment of measures of biomarkers would strengthen previously reported cross sectional associations and enable analysis of within person change in both the social environment and biology.

We are proposing two new or innovative measures to the Study that would serve to provide new biological pathways by which the environment is associated with health. Vitamin D levels are impacted by light exposure and diet, are patterned by SEP (44) and geography (45). Vitamin D data will provide an insight into diet and sun seeking/outdoor behaviour to complement other data available in the study. Recently the microbiome has emerged as a further mechanism that may underpin the association of the social environment and health (46). Methods to collect and measure the microbiome are new and require some development in longitudinal population survey settings, but promise to put *Understanding Society* at the forefront of studying associations of the social experience and health and the methodological tools to do this.

Relevant biomarkers: All biomarkers suggested for collection at Wave 16 in Understanding Society are collected to enable researchers to address this purpose (see Table 1, below for complete list). The table provides a rationale for each proposed measure and/or biomarker. Much work has been conducted to understand the minimum biological components needed to assess the allostatic load, which includes metabolic markers, inflammatory markers, neuroendocrine markers and blood pressure: we are proposing height, weight, total cholesterol, waist circumference, DHEAs, CRP and diastolic and systolic blood pressure. Measurement of DNA methylation would enable a recalculation of biological age to calculate within person 'accelerated age'.

Objective B) Prevalence of undiagnosed condition/sub-clinical measures in different social groups

Understanding Society has collected biomarkers that are related to specific disease outcomes such as total cholesterol, which is a cardiovascular risk factor and glycated haemoglobin (HbA1C) for the assessment of diabetes. Including such measures in longitudinal surveys enables the identification of the prevalence of diseases in different sub groups of the population and how this changes as people age and over time. Compared to many resources, *Understanding Society* includes participants from a number of life stages, it is the full population with complete coverage in the UK. Thus, there is the ability to examine different parts of the UK with sample sizes that should enable sub-group analyses (see Table 2, below, for expected sample sizes for biomarkers). For example, changes in policy and health service provision means that there are a varying number of untreated people in the population. Data from earlier waves of *Understanding Society* suggests that men and people aged 55-64 and people living alone are at increased risk of untreated hypertension (47), while analysis of HbA1c data suggested an increased risk of untreated diabetes in those with less education. Re-collection of these data will provide insight into the which groups remain at risk for undiagnosed or untreated disease. Additional analyses have described mismatches between self- report smoking and the biomarker of smoking by socio-economic position with implications for the contribution of smoking and health inequalities as mis-match rates are higher in those with more advantaged social position.

The collection of biological data in *Understanding Society* represents one of only a handful of studies with the potential to provide information from large and representative minority ethnic groups. This was not done at wave 2/3. An example of potential cross-sectional analyses afforded by the measurement of HbA1c, for example, at wave 16 include a description of diabetes (48mmol/mol or over) in immigrant and ethnic minority groups and associations with measures and experiences such as racism not collected in other studies. Our forecasted sample size in each group enables researchers to address questions such as whether associations of racism and health vary in different ethnic groups. We plan to collect genetic information from ethnic minority groups which will serve to contribute to on-going consortium efforts that are incorporating different ethnicity data to their studies as seen recently for genome-wide association studies (GWAS) of lipids that found increasing diversity resulted in substantial improvements in fine-mapping functional variants and portability of polygenic prediction (48).

Adding genetic data from the immigrant and ethnic minority boost samples (which was not done at Waves 2/3) will provide valuable data to investigate the health of different groups. This was mentioned by a significant number of users across all fields and would create wide ranging novel multidisciplinary research opportunities. A very strong ambition articulated by a significant subgroup of users, especially genetic epidemiologists, is the unique position *Understanding Society* would be in if we genotyped DNA from minority ethnic groups. Most existing molecular genetic studies are restricted to individuals of white European descent. This reduces the discoveries that can be made, limits generalizability to other ancestral groups, and raises serious issues of justice about who research is conducted for. This is increasingly recognised as a problem, and genotyping arrays have been developed that are much more inclusive. Such genotyping now needs to be carried out on diverse populations and given *Understanding Society's* population coverage and immigrant and ethnic minority boost samples, we could make a hugely important contribution to research in this field.

The household design of *Understanding Society* has enabled analyses (49) that used our methylation data to develop a biomarker for smoking and suggests that there are minimal biological signals with 'passive smoking' in the household. Further, the collection of biomarker data in all adults in the house enable analyses such as those that suggested that there is spousal concordance in adiposity, cardiovascular and diabetes risk. In this work, the length of the spousal relationship is not associated with biomarkers. This observation is from self-reported data and would be strengthened by repeated biomarker information captured many years apart (50).

Relevant biomarkers: Total cholesterol, Blood pressure and HbA1c, for the categorisation of diabetes, markers of kidney function and liver function tests are relevant here. Genetic and Methylation data.

Objective C) Measuring impacts associated with macro-changes in society

COVID-19 is an extreme example of a possibly unexpected event changing society, but new events that can impact both society and health happen are common (and frequently unexpected). Capturing both social and biomarkers longitudinally, *Understanding Society* will be able to track the impact these changes have on the relation between social and health outcomes. Having before and after indicators allow for stronger claims of causality these events have on the relation between social and health outcomes. *Understanding Society* collects a wide variety of social and other information annually which uniquely will enable the creation of detailed and rich histories and trajectories of these factors before and after the collection of biomarker data.

In Wave 16 (2024-2026) it is anticipated that population health will be altered as a consequence of the pandemic. With response to the pandemic, we anticipate that population level disability will be higher

overall and in younger age groups in the post pandemic period than in the pre-pandemic period and thus will the study be able to examine these associations with greater precision than we were previously able. For example, we have collected information on individual experiences in the pandemic, including infection and antibody levels. It might be expected that the virus impacts guts, kidneys, lungs and heart, given where the mediating receptor is expressed (51) and we might anticipate that measurement of these systems will provide any long term associations in the post pandemic era. In particular, the study would potentially be able to examine questions such as which of these systems are important for outcomes such as early retirement or unemployment in working age populations. Repeat measurement of these biomarkers will enable us to understand, for example the increase in subclinical disease in the light of the change health care provision following the pandemic. Analysis of biomarker data from Wave 2/3 quantified health service use prior to the pandemic (52), which we anticipate would be different in the light of greater population ill-health in the post pandemic era.

Further subsequent social and economic conditions of the country will be different to those in wave 2/3 (2010-2013) and our study will enable an examination of the differences in associations in the two time points. For example, a priori one might expect income volatility to be higher in the mid-2020s than in 2010, by having repeat measures of biomarkers at wave 16 of *Understanding Society*, researchers can examine whether great volatility has different associations with cardiovascular disease (53), cognition (54). We additionally expect that associations might vary in the context of different economic environments.

Relevant biomarkers: Cardiovascular risk factors, adiposity, blood pressure, inflammatory markers, kidney markers to examine change in inequalities (for example) and microbiome to see if it related to COVID exposure.

Objective D) National representativeness/benchmarking

The national and representative nature of the study enables it to be positioned as a benchmark to biomedical studies and clinically collected data. There is precedence for the use of *Understanding Society* in this and other ways that capitalise on the unique features of the study. Thus, methylation clocks measured in blood have been compared to these clocks in other tissues (55) and clinical sample sets (56). A repeat measurement of methylation data to calculate the methylation clocks will provide innovation in the study and also a new insight into the expected individual level changes across the adult age range.

Alongside the value of repeating biomarkers already collected, we should look to innovative measurement to stay at the forefront of bio-social research. *Understanding Society* intends to provide such innovative data to the researchers through the collection of gut microbiome, which has been implicated in a variety of diseases and conditions such as depression, anxiety, dementia, obesity and metabolic syndrome. Recent evidence suggests that the environment plays a greater role in shaping gut microbiota than genetics and thus the association of environmental and social factors with the gut microbiota requires further investigation. However, to date studies that collect gut microbiota are typically small, clinical or composed of volunteers and therefore limit generalisability. Microbiome sampling has been conducted in very biased groups, for example in the UK volunteers to the 'mapmygut ([ZOE - Understand how your body responds to food \(joinzoe.com\)](https://www.joinzoe.com))' and US 'American gut project [Home - The Microsetta Initiative \(ucsd.edu\)](https://www.microsetta.org/)'. These have been shown to be unrepresentative in that their participants are disproportionately white, female and healthy. A recent review reported that of the studies that used community based recruitment only 4 of 71 studies used population-based random sampling design (34). There has therefore been a call for the introduction of microbiota collection to

large, representative, richly phenotyped studies to understand the full complexity of previously described associations. There are also key co-variables that are largely not accounted for in this field. We have the opportunity to provide data to this research community on social and environmental factors that may or may not be confounding or playing a role in analyses. A successful collection will provide the basis of an expanded future roll out, providing both methodological knowledge and also a unique insight into population health. For example, evidence suggests that the microbiome is clustered in households (57) and our data will provide an opportunity to add further insight into these observations. Collection of these data in *Understanding Society* will provide information that will also enable population inference.

Relevant biomarkers: all biomarkers, in particular clinical risk factors, biomarkers of ageing (epigenetic clocks), and where data are available from unrepresentative groups (microbiome).

Sample Sizes

Our overall sample estimates, and estimates for specific subgroups, were provided in the Wave 16 response to reviewers and are reproduced in Table 2 below. We anticipate some n=16,000 participants with blood analytes will be well powered to address cross-sectional questions, for example, of inequality between the top 10% and the bottom 10% of the distribution of any socioeconomic indicator, or between two regions each making up about 10% of the sample. A sample size of 10,000 total as we will have with measures of blood pressure or anthropometry in wave 2/3 and wave 16, (1,000 per 10% subgroup) has power of 95% to detect as significant at the 5% level a difference between group means of 16% of the standard deviation of any continuous unimodal outcome variable. This corresponds to a difference between group means of approximately 4% of the range of measurements in the entire population (assuming the range approximates to 4SD). Studies report 1.6-2cm increase in waist circumference over ten years, in groups under the age of 70, which is greater than 0.1SD with similar differences in the most advantaged and disadvantaged SEP (58). We anticipate a substantial number of ethnic minority participants; sample size estimates by ethnic groups and biomarkers are shown in Table 3. While absolute numbers are lower than those in UK Biobank, our participants will be drawn from all age groups and regions of the UK. We will have a greater number of ethnic minority participants than other national surveys, such as Health Survey for England (HSE). For example, in 2019 (the most recent data available), the HSE had 1007 Asian respondents (of any background) and 345 Black respondents. *Understanding Society* expects 3472 Asian and 1285 Black respondents at wave 16. Similarly, the National Diet and Nutrition survey collects Vitamin D information from across the age ranges but this is limited to less than 2,000 adults in total and thus the data available in *Understanding Society* from ethnic minority participants would provide substantial additional insight than previously possible. Our study would also be one of the largest of studies available with repeat measures of methylation using the EPIC array.

Table 2 Forecast samples sizes for key bio-measures at Wave 16 (from response to Wave 16 reviewers' comments)

Measures	Wave2/3 biomarker data collection sample size*	Wave 16 forecast total	Wave 16 forecast (those who took wave 2/3)
Sample issued	35,875	42,000	
Actual/Forecast number of interviews			
Web interview		17,408	
Face-to-face interview	20,699	18,327	
Total sample	20,699	35735	11,000
BP and anthropometry, varies by measure	19,871 – 20,245	21,432	10,716
Blood analyte	13,107	16,000	8,000
Genetics data	9,500 (genome data for White Europeans only)	6,640	Total across w2/3 and w16= 16,640
	Will add 500 Ethnic Minorities (DNA already extracted)		
Epigenetic data	3,654	2,500	1,800
Metagenomic data		2,000	

*This was a follow up to the main interview for 0.81 of the GPS sample and the BHPS sample only, spread over two waves.

Table 3. Forecast samples for Ethnic Minority numbers by Biomarkers

Measures	
<u>Bloods</u>	<u>Expected n</u>
Indian	590
Pakistani	560
Bangladeshi	290
Caribbean	240
African	280
Total	1960
<u>Adiposity</u>	<u>Expected n</u>
Indian	605
Pakistani	578
Bangladeshi	300
Caribbean	254
African	294
Total	2031
<u>Blood Pressure</u>	<u>Expected n</u>
Indian	630
Pakistani	604
Bangladeshi	312
Caribbean	265
African	307
Total	2118
<u>Microbiome</u>	<u>Expected n</u>
Indian	71
Pakistani	68
Bangladeshi	35
Caribbean	30
African	35
Total	239

References

1. Spruill TM. Chronic Psychosocial Stress and Hypertension. *Curr Hypertens Rep*. 2010 Feb;12(1):10–6.
2. Chandola T, Brunner E, Marmot M. Chronic stress at work and the metabolic syndrome: prospective study. *BMJ*. 2006 Mar 4;332(7540):521–5.
3. Juster RP, McEwen BS, Lupien SJ. Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience & Biobehavioral Reviews*. 2010 Sep;35(1):2–16.
4. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. *The Lancet*. 2005 Jan 15;365(9455):217–23.
5. Allen NB, Khan SS. Blood Pressure Trajectories Across the Life Course. *American Journal of Hypertension*. 2021 Apr 2;34(3):234–41.
6. McLaren L. Socioeconomic Status and Obesity. *Epidemiologic Reviews*. 2007 May 2;29(1):29–48.
7. Hemmingsson E, Johansson K, Reynisdottir S. Effects of childhood abuse on adult obesity: a systematic review and meta-analysis. *Obesity Reviews*. 2014;15(11):882–93.
8. Hughes A, Kumari M. Unemployment, underweight, and obesity: Findings from Understanding Society (UKHLS). *Prev Med*. 2017 Apr;97:19–25.
9. Kivimäki M, Head J, Ferrie JE, Shipley MJ, Brunner E, Vahtera J, et al. Work stress, weight gain and weight loss: evidence for bidirectional effects of job strain on body mass index in the Whitehall II study. *Int J Obes*. 2006 Jun;30(6):982–7.
10. Bhaskaran K, dos-Santos-Silva I, Leon DA, Douglas IJ, Smeeth L. Association of BMI with overall and cause-specific mortality: a population-based cohort study of 3.6 million adults in the UK. *The Lancet Diabetes & Endocrinology*. 2018 Dec;6(12):944–53.
11. Caleyachetty R, Stafford M, Cooper R, Anderson EL, Howe LD, Cosco TD, et al. Exposure to multiple childhood social risk factors and adult body mass index trajectories from ages 20 to 64 years. *European Journal of Public Health*. 2021 Apr 24;31(2):385–90.
12. Shohaimi S, Boekholdt MS, Luben R, Wareham NJ, Khaw KT. Distribution of lipid parameters according to different socio-economic indicators- the EPIC-Norfolk prospective population study. *BMC Public Health*. 2014 Aug 28;14:782.
13. Chandola T, Britton A, Brunner E, Hemingway H, Malik M, Kumari M, et al. Work stress and coronary heart disease: what are the mechanisms? *European Heart Journal*. 2008 Mar 1;29(5):640–8.
14. Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, et al. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J*. 2015 Mar 1;36(9):539–50.
15. Lindbohm, J. Association between change in cardiovascular risk scores and future cardiovascular disease: analyses of data from the Whitehall II longitudinal, prospective cohort study | Elsevier Enhanced Reader [Internet]. 2021 [cited 2022 Feb 1]. Available from: <https://reader.elsevier.com/reader/sd/pii/S2589750021000790?token=0C320F68061572CEBD4A6>

3B9E6AE47ACEAE5906DDB2D149970947E8940AA91929DA559C130A3D5D8F67D256B6BA441D1
&originRegion=eu-west-1&originCreation=20220201221520

16. Kumari M, Head J, Marmot M. Prospective Study of Social and Other Risk Factors for Incidence of Type 2 Diabetes in the Whitehall II Study. *Archives of Internal Medicine*. 2004;164:18873–1880.
17. Brewer N, Wright CS, Travier N, Cunningham CW, Hornell J, Pearce N, et al. A New Zealand Linkage Study Examining the Associations Between A1C Concentration and Mortality. *Diabetes Care*. 2008 Jun 1;31(6):1144–9.
18. Luo M, Tan KHX, Tan CS, Lim WY, Tai E, Venkataraman K. Longitudinal trends in HbA1c patterns and association with outcomes: A systematic review. *Diabetes Metab Res Rev*. 2018 Sep;34(6):e3015.
19. Davillas A, Benzeval M, Kumari M. Socio-economic inequalities in C-reactive protein and fibrinogen across the adult age span: Findings from Understanding Society. *Scientific Reports [Internet]*. 2017 Jun 1 [cited 2022 Feb 1];7(1). Available from: <http://www.scopus.com/inward/record.url?scp=85020200585&partnerID=8YFLogxK>
20. Libby P. History of Discovery: Inflammation in Atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2012 Sep;32(9):2045–51.
21. Gimeno D, Brunner EJ, Lowe GDO, Rumley A, Marmot MG, Ferrie JE. Adult socioeconomic position, C-reactive protein and interleukin-6 in the Whitehall II prospective study. *Eur J Epidemiol*. 2007 Sep 20;22(10):675–83.
22. Akbaraly TN, Hamer M, Ferrie JE, Lowe G, Batty GD, Hagger-Johnson G, et al. Chronic inflammation as a determinant of future aging phenotypes. *CMAJ*. 2013 5;185(16):E763–70.
23. Barrett-Connor E, Khaw KT, Yen SSC. A Prospective Study of Dehydroepiandrosterone Sulfate, Mortality, and Cardiovascular Disease [Internet]. <http://dx.doi.org/10.1056/NEJM198612113152405>. Massachusetts Medical Society; 2009 [cited 2022 Feb 1]. Available from: <https://www.nejm.org/doi/10.1056/NEJM198612113152405>
24. Newman AB, Sanders JL, Kizer JR, Boudreau RM, Odden MC, Zeki Al Hazzouri A, et al. Trajectories of function and biomarkers with age: the CHS All Stars Study. *Int J Epidemiol*. 2016 Aug 1;45(4):1135–45.
25. Daka B, Langer RD, Larsson CA, Rosén T, Jansson PA, Råstam L, et al. Low concentrations of serum testosterone predict acute myocardial infarction in men with type 2 diabetes mellitus. *BMC Endocr Disord*. 2015 Dec;15(1):35.
26. Decaroli MC, Rochira V. Aging and sex hormones in males. *Virulence*. 2016 Nov 10;8(5):545–70.
27. Al-Qaoud TM, Nitsch D, Wells J, Witte DR, Brunner EJ. Socioeconomic Status and Reduced Kidney Function in the Whitehall II Study: Role of Obesity and Metabolic Syndrome. *American Journal of Kidney Diseases*. 2011 Sep;58(3):389–97.
28. Thomas R, Kanso A, Sedor JR. Chronic Kidney Disease and Its Complications. *Prim Care*. 2008 Jun;35(2):329–vii.
29. Guo Y, Cui L, Ye P, Li J, Wu S, Luo Y. Change of Kidney Function Is Associated With All-Cause Mortality and Cardiovascular Diseases: Results From the Kailuan Study. *J Am Heart Assoc*. 2018 Oct 24;7(21):e010596.

30. Culleton BF, Manns BJ, Zhang J, Tonelli M, Klarenbach S, Hemmelgarn BR. Impact of anemia on hospitalization and mortality in older adults. *Blood*. 2006;107(10):3841–6.
31. Oblak L, van der Zaag J, Higgins-Chen AT, Levine ME, Boks MP. A systematic review of biological, social and environmental factors associated with epigenetic clock acceleration. *Ageing Res Rev*. 2021 Aug;69:101348.
32. Hughes A, Smart M, Gorrie-Stone T, Hannon E, Mill J, Bao Y, et al. Socioeconomic Position and DNA Methylation Age Acceleration Across the Life Course. *Am J Epidemiol*. 2018 Nov 1;187(11):2346–54.
33. Belsky DW, Caspi A, Corcoran DL, Sugden K, Poulton R, Arseneault L, et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. Deelen J, editor. *eLife*. 2022 Jan 14;11:e73420.
34. Renson A, Herd P, Dowd JB. Sick Individuals and Sick (Microbial) Populations: Challenges in Epidemiology and the Microbiome. *Annual Review of Public Health*. 2020;41(1):63–80.
35. Brumpton B, Sanderson E, Heilbron K, Hartwig FP, Harrison S, Vie GÅ, et al. Avoiding dynastic, assortative mating, and population stratification biases in Mendelian randomization through within-family analyses. *Nat Commun*. 2020 Jul 14;11(1):3519.
36. Macdonald HM, Mavroeidi A, Barr RJ, Black AJ, Fraser WD, Reid DM. Vitamin D status in postmenopausal women living at higher latitudes in the UK in relation to bone health, overweight, sunlight exposure and dietary vitamin D. *Bone*. 2008 May;42(5):996–1003.
37. Reijven PLM, Soeters PB. Vitamin D: A magic bullet or a myth? *Clin Nutr*. 2020 Sep;39(9):2663–74.
38. Allen AP, Dinan TG, Clarke G, Cryan JF. A psychology of the human brain–gut–microbiome axis. *Social and Personality Psychology Compass*. 2017;11(4):e12309.
39. Dill-McFarland KA, Tang ZZ, Kemis JH, Kerby RL, Chen G, Palloni A, et al. Close social relationships correlate with human gut microbiota composition. *Sci Rep*. 2019 Jan 24;9:703.
40. Hughes A, Smart M, Gorrie-Stone T, Hannon E, Mill J, Bao Y, et al. Socioeconomic Position and DNA Methylation Age Acceleration Across the Life Course. *Am J Epidemiol*. 2018 Nov 1;187(11):2346–54.
41. Clair A, Hughes A. Housing and health: new evidence using biomarker data. *J Epidemiol Community Health*. 2019 Mar;73(3):256–62.
42. Chandola T, Zhang N. Re-employment, job quality, health and allostatic load biomarkers: prospective evidence from the UK Household Longitudinal Study. *International Journal of Epidemiology*. 2018 Feb 1;47(1):47–57.
43. Prior L, Manley D, Jones K. Stressed out? An investigation of whether allostatic load mediates associations between neighbourhood deprivation and health. *Health Place*. 2018 Jul;52:25–33.
44. Hayden KE, Sandle LN, Berry JL. Ethnicity and social deprivation contribute to vitamin D deficiency in an urban UK population. *The Journal of Steroid Biochemistry and Molecular Biology*. 2015 Apr 1;148:253–5.
45. Aspell N, Laird E, Healy M, Shannon T, Lawlor B, O’Sullivan M. The Prevalence and Determinants of Vitamin D Status in Community-Dwelling Older Adults: Results from the English Longitudinal Study of Ageing (ELSA). *Nutrients*. 2019 Jun 1;11(6):1253.

46. Bowyer RCE, Jackson MA, Le Roy CI, Ni Lochlainn M, Spector TD, Dowd JB, et al. Socioeconomic Status and the Gut Microbiome: A TwinsUK Cohort Study. *Microorganisms*. 2019 Jan 11;7(1):17.
47. Petersen J, Benzeval M. Untreated hypertension in the UK household population - Who are missed by the general health checks? *Prev Med Rep*. 2016 Dec;4:81–6.
48. Graham SE, Clarke SL, Wu KHH, Kanoni S, Zajac GJM, Ramdas S, et al. The power of genetic diversity in genome-wide association studies of lipids. *Nature*. 2021 Dec;600(7890):675–9.
49. Hulls PM, de Vocht F, Bao Y, Relton CL, Martin RM, Richmond RC. DNA methylation signature of passive smoke exposure is less pronounced than active smoking: The Understanding Society study. *Environ Res*. 2020 Nov;190:109971.
50. Davillas A, Pudney S. 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*. 2017 Dec 1;56:87–102.
51. Hamming I, Timens W, Bulthuis M, Lely A, Navis G, van Goor H. Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *The Journal of Pathology*. 2004;203(2):631–7.
52. Davillas A, Pudney S. Using biomarkers to predict healthcare costs: Evidence from a UK household panel. *Journal of Health Economics*. 2020 Sep 1;73:102356.
53. Elfassy T, Swift SL, Glymour MM, Calonico S, Jacobs DR, Mayeda ER, et al. Associations of Income Volatility with Incident Cardiovascular Disease and All-cause Mortality in a US cohort: 1990 to 2015. *Circulation*. 2019 Feb 12;139(7):850–9.
54. Grasset L, Glymour MM, Elfassy T, Swift SL, Yaffe K, Singh-Manoux A, et al. Relation between 20-year income volatility and brain health in midlife. *Neurology*. 2019 Nov 12;93(20):e1890–9.
55. Shireby GL, Davies JP, Francis PT, Burrage J, Walker EM, Neilson GWA, et al. Recalibrating the epigenetic clock: implications for assessing biological age in the human cortex. *Brain*. 2020 Oct 29;143(12):3763–75.
56. El Khoury LY, Gorrie-Stone T, Smart M, Hughes A, Bao Y, Andrayas A, et al. Systematic underestimation of the epigenetic clock and age acceleration in older subjects. *Genome Biol*. 2019 Dec 17;20:283.
57. Gacesa R, Kurilshikov A, Vich Vila A, Sinha T, Klaassen M a. Y, Bolte LA, et al. Environmental factors shaping the gut microbiome in a Dutch population. *Nature*. 2022 Apr;604(7907):732–9.
58. Zaninotto P, Lassale C. Socioeconomic trajectories of body mass index and waist circumference: results from the English Longitudinal Study of Ageing. *BMJ Open*. 2019 Apr 20;9(4):e025309.
59. Benzeval M, Andrayas A, Mazza J, Al Baghal T, Burton J, T Crossley TF, and Kumari M (2022) Does the feedback of blood results in observational studies influence response and consent? A randomised study of the Understanding Society Innovation Panel, 19 July 2022, PREPRINT (Version 1) available at Research Square <https://doi.org/10.21203/rs.3.rs-1847006/v1>

60. Kumari M, Andrayas A, Al Baghal T, Burton J, Crossley T, Jones K S, et al. Nurse- and self-collected dried blood spots for the assessment of cardiovascular risk factors: A randomised study of the Understanding Society Innovation Panel. in preparation.
61. Al Baghal T, Burton J, Benzeval M, Crossley T, Kumari M, Lynn P, et al. Collection of biomarkers using nurses, interviewers, and participants: Differences in response rates and take-up rates. Understanding Society working paper, in preparation.
62. Crossley T, Mazza J, Al Baghal T, Burton J, Kumari M, Benzeval M. A comparison of nurse and participant led blood pressure measurement in the Understanding Society Innovation Panel. In preparation.