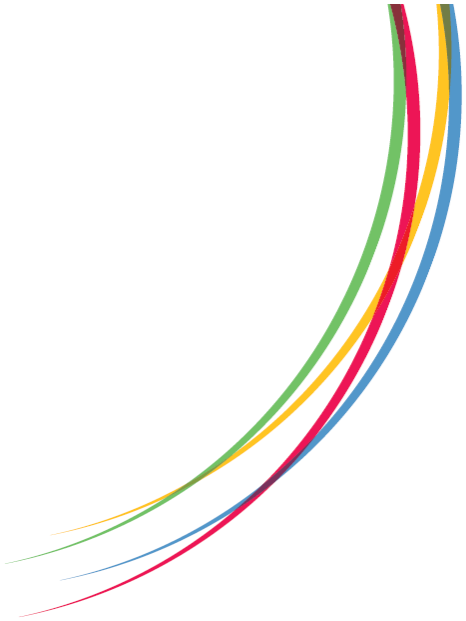


Understanding Society: The UK Household Longitudinal Study

Biomarker User Guide and Glossary

Prepared by
Michaela Benzeval, Apostolos Davillas, Meena Kumari, Peter Lynn

Institute for Social and Economic Research
University of Essex
Colchester
Essex

Version 1, December 2014

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

ACKNOWLEDGEMENTS

Understanding Society the UK Household Longitudinal Study is an initiative funded by the Economic and Social Research Council (ESRC) with scientific leadership by the Institute for Social and Economic Research (ISER), University of Essex, and survey delivery by the National Centre for Social Research (NatCen) and TNS BMRB. The study has also been supported by the Department for Work and Pensions, the Department for Education, the Department for Transport, the Department for Culture, Media and Sport, the Department for Communities and Local Government, the Scottish Government, the Welsh Government, the Department for Environment, Food and Rural Affairs, the Food Standards Agency, the Office for National Statistics, and the Department of Health.

We are grateful to all of the respondents who gave up their time to take part in the nurse health assessment interviews, completed physical measures and provided blood samples to improve our understanding of people's health. The nurse health assessment interviews were conducted by NatCen; we are grateful to the interviewers, nurses and fieldwork and survey teams both at NatCen and ISER for all of their efforts to produce high quality data for the study. The blood samples were initially received, processed and frozen by Fisher BioServices; the biomarkers reported here were analysed by Newcastle upon Tyne Hospitals NHS Foundations Trust. We are grateful to both laboratories and their staff for the efficient and high quality processing of the samples.

TABLE OF CONTENTS

LIST OF BOXES, FIGURES AND TABLES

INTRODUCTION

Understanding Society: the UK Household Longitudinal Study is a large longitudinal survey of households in the United Kingdom. Information is collected on the household; all young people aged 10-15 are asked to do a self-complete questionnaire; and, all adults 16 and over are invited to take part in an interview. Households recruited at the first round of data collection are visited each year to collect information on changes to their household and individual circumstances.

In 2010-2012 (Waves 2 or 3), after the annual survey, adult respondents were also invited to take part in a nurse health assessment interview, which included a range of physical measures and blood samples. With consent the blood samples were frozen for future analysis and DNA extracted. Some of the blood samples have now been analysed to produce a set of biomarkers, which are characteristics that are *'objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention'*[1]. We have selected a range of biomarkers that are either measures of key risk factors for diseases which are major public health problems and/or reflect key biological pathways between social and environmental factors and health.

The purpose of this guide is to outline the biomarkers currently available in *Understanding Society* and some of the factors that require consideration in their analysis. Full details about the nurse visit, and the other data collected as part of this, can be found in the Nurse Health Assessment User Guide[2], CAPI programme[3] and fieldwork protocols.[4] More information on the main annual survey can be found in the main user guide;[5] all of these guides are available on the *Understanding Society* website[6]. Below we briefly describe the sample who took part in the nurse health assessment and the procedures for taking blood. We then outline the criteria for deciding on the biomarkers to be analysed, the general approach to their analysis and quality control measures. The main part of this guide provides a glossary for the biomarkers included in *Understanding Society*. After an initial overview of issues associated with analysing biomarker data, for each biomarker we outline:

- the clinical significance of each biomarker,
- a description of the role of the biomarker in the body;
- laboratory methods and procedures used to measure the analyte;
- guidance on factors to consider when analysing the biomarker;
- its distribution in the *Understanding Society* sample.

Not all of the blood samples available have been used for these biomarkers. Significant samples remain frozen for future use. In due course we will both advertise their availability for researchers

[1] National Institute of Health Biomarkers Definitions Working Group (1998) referenced in Biomarkers Definitions Working Group (2001) Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 69: 89–95.
[2] McFall, Stephanie L., Petersen, Jakob, Kaminska, Olena, Lynn, Peter. (2014). *Understanding Society* –UK Household Longitudinal Study: Waves 2 and 3 Nurse Health Assessment, 2010-2012, Guide to Nurse Health Assessment. Colchester: University of Essex. https://www.understandingsociety.ac.uk/documentation/health-assessment
[3] https://www.understandingsociety.ac.uk/documentation/health-assessment/questionnaires
[4] NatCen (2010) Nurse Protocols for Measurements and samples used by the National Centre for Social Research, London: NatCen. Understanding Society and NatCen (2010) Understanding Society Nurse Visit Nurse Project Instructions, Colchester: University of Essex. https://www.understandingsociety.ac.uk/documentation/health-assessment/fieldwork-documents
[5] Knies, Gundi (2014) *Understanding Society* –UK Household Longitudinal Study: Wave 1-4, 2009-2014, User Manual. Colchester: University of Essex, https://www.understandingsociety.ac.uk/documentation/mainstage.
[6] https://www.understandingsociety.ac.uk/documentation

to apply to analyse them, but also, based on consultation, seek further funding ourselves for analyte users tell us are important. DNA was also extracted from the blood samples and analysed by a genome wide scan by the Wellcome Trust Sanger Institute. These data are also available for researchers to analyse, information on how to apply for them is on the *Understanding Society* website[7].

OVERVIEW OF THE NURSE HEALTH ASSESSMENT

Understanding Society has four samples[8]:

- the General Population Sample (GPS)– a stratified clustered sample of households representative of the general population of the United Kingdom in 2009;
- the Ethnic Minority Boost sample – an additional sample of 1,000 adults in each of the five largest ethnic minority groups in the UK;
- the Innovation Panel – 1,500 households – a sample for methodological research;
- the British Household Panel Survey (BHPS) – a longitudinal study begun in 1991, with 8,000 households incorporated into *Understanding Society*.

The nurse health assessments were conducted with adult participants from the GPS and BHPS samples only. For the GPS sample, the nurse health assessment was undertaken in Wave 2 and for the BHPS sample it was conducted in Wave 3. In both cases the nurse visit took place approximately 5 months after the main interview. Data collection began in May 2010 and was completed in July 2012 for the Wave 2 nurse assessment. For the Wave 3 nurse assessment data collection began in June 2011 and ended in July 2012. Respondents were eligible for a nurse interview if they had completed a full face-to-face interview in the corresponding Wave, were aged 16 or older, lived in England, Scotland or Wales, completed their interview in English[9] and for women were not pregnant. Given limitations with the number of nurse interviewers, in the second year of Wave 2, eligibility was further restricted to 0.81 of the primary sampling units (PSUs) in England. Ethical approval for the nurse health assessment was obtained from the National Research Ethics Service (*Understanding Society* - UK Household Longitudinal Study: A Biosocial Component, Oxfordshire A REC, Reference: 10/H0604/2).

Overall, 57.9% of those eligible for the nurse health assessment took part from the GPS sample and 56.6% from the BHPS sample. The data from the survey questions and physical measures for both Wave 2 and 3 (SN 7251) have been deposited at the UK Data Service (UKDS) under an End User Licence (EUL)[10]. The biomarker variables outlined below have been added to these data files.

[7] https://www.understandingsociety.ac.uk/
[8] Lynn P. Sample design for Understanding Society. Understanding Society Working Paper Series 2009: Available from: https://research.understandingsociety.ac.uk/publications/working-paper/2009-01.
[9] Respondents could request survey materials and a nurse visit in Welsh, as required by the Welsh Language Act. However, such requests were not made.
[10] http://discover.ukdataservice.ac.uk/catalogue/?sn=7251&type=Data%20catalogue

THE BLOOD SAMPLES AND ANALYSIS

All respondents were eligible to give blood except individuals who volunteered that they were HIV positive or had hepatitis B or C and people with clotting or bleeding disorder such as haemophilia or low platelets. Clotting disorders did not include a history of thrombophlebitis, a deep venous thrombosis, a stroke caused by a clot, a myocardial infarction or an embolus. Finally people who have ever had a fit, or those taking anti-clotting medication, e.g. warfarin, were excluded. Aspirin was not counted as an anti-clotting medication.

Respondents gave written consent for their blood to be taken, the storage of blood for future scientific analyses and for genetic analysis as outlined in Box 1 below. The information leaflet given to respondents about *Giving a blood sample* can be seen at https://www.understandingsociety.ac.uk/documentation/mainstage/fieldwork-documents. Both the participant and the nurse signed the consent form. Respondents aged 16 and 17 years old were asked to consent to their own participation. However, nurses were advised to check with parents (when present), as a matter of courtesy, before taking a blood sample from this age group.

Box 1 Consent wording for blood samples

I consent to a qualified nurse taking a sample of my blood on behalf of the Institute for Social and Economic Research/National Centre for Social Research.

1. I have read and understood the Information for Participants leaflet about the second stage of the survey. The nurse has explained the procedures, and I have had an opportunity to discuss these with him/her.

2. I consent to my blood being taken, stored and used in scientific research. I understand that all blood test results and related information will be coded so I cannot be identified. For purposes of scientific analyses, links to my name will be held separately and securely from any data collected. The sample will not be tested for HIV. I also understand my right to withdraw consent for storing the blood sample.

3. I give my consent for a sample of my DNA to be taken from my blood, stored and used in scientific research.

I understand that:

• the DNA samples and related information will be coded to ensure that my personal identity is not revealed to researchers carrying out scientific analysis
• links to my name will be held separately and securely, for administering the study and data collection
• that no personal test results from my DNA will be given to me
• the data and samples will be owned by the Study and the ESRC. No samples or information will be sold
• The DNA analyses will not be used for paternity analysis, life insurance, mortgage applications or police records

I also understand my right to withdraw consent for storing the blood sample

Following written consent from eligible participants, non-fasting blood samples were collected into the following tubes: 1 x 6 ml red plain tube for subsequent extraction of serum, 1 x 1.8 ml

light blue tube with citrate for plasma extraction, 3 x 4 ml purple EDTA tubes (to prevent clotting for plasma, whole blood and DNA). The tubes were labelled with the participant's serial number and date of birth before taking the sample. They were packed in a 6 tube transport container and despatched using Royal Mail to the Fisher BioServices secure storage facility. Storage facility staff reconciled the sample with consent forms and visually inspected the tubes. They applied a unique bar code which is used in sample retrieval. Samples were processed - ie to separate plasma and serum and extract DNA - placed in smaller tubes (aliquots) and stored in freezers at -80 degrees C.

On average it took 2.6 days from the time of blood collection to the samples being processed by the Lab; 90% of samples were processed within 4 days. Below we note issues with the length of time to processing, and we are planning some small scale experiments to further understand the robustness of analyte measurement to long delays in sample processing. The results of which will be made available in due course.

ANALYSIS OF SAMPLES

In 2013, after successfully securing funding, ISER issued an invitation to tender for the analysis of the *Understanding Society* blood samples. The tender was won by Newcastle upon Tyne Hospitals NHS Foundations Trust (NUTH), which has considerable experience of undertaking blood analysis for research purposes, having conducted the analysis of samples from the Health Survey for England (HSE) and English Longitudinal Study of Ageing (ELSA) for a number of years as well as other studies.

Batches of frozen samples for 2000 respondents were transferred to NUTH per month between December 2013 and July 2014. On delivery the bar codes were scanned and samples aliquoted into relevant tubes for the different analytical machines and labelled with a unique bar code in the NUTH system. In order to minimise the use of blood, all serum analytes were run on a single Roche machine requiring 2x250uL of serum, with the exception of one biomarker which had to be analysed on a separate analyser, and therefore required a further 250uL of serum. Two 250 uL of citrated plasma were required and 204uL of whole blood. The latter was sampled from the 4mL EDTA storage tube, which was then refrozen and returned to the storage facility. Results were transferred electronically from the analysers into a patient management information system and exported and sent to ISER.

All tests were undertaken according to the Standard Operating Procedures by HCPC Registered Biomedical Scientists. Internal Quality Controls (IQC) were run on each machine at regular intervals per day. External Quality Assurance (EQA) systems were in place to monitor all tests. Both internal and external quality assessments were regularly reviewed by the quality team and any trends identified escalated through internal governance arrangements; any non-conformities with EQAs were raised and investigated with the scheme provider. The results from the IQC and EQA for the period *Understanding Society* samples were analysed at NUTH are summarised below in Appendix 1, and key points highlighted in the individual biomarker glossary entries. A full technical report is being drawn together and will be made available on the *Understanding Society* website in due course.

THE RESPONDENTS WHO TOOK PART

Full details of the eligibility and exclusions for the overall nurse visit can be found in the Nurse Health Assessment User Guide[11]. Here we provide a broad summary of those respondent who were eligible and took part, and more detail on the eligibility, consent and valid samples for the data arising from the blood samples. As shown in Table 1, across both the GPS and BHPS samples, of the 35,937 respondents eligible for the nurse health assessment, 20,700 took part (a response rate of 57.6%). Of those participating in the nurse health assessment, 1,579 (7.6%) were ineligible to give blood, and a further 22.6% (4,688 people) refused. Reasons given by respondents for non-consent (not mutually exclusive) included: dislike/fear of needles (42.7%), recently had blood test/health check (14%), previous difficulties with venipuncture (14%), no information about what blood will be tested for (12.7%), no feedback of results (12.2%), current illnesses (3.1%) and other reasons (8.8%). Of those eligible and consenting to give blood samples to be stored for future analysis, samples were obtained and successfully processed (at least one biomarker available) for 13,107 respondents. This represents a response of 36.5% of those eligible for the nurse health assessment; 68.5% of those who participated in the nurse interview, and 90.8% of those who consented. All of the response rates were slightly higher among men than women.

[11] McFall, Stephanie L., Petersen, Jakob, Kaminska, Olena, Lynn, Peter. (2014). *Understanding Society* –UK Household Longitudinal Study: Waves 2 and 3 Nurse Health Assessment, 2010-2012, Guide to Nurse Health Assessment. Colchester: University of Essex. https://www.understandingsociety.ac.uk/documentation/health-assessment

Table 1 Eligibility, missing cases and participation in the blood sample: Combined GPS and BHPS sample components.

		Gender	
	Whole sample	Male	Female
Eligible for the nurse visit	**35,937 (100%)**	**15,864 (100%)**	**20,073 (100%)**
Reasons for non-participation in nurse visit			
Pregnant, ill, died, out of scope	349 (1.0%)	80 (0.5%)	269 (1.3%)
No contact	5,534 (15.4%)	2,581 (16.3%)	2,953 (14.7%)
Refused the nurse visit	9,354 (26.0%)	4,103 (25.9%)	5,251 (26.2%)
People participating in the nurse visit	***20,700***	***9,100***	***11,600***
Reasons for no blood sample			
On anti-coagulant drugs, had clotting/bleeding disorder or have had a fit in the last 5 years	1,579 (4.4%)	750 (4.7%)	829 (4.1%)
People eligible for giving blood	**19,121**	**8,350**	**10,771**
No consent to take or store blood sample	4,387 (12.2%)	1,900 (12.0%)	2,487 (12.4%)
No consent to take blood: Reported inability to give blood sample	301 (0.8%)	120 (0.8%)	181 (0.9%)
People eligible to give blood who consented	**14,433**	**6,330**	**8,103**
Unable to give blood sample during the fieldwork (no suitable/palpable veins, collapsed veins, anxious/nervous, felt faint/fainted)	1105 (3.1%)	380 (2.4%)	725 (3.6%)
Unable to process samples/other missing cases	221 (0.6 %)	100 (0.6 %)	121 (0.6 %)
At least one biomarker available	***13,107***	***5,850***	***7,257***
Response rates			
At least one biomarker available as % of total eligible	36.5 %	36.9 %	36.2 %
At least one biomarker available as % of people eligible for giving blood	68.5 %	70.0 %	67.4 %
At least one biomarker available as % of people who gave blood consent	90.8 %	92.4 %	89.6 %

CHOICE OF BIOMARKERS

A number of criteria were considered in identifying biomarkers that should be included in the first set of blood analytes from the *Understanding Society* blood samples. We felt it was important to include biomarkers where there was:

- an environmental (socioeconomic, physical, psychosocial) and/or behavioural effect on marker;
- evidence of pathways to important health outcomes or it was a marker for important health conditions;
- a reasonable proportion of general population affected by the biomarker and a reasonable prevalence among those affected.

More practically, it was important to ensure that the measures chosen were robust to the sampling and storage processes undertaken by the study, i.e. as noted above the blood was unfasted; sent to the storage laboratory by post and therefore at room temperature for a number of days before processing; and, frozen for 3-4 years before analyzing.

A long list of possible biomarkers was peer reviewed as part of the ESRC grant application process and considered by the Health and Biomarker Advisory Committee of *Understanding Society*. We also consulted key researchers in this field and those responsible for biomarker data collection and analysis in other major UK longitudinal studies. The final set of biomarkers included in this data release are listed in Table 2, and described in more detail in the glossary below.

Of the 13,107 respondents with at least one biomarker, not all measures are available, mainly due to problems processing specific samples. Table 3 describes valid cases, and reasons for missingness, for each biomarker in the combined GPS and BHPS sample. On average valid results are available for 97.5% of the samples processed. The key reason for missingness (column NBA) was the lack of availability of a particular type of blood (ie serum, citrate plasma or whole blood). Haemolysis of the serum sample caused the next highest level of invalid results; liver function tests were particular susceptible to this problem. For a number of biomarkers the analysers had set low or high detection limits (column 2) under/over which the result could not be measured accurately. Table 4 specifies the specific detection limits for each relevant analyte. For testosterone this was true for most women, where the lower detection limit was 1 nmol/l (n=4,978 females). We note below in the glossary how to accommodate such measurement issues in analyses.

For CMV, as explained in more detail in the glossary, an additional CMV avidity test was performed on those who had a positive or indeterminant value on the CMV IgM test. This was only done on 371 cases.

In the data file, the missingness above is separately identified, as described in Table 5, so researchers can decide how they wish to incorporate and report in their results.

Table 2 Biomarkers available in *Understanding Society*

Biomarker	Applications
Cholesterol (total and HDL)& triglycerides	'Fat in the blood' associated heart disease
Glucose intolerance: Glycated haemoglobin - HbA1c	Undiagnosed or poorly managed diabetes
Inflammatory markers - C-reactive protein (CRP), fibrinogen	Measures of inflammation – due injury or infection – acute or chronic – response to stress
- Cytomegalovirus (CMV) seropositivity	Immunoscenence - wear & tear on immune system, chronic stress, associated diabetes
Anaemia – haemoglobin (Hb), ferritin	Marker for poor nutrition; increases with age, sig. health consequences
Liver function tests (LFTs) - Alkaline phosphatase (ALP), Alanine transaminase (ALT), Aspartate transaminase (AST), Gamma glutamyl transferase (GGT), albumin	Associated alcohol, drugs, obesity, consequence of other diseases
Kidney function – creatinine, urea	Kidney diseases increases with age, associated other diseases
Hormones	Associated with stress processes, building muscles, ageing
Testosterone	Marker aggression
Insulin-like growth factor 1 (IGF-1)	Growth & development - associated diet, diabetes and cancer
Dihydroepiandrosterone sulphate (DHEAs)	Associated CVD, muscle strength, cognition

Table 3 Valid cases for each biomarker and reasons for missingness (Combined GPS and BHPS samples with at least one biomarker n=13,107)

Biomarker	Variable name	Valid	Detection Limit‡	%non-missing*	Missing cases													
					ANAERR	CCOAG	CHOHDU	CLOT	FCOAG	HASTND	INSUFF	NOCLOT	NBA	SDIS	UNAVB	UNAVH	UNAVL	UNUS
Cholesterol & triglycerides																		
Cholesterol	chol	12,895		98.4%				2					132	3		71	4	
HDL cholesterol	hdl	12,876		98.2%			15	2					132	3		79		
Triglycerides	trig	12,898		98.4%				2			1		132	3		71		
Glycated haemoglobin	hba1c	12,162		92.8%	1			1					934	1				8
Inflammatory markers																		
C-reactive protein	hscrp	12,530	381	98.5%				2			2		132	3	3	45	9	
Cytomegalovirus IgG	uscmg	12,896		98.4%				2			4		132	3			70	
Cytomegalovirus IgM	uscmm	12,896		98.4%				2			4		132	3			70	
CMV avidity tests†	cmvavc	371		100 %									0					
Clauss fibrinogen	cfib	12,837	1	97.9 %	1	9				1	3	15	215	4		21		
Markers of anaemia																		
Haemoglobin	hgb	12,156		92.7%	8				6		2		933	2				
Ferritin	rtin	12,894	1	98.4%				2			5		132	3		70		
Liver function																		
Albumin	alb	12,920		98.6%				2			2		132	3		44	4	
Alkaline phosphatase	alkp	12,785	1	97.6%				2					132	3		184		
Alanine transaminase	alt	12,778	1	97.5%				2			1		132	3		183	8	
Aspartate transaminase	ast	12,386		94.5%				2			3		132	3		575	6	
Gamma glutamyl transferase	ggt	12,816	10	97.9%				2					132	3		142	2	
Kidney function																		
Creatinine	ecre	12,918		98.6%				2			2		132	3		46	4	
Urea	ure	12,923		98.6%				2					132	3		44	3	
Hormones																		
Testosterone	testo	7,835	5,054	98.3%				2			4		132	3		79	3	
Insulin-like growth factor 1	igfi	12,831		97.9%				2			45		135	2		92		
Dihydroepiandrosterone sulphate	dheas	12,873	23	98.4%				2			5		132	3		69		

† Test only done on those positive or indeterminant on the CMV IgM test 12,736 not applicable

‡ Below/above the lowest/highest detection limit.

* Includes the valid cases and those cases below/above the low/high detection limit.

Explanation of the indicators used for the different types of missing cases:

ANAERR, Unable to perform analysis due to analyser error; **CCOA**, Coagulation studies sample coagulated; **CHOHDU**, Unable to measure Hdl as trig greater than 10.0 mmol/L; **CLOT**: Specimen clotted; **FCOAG**: Full blood count sample coagulated; **HASTND**, Plasma appears haemoglobin stained; **INSUFF**, Insufficient sample for analysis; **NOCLOT**, No clot detected; **NBA**, no relevant type of blood; **SDIS**: Sample discarded prior to analysis; **UNAVB**, Unavailable due to Bilirubin interference; **UNAVH**, Unavailable due to Haemolysis; **UNAVL**: Unavailable due to interference by Lipaemia; **UNUS**: Unusual chromatography.

Table 4 Missing codes for biomarker variables

Reason for missingess	Missing code in data file
Inapplicable	-8
Tissue sample obtained unable to process sample	-22
Below detection limit	-31
Above linear range for analytical method	-32
No blood serum available	-41
No full blood available	-42
No blood plasma available	-43

Table 5 Lowest and highest detection limits for biomarkers where applicable

Biomarker	Lowest detection limit (number of individuals in parenthesis)	Highest detection limit (number of individuals in parenthesis)
CRP	0.2 mg/l (381)	-
Clauss fibrinogen	0.5 g/l (1)	-
Ferritin	3 ug/l (1)	-
ALP	5 u/l (1)	-
ALT	4 u/l (1)	-
GGT	5 u/l (10)	-
Testosterone	1 nmol/l (5,049)	52 nmol/l (5)
DHEAs	0.1 µmol/l (20)	27 µmol/l (3)

ANALYSIS OF BIOMARKER DATA

In analysing biomarkers a range of factors need to be considered.

Biomarkers may be affected by the time of day the blood is taken. For example, some hormones such as testosterone are much higher in the morning than later in the day.

Biomarkers can also be affected by the way the blood has been collected how long it has been stored or the quality of the blood when analysed. For example, while some biomarkers are robust to being stored at room temperature for a number of days (between collection, posting and processing) others may be sensitive. Although those biomarkers that cannot be analysed on 'old' blood at all are known, others may become less accurate if stored at room temperature for a few days, but there is no systematic information on this for different analytes, although as outlined above we are planning a set of experiments to investigate this. Similarly, bloods frozen for a long time may make some analytes less accurate, although based on the comparisons with other studies outlined below; we do not believe this is a concern with these biomarkers.

Some biomarkers are affected by health conditions not related to those that they are representing. This can make it hard to understand which condition the respondent may have. It may be important to control for such co-morbidities in analyses.

Other biomarkers may be affected by substances that the respondent has recently consumed such as food, drink or medications. For example there has been considerable debate as discussed below in whether triglycerides can be accurately measured if someone has recently eaten. In many cases biomarker levels are influenced by medications, these may be being deliberately prescribed to influence the level of that biomarker – for example statins to reduce cholesterol. However, medications may also affect other biological processes. Derived variables for the specific medication categories listed below which need to be considered in the analysis of these biomarkers have been produced.

In all of these cases the analyst will need to decide how to address factors such as these.

As noted above, some of the analysers employed had set lower and/or upper detection limits for specific biomarkers. Analysts may choose to exclude cases outwith these limits. Alternatively, common practice is to include those cases below the lower detection level with values set half way between the detection limit and zero.

Sensitivity analyses which investigate these factors are advised. For example, it may be advisable to exclude samples that took a significant amount of time to be processed. For those biomarkers affected by time of day, it is important to control for this as a confounder in any models. For medication data, analysts take a range of different approaches. Some adjust the biomarker concerned by estimated effect size from the literature for specific medications, this is an approach often employed for blood pressure. Alternatively analysts may control for medication use in their models or conduct sensitivity analyses excluding those taking the medications of concern for a particular biomarker.

GLOSSARY

In this glossary we provide key information for each biomarker to help users with their analysis and to provide relevant information to be considered when these data are used for publications. For each biomarker we outline:

- the clinical significance of each biomarker,
- a description of the role of the biomarker in the body;
- laboratory methods and procedures used to measure the analyte;
- guidance on factors to consider when analysing the biomarker;
- distributions for each biomarker and where available[12], these distributions are compared to equivalent data available from the HSE or ELSA.

As outlined above, NUTH undertake both internal and external quality assurance programmes and have provided the data from these to us. Below for each analyte we provide quality control information given by co-efficient of variation data. This shows the extent of variation within an assay (intra) and between assays (inter). Values less than 5% are considered good quality. Appendix 1 provides definitions and exact details of internal and monthly external quality control values that are used in each biomarker description.

The distribution of each biomarker in *Understanding Society* is illustrated with a Kernel density estimate, a non-parametric estimator that smooths observed data over local neighbourhood points. This shows how each biomarker is distributed, so that analysts can consider if and how to transform variables for analysis. Additionally, where data from HSE or ELSA are available, the age-sex distributions of the biomarkers are compared to the English sample of *Understanding Society* only. This comparison is to illustrate how well the mean of the biomarkers in the different studies, analysed by the same laboratory, match; where they are differences we have tried to identify the reasons for this.

All data presented are weighted by sample weights and do not exclude cases because of co-morbidities or medications, but do exclude top and bottom 0.5% of outliers.

[12] Not available for triglycerides, LFTs, Urea, CMV.

CHOLESTEROL AND TRIGLYCERIDES

Total cholesterol and HDL-cholesterol

What is the clinical significance of total cholesterol and HDL-cholesterol?

Total cholesterol is a risk factor for cardiovascular disease (CVD), while HDL-cholesterol is thought to be protective against it.

What are total cholesterol and HDL-cholesterol?

Cholesterol is a steroid that is a vital component of the lining of cells. Because it is not soluble in blood, it is transported around the body as cargo in cells known as lipoprotein particles. There are two kinds of lipoproteins – Apolipoprotein A and Apolipoprotein B. The first of these (Apolipoprotein A) contains high density lipoproteins (HDL) which are involved in the delivery of cholesterol to the liver for breakdown, and are hence beneficial for the body. Apolipoprotein B carries low density lipoproteins (LDL) – 'bad cholesterol' which are taken up by blood vessels to cause narrowing of arteries.

How is it measured?

Total cholesterol and HDL-cholesterol were measured from blood serum using enzymatic methods with a Roche Modular P analyser calibrated to the Centre for Disease Control guidelines.

Intra and inter assay coefficients of variation (%CV) were less than 2%.

Are there clinical cutpoints?

Total cholesterol should be 5mmol/L or less for healthy adults

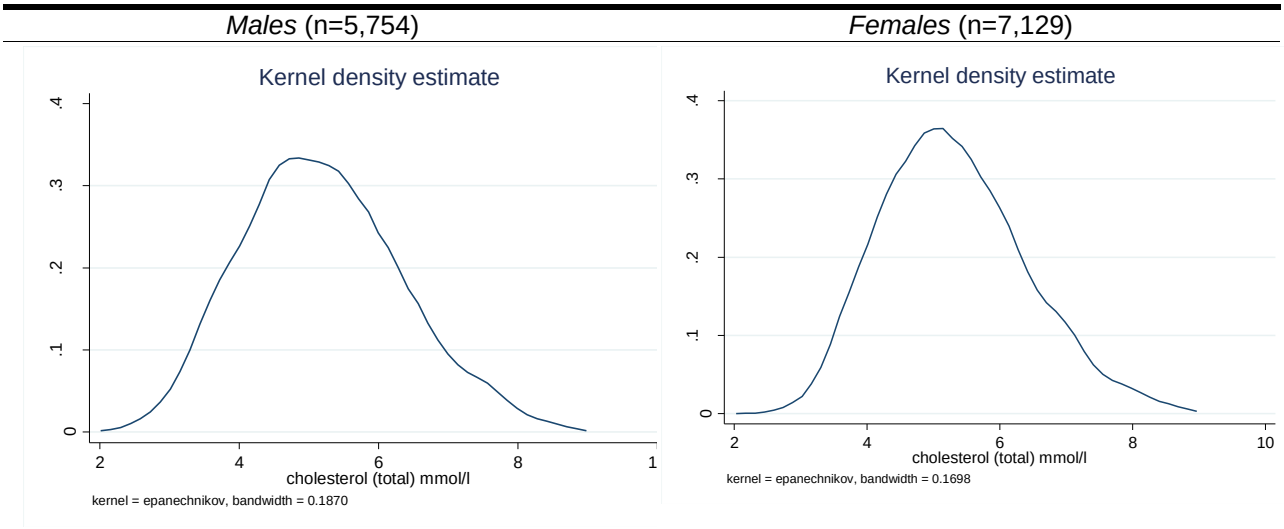
HDL-cholesterol should be above 1mmol/L

What should be considered in analyses?

Cholesterol is treated with a number of lipid regulating drugs eg statins (BNF: chapter 2.12). A derived variables has been created to indicate whether or not respondents reported taking one of these prescribed medications in the previous 7 days (variable = bnf7_statins).

Distribution in *Understanding Society*

Figure 1 Distribution of Total Cholesterol levels [mmol/l] by gender



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

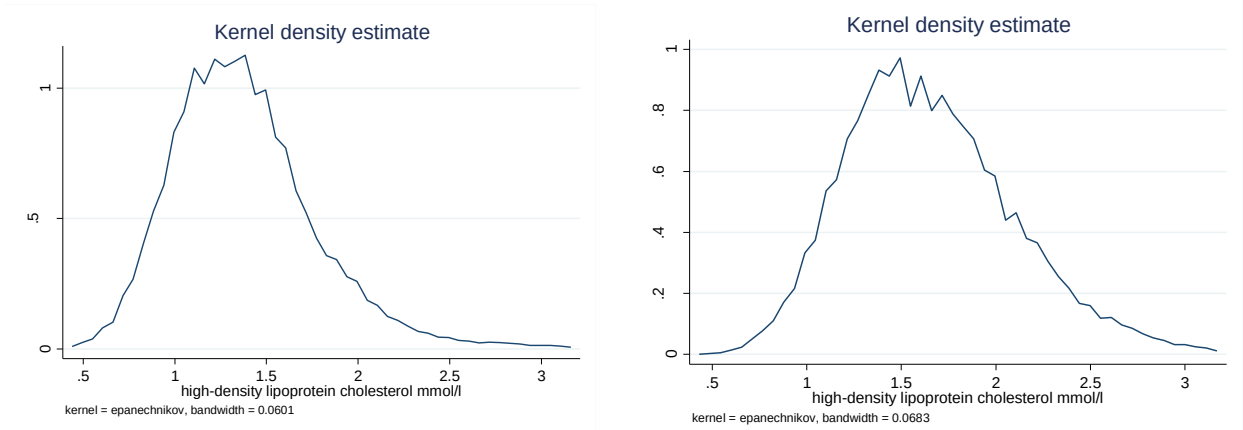
Table 6 Total Cholesterol levels [mmol/l] by gender and age

Males								
Health Survey for England 2011 (n=1,733)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Total Cholesterol	4.5 (0.10)	5.0 (0.06)	5.5 (0.06)	5.6 (0.07)	5.3 (0.06)	4.9 (0.07)	4.6 (0.09)	5.1 (0.03)
Understanding Society (England sub-sample; n= 4,848)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Total Cholesterol	4.3 (0.07)	5.1 (0.06)	5.6 (0.04)	5.7 (0.04)	5.4 (0.04)	5.0 (0.04)	4.7 (0.05)	5.2 (0.02)
Females								
Health Survey for England 2011 (n=2,184)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Total Cholesterol	4.4 (0.07)	4.8 (0.05)	5.1 (0.05)	5.4 (0.05)	5.8 (0.05)	5.7 (0.07)	5.4 (0.08)	5.2 (0.03)
Understanding Society (England sub-sample; n= 5,965)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Total Cholesterol	4.6 (0.06)	4.8 (0.04)	5.1 (0.04)	5.6 (0.03)	5.9 (0.04)	5.7 (0.04)	5.3 (0.06)	5.3 (0.02)

Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded. Sample weights are employed for each dataset. Standard errors of the mean in parentheses.

Figure 2 Distribution of HDL Cholesterol levels [mmol/l] by gender

Males (n=5,759)	*Females (n=7,119)*



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

Table 7 HDL Cholesterol levels [mmol/l] by gender and age

Males								
Health Survey for England 2011 (n=1,735)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean HDL	1.3 (0.03)	1.4 (0.02)	1.3 (0.02)	1.4 (0.02)	1.3 (0.02)	1.4 (0.02)	1.4 (0.04)	1.3 (0.01)
Understanding Society (England sub-sample; n=4,850)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean HDL	1.4 (0.03)	1.4 (0.02)	1.4 (0.02)	1.4 (0.02)	1.4 (0.02)	1.4 (0.02)	1.4 (0.02)	1.4 (0.01)
Females								
Health Survey for England 2011 (n=2,185)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean HDL	1.5 (0.03)	1.6 (0.02)	1.6 (0.02)	1.6 (0.02)	1.7 (0.02)	1.6 (0.02)	1.7 (0.03)	1.6 (0.01)
Understanding Society (England sub-sample; n=5,954)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean HDL	1.6 (0.03)	1.6 (0.02)	1.6 (0.02)	1.7 (0.02)	1.7 (0.02)	1.7 (0.02)	1.7 (0.02)	1.7 (0.01)

Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded. Sample weights are employed for each dataset. Standard errors of the mean in parentheses.

Tables 6 and 7 show that mean total and HDL-cholesterol values from the England component of *Understanding Society* are very close to those obtained from the HSE (2011) by gender and across age groups.

Triglycerides

What is the clinical significance of Triglycerides?

Triglycerides levels are predictive of CVD.[13]

What is it?

Triglycerides are fats that are transported in the blood. They come from dietary sources or from the liver. They may be taken up by cells and used for energy or stored as fat. High levels of triglycerides are often found with low levels of HDL cholesterol.

How is it measured?

There has been a debate about whether triglyceride levels should be assessed from participants who have or have not fasted. In *Understanding Society* blood samples were collected from participants that were not requested to fast. Levels of triglycerides are influenced by recent food intake, however evidence suggests that these changes are small and do not obscure associations with CVD.

Triglycerides are measured from serum blood using an enzymatic method, on a Roche P module analyser.

Inter and intra coefficients of variation were less than 3%.

Are there clinical cutpoints?

The desirable non-fasting triglyceride level is <2mmol/l.[14]

What should be considered in analyses?

Triglyceride levels are influenced by statins (BNF chapter 2.12; derived variable bnf7_statins).

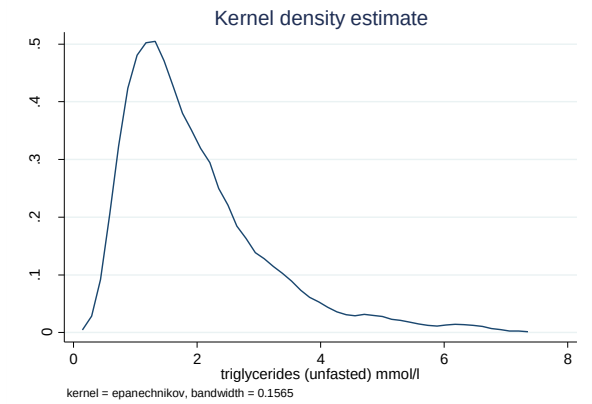
[13] *Nordestdgaard BG, Varbo A. Triglycerides and cardiovascular disease. Lancet 2014;384: 626–635*

[14] Kolovou, GD; Mikhailidis, DP; Kovar, J; Lairon, D; Nordestgaard, BG; Chye Ooi, T; Perez-Martinez, P; Bilianou, H; Anagnostopoulou, K; Panotopoulos, G. Assessment and Clinical Relevance of Non-Fasting and Postprandial Triglycerides: An Expert Panel Statement. Current Vascular Pharmacology 2011: 9(3); *258-270(13)*

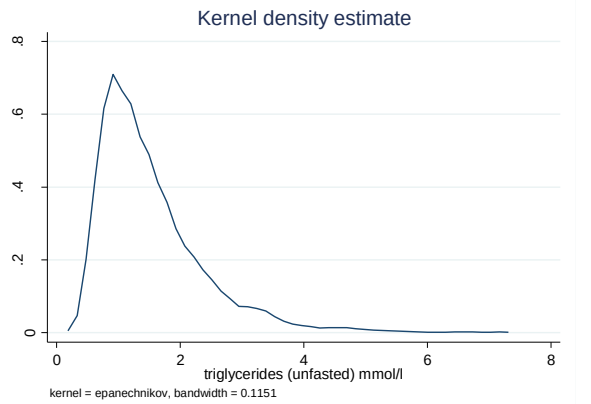
Distribution in *Understanding Society*

Figure 3 Distribution of Triglycerides [mmol/L] by gender

Males (n=5,737)



Females (n=7,155)



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

GLYCATED HAEMOGLOBIN (HbA1c)

What is its clinical significance?

HbA1c has recently been highlighted as a 'gold standard' indicator of diabetes risk[15]. It can be used both to identify those who might be suffering from diabetes as well as highlight those people who may not be managing their diabetes consistently.

What is it?

Glycated haemoglobin (HbA1c) is a measure of the level of sugar in the blood over the previous 8 to 12 weeks before measurement. Technically it is the proportion of haemoglobin proteins that have been bound by glucose.[16]

How is it measured?

Glycated haemoglobin is measured from whole blood using HPLC cation exchange on a Tosoh G8 analyser.

HbA1c can be expressed as a percentage or as a value in mmol/mol. Since 2009, mmol/mol has been the default unit to use in the UK.

Intra and inter assay coefficients of variation were less than 4%.

Are there clinical cutpoints?

HbA1c values > 48 mmol/mol (≥ 6.5%) indicates diagnosis of diabetes[17,18]. Values between 5.7% and 6.4% indicate pre-diabetes risk

What should be considered in analyses?

A number of factors are associated with decreased HbA1c measurements[19] - chronic liver disease, taking aspirin (BNF chapter 2.9; derived variable bnf7_aspirin) and anti-inflammatory medications (BNF chapter 10.1; derived variable: bnf7_antiinflam). High levels of triglycerides may lead to artefactually low measurements.

[15] *International Expert Committee (2009). International Expert Committee report on the role of the A1C assay in the diagnosis of diabetes. Diabetes Care 32, 1327–1334.*

[16] *Nathan DM, Kuenen J, Borg R, Zheng H, Shoenfeld D, Heine RJ. Translating the A1C assay into estimated average glucose values. Diabetes Care 2008; 31: 1473-1478*

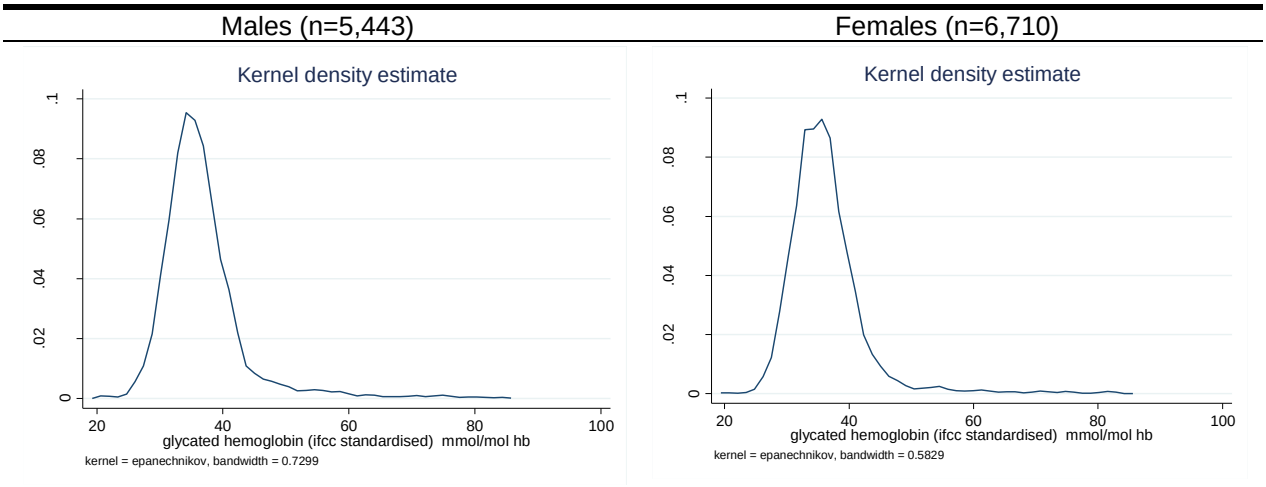
[17] *American Diabetes Association. Executive summary: standards of medical car in diabetes – 2010. Diabetes Care 33 (Suppl 1): S4-S10*

[18] *World Health Organisation (2011). Use of glycated haemoglobin (HbA1c) in the Diagnosis of Diabetes Mellitus. Geneva: World Health Organisation*

[19] *Gallagher EJ, Le Roith, D Bloomgarden Z. Review of haemoglobin A1c in the management of diabetes. J. Diabetes; 2009: 1: 9-17*

Distribution in *Understanding Society*

Figure 4 Distribution of Glycated Haemoglobin (HbA1c) [mmol/mol] by gender



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

Table 8 Glycated Haemoglobin(HbA1c) [mmol/mol] by gender and age

Males								
Health Survey for England 2011 (n=1,713)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean HbA1c	34.9 (0.34)	35.9 (0.40)	37.5 (0.33)	39.6 (0.50)	41.7 (0.53)	41.7 (0.61)	43.4 (0.67)	38.7 (0.20)
Understanding Society (England sub-sample; n=4,583)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean HbA1c	32.7 (0.26)	33.6 (0.25)	35.2 (0.20)	36.9 (0.23)	39.3 (0.35)	41.0 (0.48)	40.8 (0.43)	36.8 (0.14)
Females								
Health Survey for England 2011 (n=2,173)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean HbA1c	34.4 (0.36)	34.8 (0.29)	36.6 (0.40)	38.8 (0.40)	41.0 (0.48)	42.2 (0.58)	42.9 (0.46)	38.3 (0.18)
Understanding Society (England sub-sample; n=5,604)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean HbA1c	33.2 (0.32)	33.4 (0.23)	34.6 (0.31)	36.4 (0.24)	38.4 (0.22)	39.6 (0.25)	40.9 (0.46)	36.3 (0.11)

Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded. Sample weights are employed for each dataset. Standard errors of the mean in parentheses.

Table 8 shows that mean HbA1c values from the England component of *Understanding Society* are close to those obtained from the HSE (2011) by gender and across age groups.

INFLAMMATORY MARKERS (MARKERS OF INFLAMMATION/IMMUNE FUNCTION)

C-reactive protein (CRP)

What is its clinical significance?

CRP is a marker of inflammatory load; high values are associated with adverse CVD and mortality[20].

Systemic inflammation is defined as CRP > 3 mg/L levels. This dichotomization was selected based on the clinical guidelines of the joint scientific statement from the Centers for Disease Control and Prevention and American Heart Association that CRP levels above 3 mg/L be used to indicate high risk of cardiovascular diseases.[21]

What is it?

CRP is an acute phase protein in the blood that rises in response to inflammation. It is part of the body's defence mechanism against harmful stimulus.

How is it measured?

CRP was analyzed from serum using the N Latex CRP mono Immunoassay on the Behring Nephelometer II Analyzer (Dade Behring, Milton Keynes, UK).

Intra and inter assay coefficients of variation were less than 2%.

Are there clinical cutpoints?

Values of >3 mg/L are considered a risk factor for CVD[19]

What should be considered in analyses?

Values >10mg/L are considered to reflect recent infection. It is recommended that these data should be removed prior to analyses.

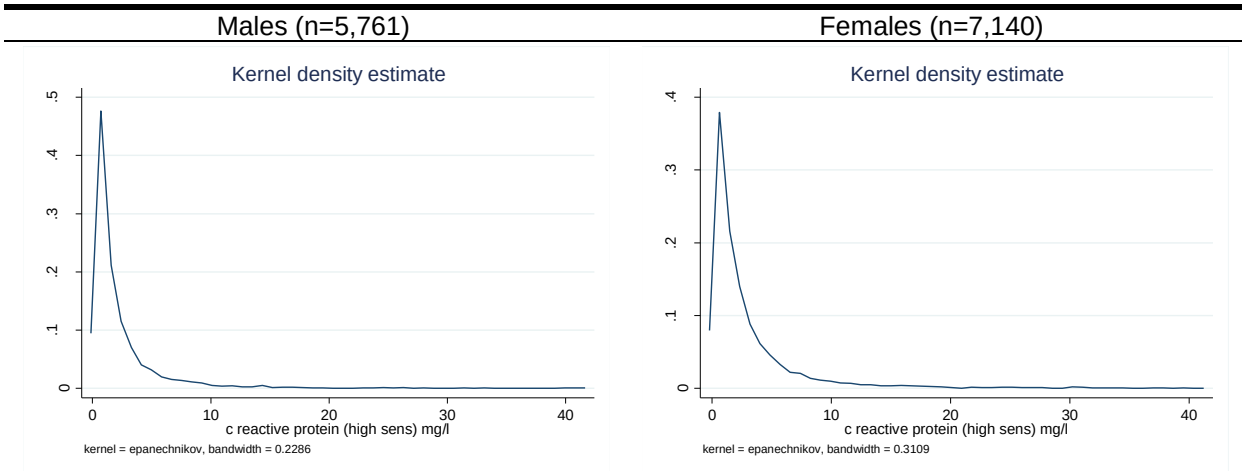
CRP is influenced by medication: anti-inflammatory medications (BNF chapter 10.1; derived variable bnf7_antiinflam), statins (BNF chapter 2.12; derived variable bnf7_statins) and contraception and hormone replacement therapy (BNF chapter 6.4.1; derived variable bnf7_conhrt).

[20] *Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the biology of atherosclerosis Nature, 473 (2011), pp. 317–325*

[21] *Pearson TA, Mensah GA, Alexander RW, Anderson JL, Cannon 3rd RO, Criqui M, et al. Markers of inflammation and cardiovascular disease: application to clinical and public health practice: A statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association Circulation, 107 (2003), pp. 499–511*

Distribution in *Understanding Society*

Figure 5 Distribution of C-reactive protein [mg/l] by gender



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

Table 9 C-reactive protein (CRP) [mg/l] by gender and age

Males								
Health Survey for England 2009 (n=1,525)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean CRP	1.4 (0.28)	1.9 (0.30)	2.0 (0.21)	2.8 (0.31)	3.1 (0.35)	3.7 (0.40)	3.8 (0.40)	2.5 (0.12)
Understanding Society (England sub-sample; n=4,852)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean CRP	1.2 (0.11)	1.8 (0.21)	2.3 (0.13)	2.3 (0.13)	2.8 (0.20)	3.0 (0.21)	3.5 (0.26)	2.3 (0.07)
Females								
Health Survey for England 2009 (n=1,271)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean CRP	2.8 (0.50)	2.8 (0.31)	2.7 (0.24)	3.0 (0.25)	3.8 (0.34)	3.3 (0.28)	3.9 (0.44)	3.1 (0.13)
Understanding Society (England sub-sample; n=5,971)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean CRP	2.5 (0.30)	2.8 (0.22)	2.7 (0.20)	2.8 (0.13)	3.6 (0.21)	3.6 (0.20)	3.6 (0.28)	3.1 (0.09)

Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded. Sample weights are employed for each dataset. Standard errors of the mean in parentheses.

Table 9 shows that mean CRP values from the England component of *Understanding Society* are close to those obtained from the HSE (2009) by gender and across age groups.

Fibrinogen

What is its clinical significance?

Fibrinogen is a marker of inflammation and it helps the body to stop bleeding by helping blood clots to form. Higher levels of fibrinogen are implicated in the development of CVD.[22]

What is it?

Fibrinogen is a glycoprotein. Through a series of enzymatic steps is converted into fibrin in the clotting process. Fibrinogen is also an 'acute phase protein' and therefore reflects inflammatory processes.

How is it measured?

Fibrinogen was analyzed from citrate plasma samples using a modification of the Clauss thrombin clotting method on the IL-ACS-TOPS analyser.

Intra and inter-assay coefficients of variation were less than 7%.

Are there clinical cutpoints?

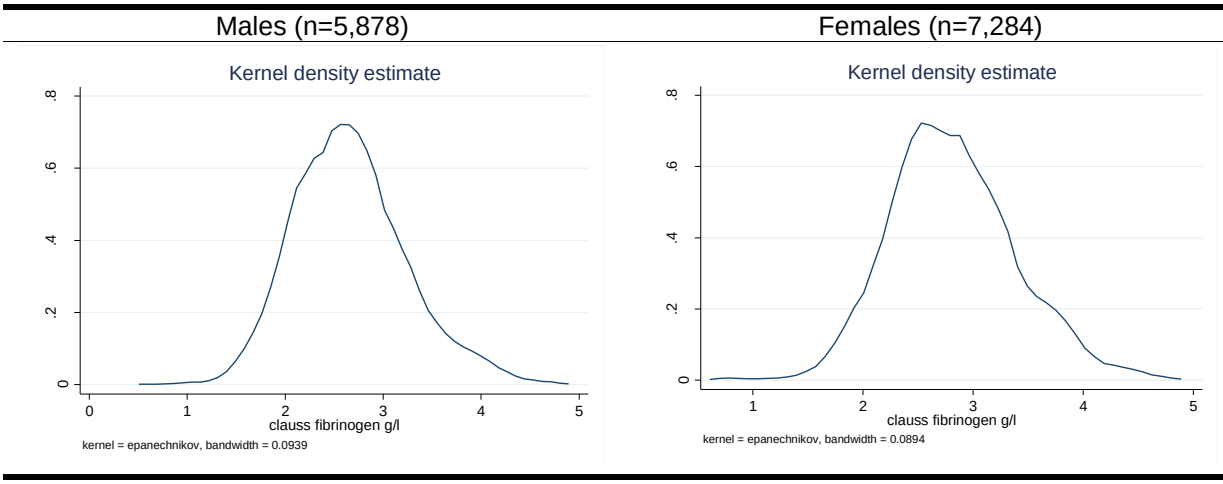
Data are continuous and there are no established clinical cutpoints.

What should be considered in analyses?

Fibrinogen can be influenced by contraception and hormone replacement therapy (BNF chapter 6.4.1; derived variable bnf7_conhrt) and antifibrinolytic drugs and haemostatics medications (BNF chapter 2.11; derived variable bnf7_antfibs).

Distribution in *Understanding Society*

Figure 6 Distribution of Fibrinogen [g/L] by gender



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

[22] *Danesh J, Lewington S, Thompson SG, Lowe GDO, Collins R. Plasma fibrinogen level and the risk of major cardiovascular diseases and nonvascular mortality. J. Am. Med. Assoc., 294 (2005), pp. 1799–1809*

Table 10 Fibrinogen [g/L], by gender and age

Males								
Health Survey for England 2009 (n= 1,033)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Fibrinogen	2.5 (0.06)	2.7 (0.05)	2.9 (0.05)	3.1 (0.04)	3.2 (0.04)	3.4 (0.05)	3.5 (0.07)	3.0 (0.03)
Understanding Society (England sub-sample; n=4,829)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Fibrinogen	2.3 (0.04)	2.4 (0.03)	2.6 (0.02)	2.7 (0.02)	2.8 (0.02)	2.9 (0.02)	3.0 (0.03)	2.7 (0.02)
Females								
Health Survey for England 2009 (n=1,176)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Fibrinogen	3.1 (0.06)	3.0 (0.05)	3.0 (0.04)	3.2 (0.04)	3.3 (0.04)	3.4 (0.05)	3.5 (0.06)	3.2 (0.02)
Understanding Society (England sub-sample; n=5,934)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Fibrinogen	2.6 (0.04)	2.7 (0.03)	2.7 (0.02)	2.8 (0.02)	2.9 (0.02)	3.0 (0.02)	3.1 (0.03)	2.8 (0.01)

Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded. Sample weights are employed for each dataset. Standard errors of the mean in parentheses.

Table 10 shows that mean Fibrinogen values from the England component of *Understanding Society* are close to those obtained from the HSE (2009) by gender and across age groups.

Cytomegalovirus antibody measurement (CMV)

What is its clinical significance?

Measurement of Cytomegalovirus (CMV) antibodies provides information about the people's immune function, which declines with illness and age.

What is it?

Cytomegalovirus (CMV) is a herpes virus that is often asymptomatic. This type of virus is unusual as it is lays dormant in the body[23]; approximately half of the general population have ever been infected with it (Table 12 below). With ageing or weakening of the immune system the virus can develop and this causes the release of antibodies, which attempt to protect the body from the virus. These antibodies are used as biomarkers for the virus.

How is it measured?

There are two antibodies that are important. The presence of immunoglobulin G (IgG) shows that someone has at some time had a CMV infection; while Immunoglobulin M (IgM) indicates recent infection.

In *Understanding Society*, we have measured IgG and IgM from serum samples with an electrochemiluminsecent immunoassay on the Roche E170 analyser.

A positive CMV IgG result indicates a CMV infection at some point in time, while a negative CMV IgG indicates that the participant has never been exposed or been infected with CMV. A positive Immunoglobulin M (IgM) indicates a recent or current infection. Indeterminate CMV occurs during current or acute infection or may be due to non-specific binding.

For those people who had a positive IgM test or whose result was indeterminate, an additional test was performed to confirm recent CMV infection. This confirmatory assay was an avidity test on the Mini VIDAS immunoassay analyser.

Inter and intra assay coefficients of variation were less than 4%.

Are there clinical cutpoints?

No quantitative assessments of viral load were made as these methods are semi-quantitative and indirect.

Data for each antibody are presented as virus detected, not detected or indeterminate. Data for the avidity test is presented as high, low and indeterminate. In combination these three variables (see Table 11 below) demonstrate the presence of the virus in the body and how recently the person experienced an infection.

[23] *Sinclair J. Human cytomegalovirus: Latency and reactivation in the myeloid lineage. J Clin Virol. 2008;41(3):180-5*

What should be considered in analyses?

To our knowledge no other factors require consideration in the analysis.

Descriptive statistics in *Understanding Society*

Table 11 CMV infection and if it recently occurred (interpretation of combined variables)

CMV IgM	CMV IgG	CMV Avidity	INTERPRETATION	N`	%
Negative	Negative		No evidence of past or current CMV infection	6428	49.95
Negative	DETECTED		Past CMV infection. (No evidence of recent primary CMV infection)	5998	46.61
DETECTED/indeterminate	DETECTED	HIGH	Past CMV infection. (No evidence of recent primary CMV infection)	366	2.84
DETECTED/indeterminate	DETECTED	LOW	Consistent with recent primary CMV infection	4	0.03
DETECTED/indeterminate	DETECTED	Indeterminate	Evidence of CMV infection at some time. (Cannot confirm or exclude recent primary CMV infection)	3	0.02
DETECTED/indeterminate	DETECTED		Evidence of CMV infection at some time (Further sample needed to confirm, samples not available)	43	0.33
DETECTED/indeterminate	Negative		Possible very recent primary CMV infection or non-specific result. (Further sample needed to confirm, samples not available)	27	0.21

MARKERS OF ANAEMIA

Haemoglobin (Hb)

What is the clinical significance of haemoglobin?

Low levels of Hb is suggestive of anaemia, a lack of iron in the blood, which is prevalent in the elderly,[24] and associated with longer hospitalization and greater risk of mortality and CVD.[25]

What is it?

Hb is the iron-containing molecule responsible for carrying oxygen from the respiratory organs to the rest of the body, and low levels are usually indicative of anaemia.

How is it measured?

Hb levels were measured from whole blood samples with a spectrophotometric assay on a Sysmex XE-2100 analyser.

Inter and intra assay coefficients of variation were less than 1%.

Are there clinical cutpoints?

Anaemia status is defined (based on WHO guidelines) as Hb levels <13 g/dL for men and <12 g/dL for women[26].

What should be considered in analyses?

Hb levels are influenced by a number of factors, such as pregnancy and high altitude but these are not applicable in our population.

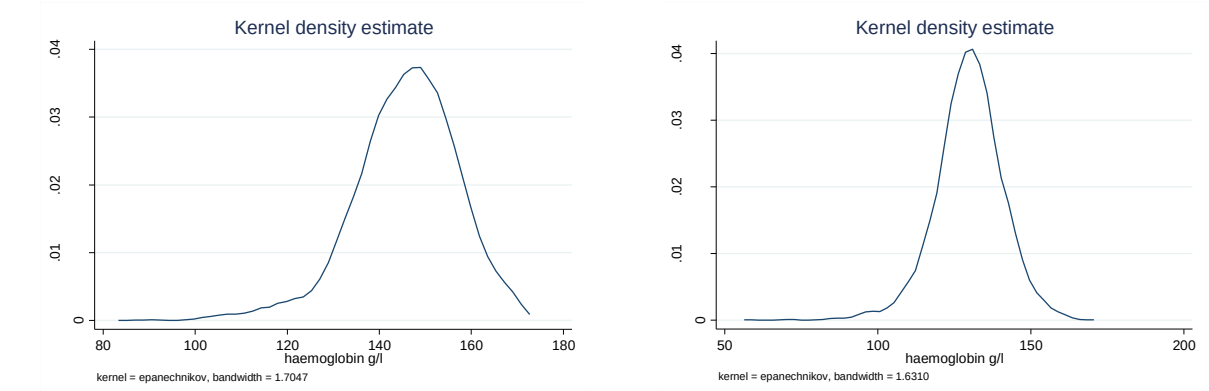
[24] *Nilsson-Ehle H, Jagenburg R, Landahl S, Svanborg A Blood haemoglobin declines in the elderly: implications for reference intervals from age 70 to 88 Eur. J. Haematol., 65 (2000), pp. 297–305*

[25] *Culleton BF, Manns BJ, Zhang JG, Tonelli M, KlarenbachS, Hemmelgarn BR. Impact of anemia on hospitalization and mortality in older adults Blood, 107 (2006), pp. 3841–3846*

[26] *WHO (World Health Organization) Nutritional Anaemias: Report of a WHO Scientific Group. World Health Organization, Geneva (1968)*

Distribution in *Understanding Society*

Figure 7 Distribution of Haemoglobin level [g/l] by gender



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

Table 12 Haemoglobin (Hb)level [g/l] by gender and age

				Males				
Health Survey for England 2009 (n=1,203)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Hb	150.9 (0.92)	151.1 (0.90)	150.5 (0.64)	149.4 (0.67)	149.2 (0.75)	147.2 (0.87)	140.4 (1.65)	149.2 (0.34)
Understanding Society (England sub-sample; n=4,570)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Hb	148.5 (0.71)	149.0 (0.57)	147.5 (0.49)	146.8 (0.41)	145.1 (0.47)	142.8 (0.56)	135.5 (0.88)	145.7 (0.22)
				Females				
Health Survey for England 2009 (n=1,044)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Hb	132.5 (1.38)	130.4 (1.18)	131.3 (0.65)	133.2 (0.86)	134.9 (0.58)	135.1 (0.89)	131.6 (1.05)	132.7 (0.36)
Understanding Society (England sub-sample; n=5,625)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Hb	129.2 (0.91)	129.1 (0.60)	128.6 (0.45)	129.2 (0.45)	132.0 (0.40)	131.2 (0.46)	126.4 (0.70)	129.8 (0.22)

Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded. Sample weights are employed for each dataset. Standard errors of the mean in parentheses.

Table 12 shows that mean Hb values from the England component of *Understanding Society* are close to those obtained from the HSE (2009) by gender and across age groups.

Ferritin

What is it?

Levels of ferritin reflect the size of the body iron stores and therefore it is indicative of anaemia. A low ferritin level is predictive of uncomplicated iron deficiency anaemia. However, high ferritin levels suggest excess body iron, which is also problematic for health.

How is it measured?

Ferritin is measured from serum samples by an electrochemiluminescent immunoassay on the Roche Modular E170 analyser.

Inter and intra assay coefficients of variation were less than 3%.

Are there clinical cutpoints?

Both high and low levels of ferritin are associated with adverse outcomes the following cut points are suggested[27]

[27] *World Health Organisation. Serum ferritin concentrations for the assessment of iron status and iron deficiency in populations. Vitamin and Mineral Nutrition Information System. Geneva, World Health Organization, 2011*

Ferritin levels below ≤ 20 ug/L indicate depletion of iron, while levels ≤ 12 indicate complete absence of stored iron

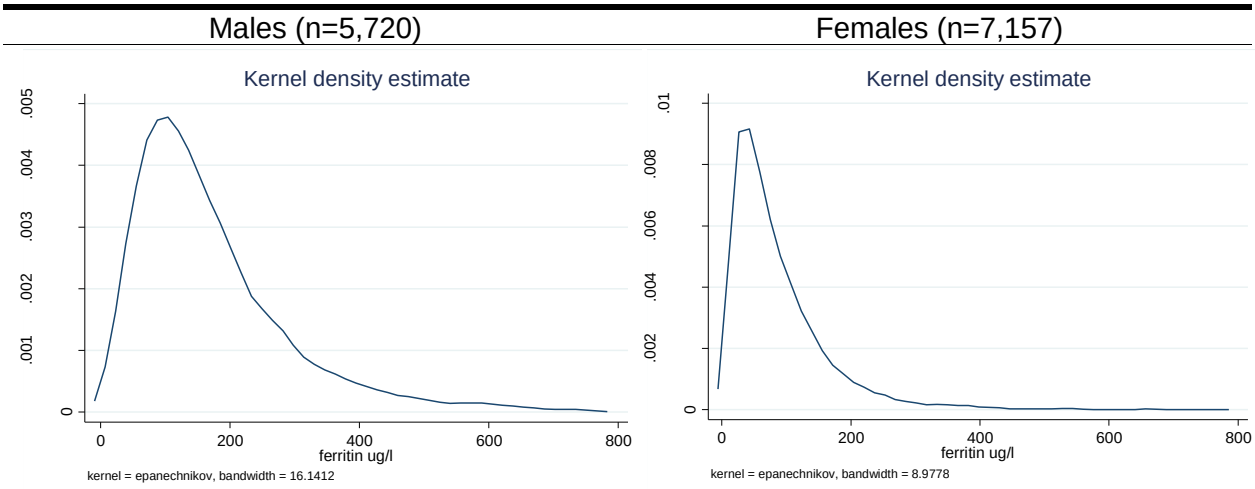
Ferritin levels >300 ug/L may indicate iron overload in men and postmenopausal women and >200 may indicate iron overload in premenopausal women.

What should be considered in analyses?

Ferritin levels are influenced by taking aspirins and anti-inflammatory medication[28] (BNF chapter 2.9, derived variable bnf7_aspirin; BNF chapter 10.1, derived variable bnf7_antiinflam).

Distribution in *Understanding Society*

Figure 8 Distribution of Ferritin [ng/ml or ug/l] by gender



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

[28] *Fleming et al., Aspirin intake and the use of serum ferritin as a measure of iron status Am J Clin Nutr 2001;74:2 219-226*

Table 13 Ferritin [ng/ml or ug/l] by gender and age

Males								
Health Survey for England 2009 (n= 1,067)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Ferritin	88.3 (7.0)	147.3 (9.9)	155.0 (8.4)	149.6 (7.1)	155.8 (8.2)	162.0 (9.1)	130.0 (13.4)	141.5 (3.7)
Understanding Society (England sub-sample; n=4,816)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Ferritin	108.6 (5.55)	166.2 (6.96)	184.1 (6.45)	190.8 (5.14)	194.8 (5.40)	177.9 (6.18)	168.0 (6.68)	172.02 (2.47)
Females								
Health Survey for England 2009 (n= 1,217)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Ferritin	36.3 (3.3)	46.8 (3.7)	55.3 (4.5)	65.8 (3.7)	93.0 (4.2)	111.0 (7.3)	94.3 (7.5)	70.0 (1.9)
Understanding Society (England sub-sample; n= 5,985)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Ferritin	50.4 (2.82)	60.8 (2.20)	67.3 (2.24)	81.4 (2.26)	114.9 (3.04)	124.3 (3.84)	110.0 (4.81)	85.2 (1.21)

Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded. Sample weights are employed for each dataset. Standard errors of the mean in parentheses.

Table 13 shows that mean ferritin values from the England component of *Understanding Society* differ to those obtained from the HSE (2009). However, the analyser employed by NUTH was changed between the HSE analyses and those for *Understanding Society*, which may explain these differences.

LIVER FUNCTION TESTS (LFTS)

What is the clinical significance of liver function tests?

A panel of liver function tests were measured in *Understanding Society,* which can reflect how well the liver is functioning.[29]

What are these tests?

- Alanine Transaminase (ALT) – an enzyme mainly found in the liver; the best test for detecting hepatitis, raised levels indicate liver damage
- Aspartate Transaminase (AST) – an enzyme found in the liver and a few other places, particularly the heart and other muscles in the body, raised levels indicate liver damage
- Alkaline Phosphatase (ALP) – an enzyme related to the bile ducts; often increased when they are blocked, either inside or outside the liver
- Gamma Glutamyl Transferase (GGT)- an enzyme raised levels of which help to detect liver disease and bile duct injury.
- Albumin – measures the main protein made by the liver and tells how well the liver is making this protein, low levels may be indicative of a loss of liver function

How are they measured?

The liver function tests are conducted with serum samples.

ALT is measured with the International Federation of clinical chemistry (IFCC) UV with Pyridoxal phosphate activation method on the Roche P module analyser.

AST is measured with the IFCC UV with Pyridoxal phosphase activation method on the Roche P module analyser.

ALP is measured with the IFCC colourimetric PNP method on the Roche P module analyser.

GGT is measured with an enzymatic method on the Roche P module analyser.

Albumin is measured with a BCG colourimetric method on the Roche P module analyser.

Inter and intra assay coefficients of variation were less than 5% for all of these assays.

[29] *Limdi J, Hyde G. Evaluation of abnormal liver function tests.* Postgrad Med J. Jun 2003; 79(932): 307–312

Are there clinical cutpoints?

Table 14: Clinical cutpoints for the liver function tests

Test	Units of measurement	Age range	Poor liver function level
ALT	U/L		>40
AST	U/L		>40
ALP	U/L	20-70 yrs	<30, >130
		>70 yrs	<30, >150
GGT	U/L		M>70, F>45
Albumin	g/L		<35, >50

What should be considered in analyses?

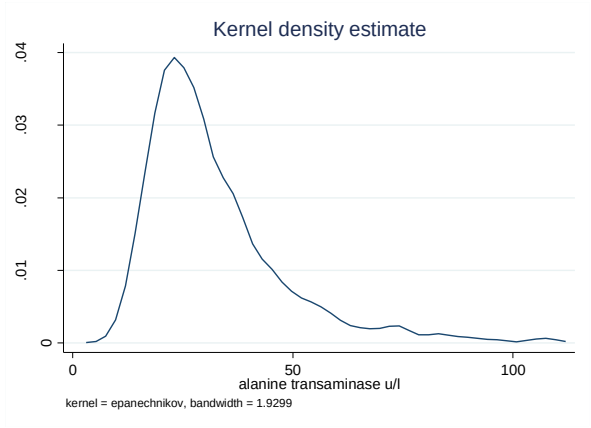
Recent alcohol intake influences the measures of these analytes. It is recommended that this is taken into account in analyses (using nurse visit question for recent consumption). Some medications may be associated with raised LFTs for example anti-epilepsy medications (BNF chapter 4.8; derived variable bnf7_antiep) or statins (BNF chapter 2.12; derived variable bnf7_statins).

Distribution in *Understanding Society*

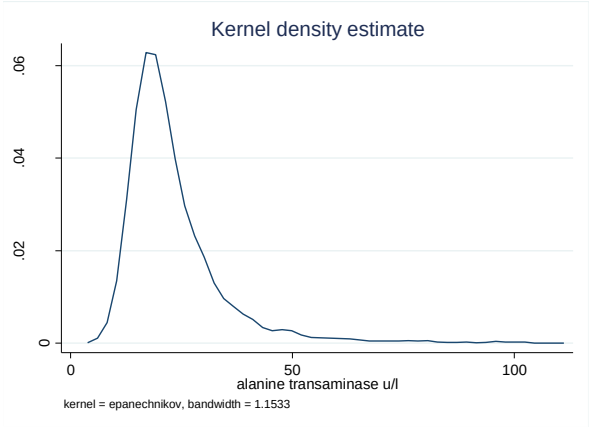
Figure 9 Distribution of LFTs by gender

Distribution of Alanine Transaminase (ALT) [IU/L] by gender

Males (n= 5,682)

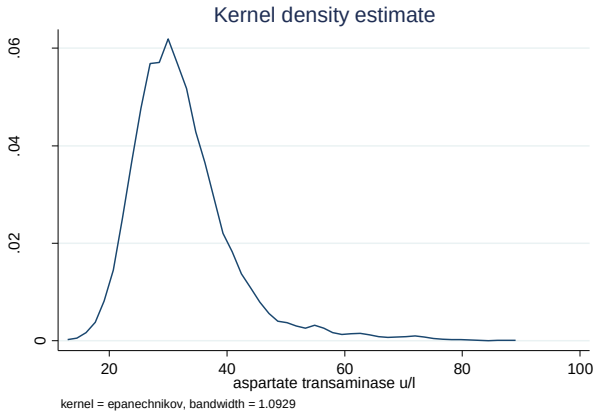


Females (n= 7,089)

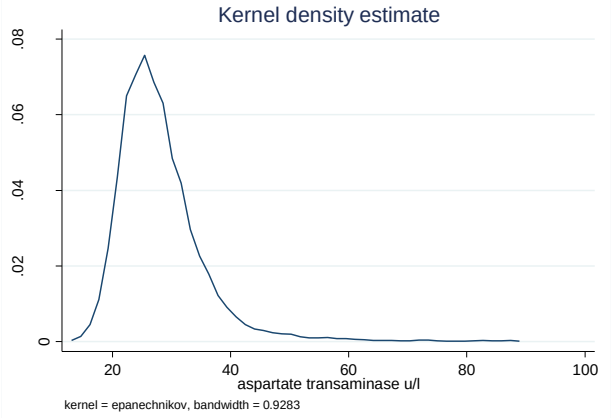


Distribution of Aspartate Transaminase (AST)[IU/L] by gender

Males (n=5,878)

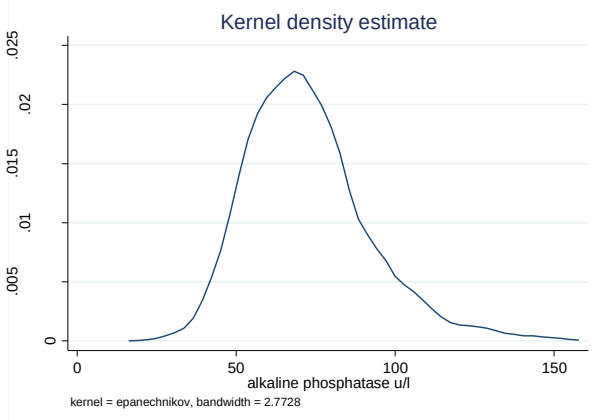


Females (n=7,284)

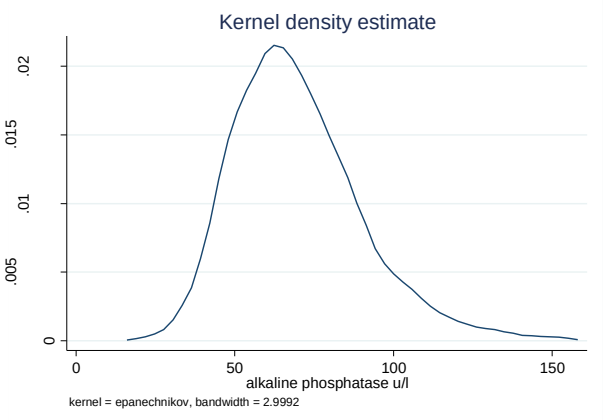


Distribution of Alkaline Phosphatase (ALP) [IU/L] by gender

Males (n=5,696)

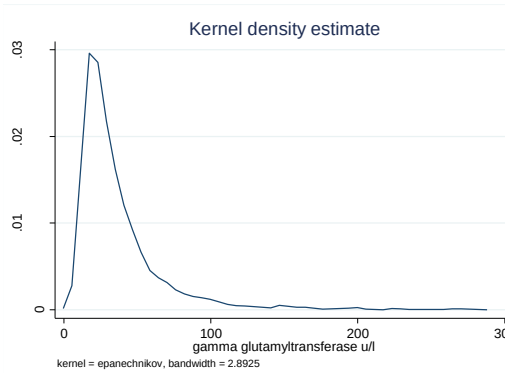


Females (n=7,078)

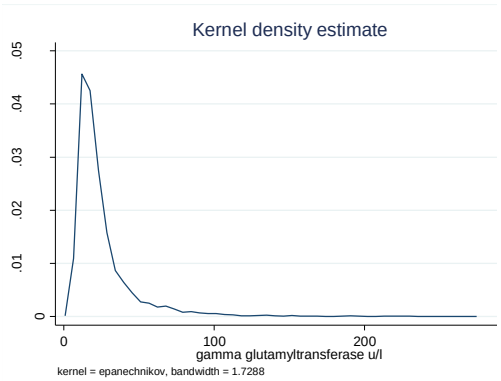


Distribution of Gamma Glutamyl Transferase (GGT) [IU/L] by gender

Males (n=5,710)

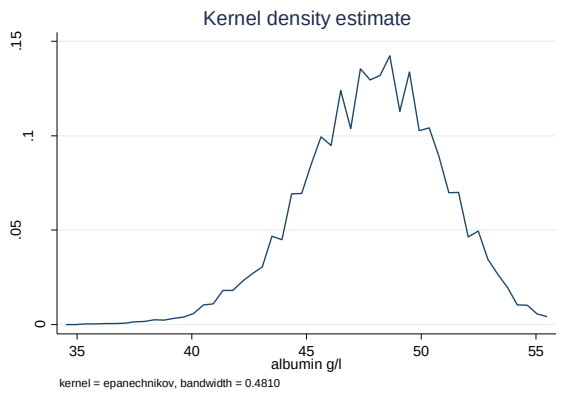


Females (n=7,111)

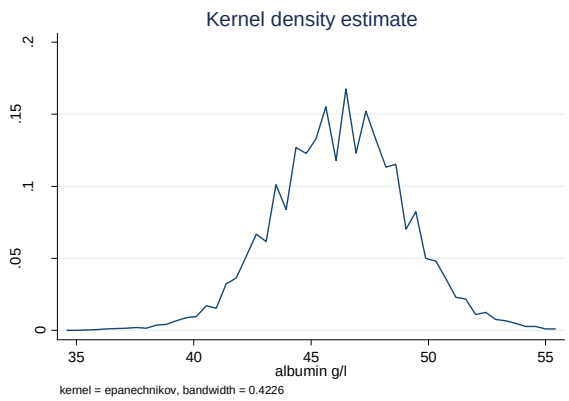


Distribution of Albumin g/L by gender

Males (n=5,710)



Females (n=7,111)



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

KIDNEY FUNCTION

A kidney function panel was undertaken, which includes creatinine and urea. Creatinine is the main indicator employed to assess kidney function.

Creatinine

What is its clinical significance?

Creatinine is used to estimate glomerular filtration rate (eGFR), which is a standard measure of kidney function. Chronic kidney disease is an increasing health problem.

What is it?

Creatinine is a chemical waste product of muscle function, which is passed through the kidneys and excreted in urine. Levels, therefore, indicate how effectively the kidneys are 'cleaning' the blood.

How is it measured?

Creatinine was measured from serum samples using an enzymatic method on the Roche P module analyser.

Inter and intra assay coefficients of variation were less than 4%.

Are there clinical cutpoints?

Equations to calculate eGFR based on creatinine have recently been published to identify increasing levels of kidney disease, dependent on age, gender and levels[30]:

- white men with a creatinine level <0.9 mg/dL, 141 x (serum creatinine/0.9)$^{-0.411}$ x(0.993)^{age};
- for serum creatinine level > 0.9 mg/dL, 141 x (serum creatinine/0.9)$^{-1.209}$ x (0.993)^{age}.
- white women with a serum creatinine level <0.7 mg/dL, 144 x (serum creatinine/0.7)$^{-0.329}$ x (0.993)^{age};
- for serum creatinine level >0.7 mg/dL, 144 x (serum creatinine/0.7)$^{-1.209}$ x (0.993)^{age}

[30] *Levey AS, Stevens LA, et al. A New Equation to Estimate Glomerular Filtration Rate. Ann Intern Med. 2009; 150:604-612.*

Table 15: Classification of kidney function[31] from eGFR equations

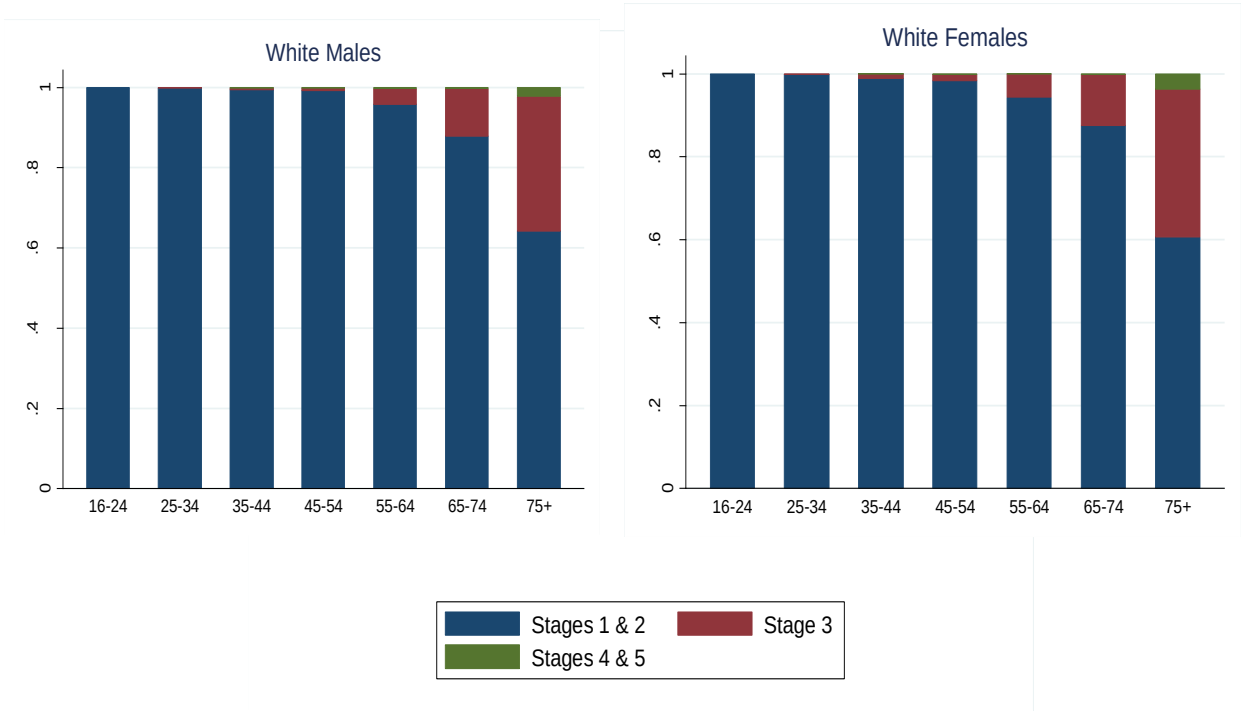
Stage of chronic kidney disease	Kidney function	eGFR
1	Normal	90+
2	Mildly reduced	60-89
3	Moderately reduced	30-59
4	Severely reduced	15-29
5	Very severely reduced	<15

What should be considered in analyses?

To our knowledge no other factors require consideration in the analysis.

Distribution in *Understanding Society*

Figure 10 Stages of kidney disease (defined by eGFR) by age and gender



[31] *Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group (2013) KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney International (Suppl. 3): 1–150*

Table 16 Creatinine [µmol/L], **by gender and age**

Males								
Health Survey for England, 2009 (n=1,027)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Creatinine	77.4 (1.35)	81.0 (1.25)	81.3 (1.0)	83.0 (1.04)	80.5 (0.96)	85.9 (1.46)	96.2 (2.96)	82.4 (0.55)
Understanding Society (England sub-sample; n=4,844)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Creatinine	79.7 (0.67)	82.8 (0.57)	82.6 (0.50)	84.4 (0.48)	84.1 (0.60)	88.2 (0.61)	95.9 (1.21)	84.6 (0.26)
Females								
Health Survey for England, 2009 (n=1,165)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Creatinine	60.6 (0.94)	61.3 (0.88)	62.7 (0.73)	64.2 (0.90)	66.1 (1.36)	69.2 (1.73)	74.4 (1.68)	64.9 (0.45)
Understanding Society (England sub-sample; n=5,984)								
Age group	16-24	25-34	35-44	45-54	55-64	65-74	75+	Total sample
Mean Creatinine	63.2 (0.61)	64.6 (0.55)	64.7 (0.47)	66.8 (0.38)	67.3 (0.43)	70.3 (0.64)	79.6 (1.24)	67.3 (0.24)

Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded. Sample weights are employed for each dataset. Standard errors of the mean in parentheses.

Table 16shows that mean creatinine values from the England component of *Understanding Society* are close to those obtained from the HSE (2009) by gender and across age groups.

Urea

What is its clinical significance?

High urea levels indicate poor kidney function which may be due to acute or chronic kidney disease. However, its use has generally been replaced as a biomarker by the more robust eGFR measure.

What is it?

Urea is a waste product of the breakdown of proteins. High levels indicate that the kidneys are not functioning effectively.

How is it measured?

Urea was measured from serum samples with a kinetic UV assay on a Roche P module analyser.

Inter and intra assay coefficient of variation was less than 3%.

Are there clinical cutpoints?

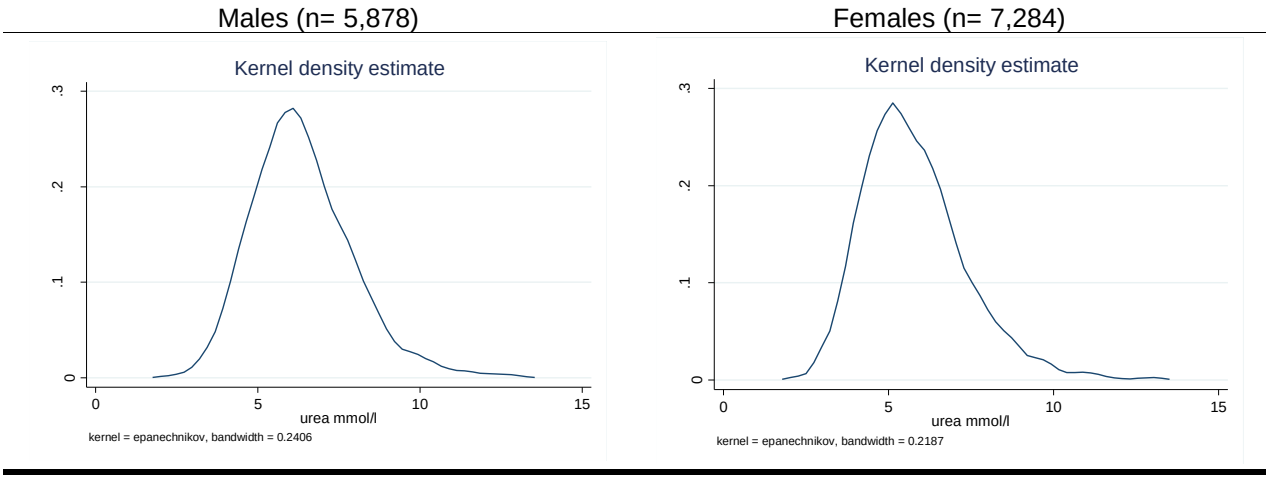
The normal range of urea is 2.5-7.8 mmol/L.

What should be considered in analyses?

Other conditions, besides kidney disease can affect urea levels such as congestive heart failure and recent heart attack.

Distribution in *Understanding Society*

Figure 11 Distribution of Urea [mmol/L] by gender



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

Hormones

Testosterone

What is its clinical significance?

Testosterone is a steroid that plays a central role in the development of secondary sexual characteristics in men. It is related to libido, building muscle mass and with aggression and competitive behaviours. Evidence suggests that low testosterone levels are associated with diabetes in men[32]. In women, high levels are associated with conditions such as polycystic ovarian syndrome.

What is it?

Testosterone is anabolic steroid, which build up muscles and tissues.

How is it measured?

Serum testosterone is measured by an electrochemiluminescent immunoassay on the Roche Modular E170 analyser.

Intra and inter assay coefficient of variation is less than 4%.

Are there clinical cutpoints?

Testosterone levels above or below the normal range are considered by many to be out of balance. In men testosterone levels are broad and considered within a normal range between 9-25 nmol/L and in women testosterone values are low and considered above normal at greater than 3.2 nmol/L. In *Understanding Society* the majority of values for women are below the lowest detection level for the analyser of 1 nmol/L.

What should be considered in analyses?

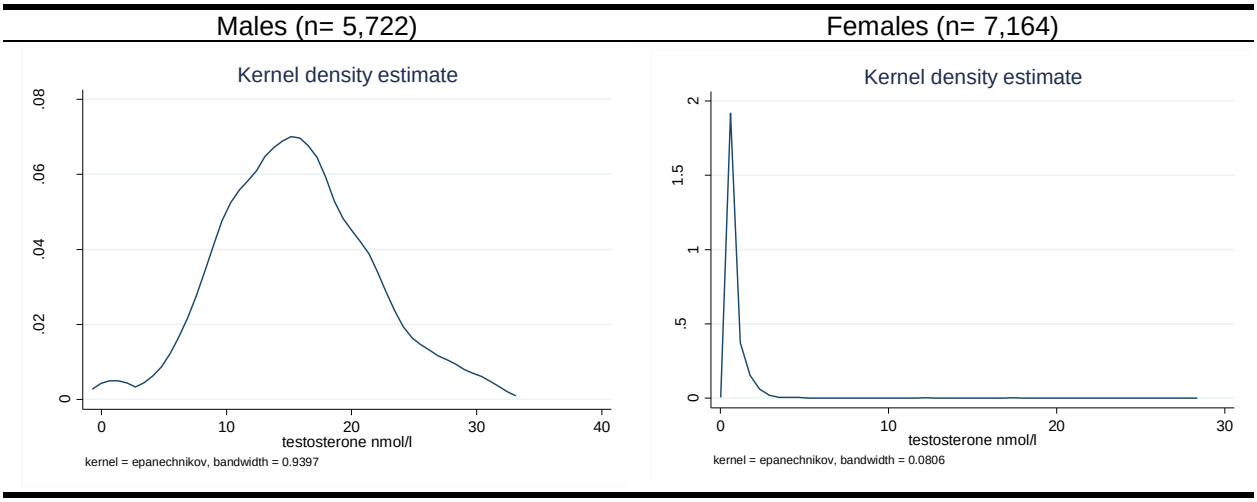
In *Understanding Society* we have measured total testosterone, which is independently associated with a number of outcomes such as diabetes. Testosterone varies by time of day such that values in the morning are higher than those found in the afternoon or evening[33].

Users should be aware that testosterone is bound by carrier proteins in the circulation[34] [35] . However, we have not measured steroid hormone binding globulin, the chief carrier protein that binds circulating testosterone.

[32] Beatrice AM, Dutta D, Kumar M, Kumbenahalli Siddegowda S, Sinha A, Ray S, Chowdhury S. Testosterone levels and type 2 diabetes in men: current knowledge and clinical implications. Diabetes Metab Syndr Obes. 2014 Oct 20;7:481-6.
[33] *Brambilla DJ, Matsumoto AM, Aroujo AB, McKinlay JB. The effects of diurnal variation on clinical measurement of serum testosterone and other sex hormone levels in men. J Clin Endocrinol Metab. 94: 907-913*
[34] *Sodergard R, Backstrom T, Shanbhag V,Carstensen H: Calculation of free and bound fractions of testosterone and estradiol-17 beta to human plasma proteins at body temperature. J Steroid Biochem16:801–810, 1982*
[35] Vermeulen A, Verdonck L, Kaufman JM. A critical evaluation of simple methods for the estimation of free testosterone in serum. J Clin Endocrinol. Metab. 84: 3666-3672, 1999

Distribution in *Understanding Society*

Figure 12 Distribution of Testosterone [nmol/L] by gender



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

Insulin-like growth factor-1 (IGF-1)

What is its clinical significance?

Low IGF-1 levels have been shown to be associated with heart disease and high levels have been shown to be predictive of some cancers.[36, 37]

What is it?

IGF-1 is a hormone, specifically an anabolic protein, which builds up organs and tissues. It plays an important role in growth and development in childhood and continues to affect adult anabolic processes.

How is it measured?

Serum IGF-1 is measured by an electrochemiluminescent immunoassay on IDS ISYS analyser.

Inter and intra assay coefficient of variation was less than 14%.

Are there clinical cutpoints?

There are no published clinical cutpoints for insulin-like growth factor-1. Normal reference values for IGF-1 vary in men and women and, because of the strong association of IGF-1 with age, these values are provided by age group in Table 17.

[36] *Seccareccia E, Brodt P. The role of the insulin-like growth factor-I receptor in malignancy: An update. Growth Hormone & IGF Research 22:2012;193–199*
[37] *Troncoso R, Ibarra C, Vicencio JM, Jaimovich E, Lavandero S. New insights into **IGF-1** signaling in the heart. Trends Endocrinol Metab. 2014; 25:128-37*

Table 17: IGF-1 reference values (nmol/L) in men and women by age

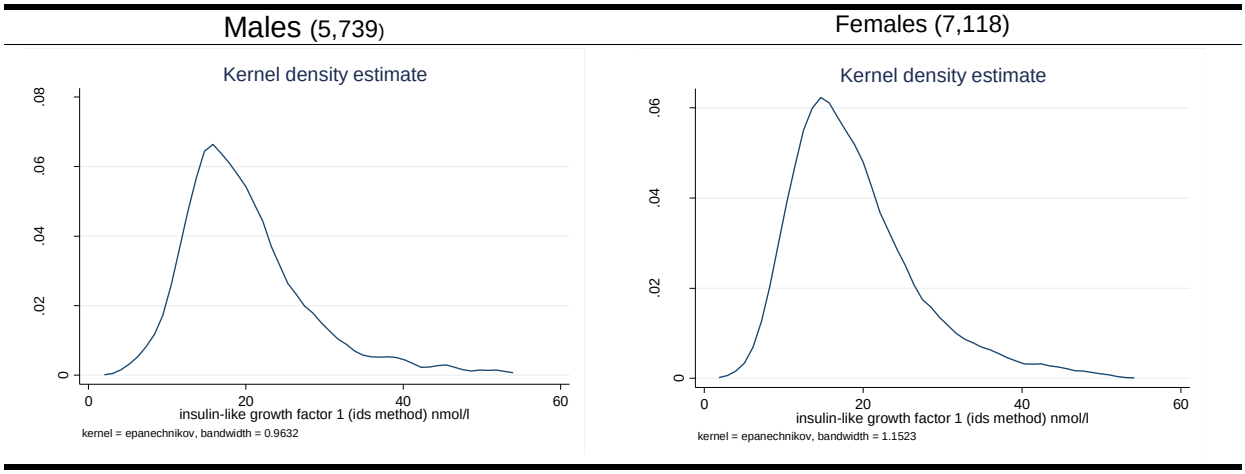
Age groups	Men	Women
17 - 18 yrs	20 - 56	35 - 73
19 - 20 yrs	21 - 85	21 - 51
21 - 25 yrs	18 - 42	12 - 44
26 - 39 yrs	15 - 37	12 - 44
40 - 54 yrs	14 - 32	12 - 44
55 - 88 yrs	11-30	12 - 44

What should be considered in analyses?

To our knowledge there are no factors that require consideration in the analysis.

Distribution in *Understanding Society*

Figure 13 Distribution of Insulin-like growth factor-1 (IGF-1)[nmol/l] by gender



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

Table 18 Insulin-like growth factor-1 (IGF-1)[nmol/l], by gender and age

Males							
ELSA, Wave 4 (2008-2009) (n= 2,814)							
Age group	50-54	55-59	60-64	65-69	70-74	75+	Total sample
Mean IGF1	17.3 (0.32)	16.3 (0.25)	16.7 (0.23)	16.4 (0.25)	16.1 (0.30)	14.1 (0.28)	16.1 (0.11)
Understanding Society (England sub-sample; n=2,737)							
Age group	50-54	55-59	60-64	65-69	70-74	75+	Total sample
Mean IGF1	17.1 (0.27)	17.0 (0.31)	16.3 (0.28)	16.3 (0.27)	15.2 (0.33)	14.5 (0.40)	16.1 (0.13)
Females							
ELSA, Wave 4 (2008-2009) (n=3,466)							
Age group	50-54	55-59	60-64	65-69	70-74	75+	Total sample
Mean IGF1	16.8 (0.34)	15.5 (0.23)	15.3 (0.19)	14.6 (0.22)	14.4 (0.24)	13.3 (0.22)	14.8 (0.10)
Understanding Society (England sub-sample; n=3,208)							
Age group	50-54	55-59	60-64	65-69	70-74	75+	Total sample
Mean IGF1	15.9 (0.23)	15.4 (0.30)	14.8 (0.24)	14.4 (0.426)	14.3 (0.33)	13.6 (0.29)	14.7 (0.10)

Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded. Sample weights are employed for each dataset. Standard errors of the mean in parentheses.

Table 18 shows that mean IGF-1 values from the England component of *Understanding Society* are close to those obtained from the ELSA study by gender and across age groups.

Dihydroepiandrosterone suphate (DHEAs)

What is its clinical significance?

DHEAs has been implicated in cardiovascular health; low levels are associated with CVD and all-cause mortality in older men,[38] whereas higher levels are related to better health outcomes such as lower risk of metabolic syndrome.[39]

What is it?

Dihydroepiandrosterone (DHEA) and its sulfate form DHEAs and are the most common steroid hormones in the body, and their levels decline with age[40] .

How is it measured?

DHEAs measures were performed using serum samples on a competitive immunoassay on the Roche E module analyser.

Intra and inter assay coefficients of variation were less than 4%.

[38] *Barrett-Connor E, Khaw KT, Yen SCC. A prospective-study of dehydroepiandrosterone sulfate, mortality, and cardiovascular-disease N. Engl. J. Med., 315 (1986), pp. 1519–1524*

[39] *Phillips AC, Carroll D, Gale CR, Lord JM, Arlt W, Batty GD. Cortisol, DHEAS, their ratio and the metabolic syndrome: evidence from the vietnam experience study Eur. J. Endocrinol., 162 (2010), pp. 919–923*

[40] *Labrie F, Belanger A, Luu-The V, Labrie C, Simard J, Cusan L, Gomez JL, Candas B. DHEA and the intracrine formation of androgens and estrogens in peripheral target tissues: its role during aging Steroids, 63 (1998), pp. 322–328*

Are there clinical cutpoints?

Data are continuous and there are no established clinical cutpoints. The expected ranges are shown in Table 19 below.

Table 19: Expected ranges of DHEAs (µmlol/L) in men and women by age group

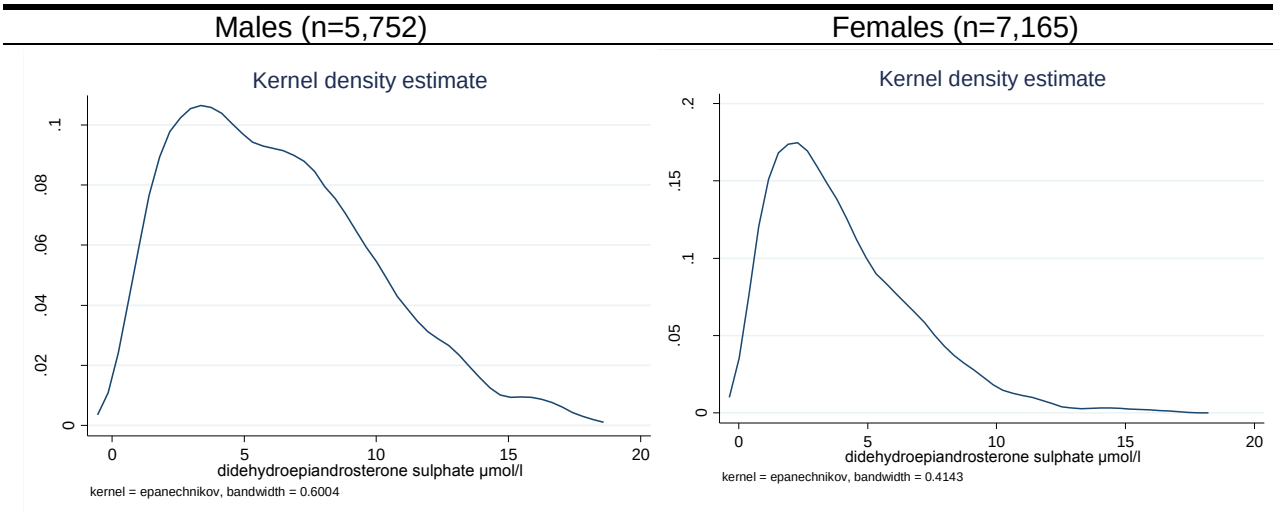
Age range	Men	Women
15 - 19 yrs	1.9 - 13.4	1.8 - 10.0
20 - 24 yrs	5.7 - 13.4	4.0 - 11.0
25 - 34 yrs	4.3 - 12.2	2.7 - 9.2
35 - 44 yrs	2.4 - 11.6	1.7 - 9.2
45 - 54 yrs	1.2 - 9.0	1.0 - 7.0
55 - 64 yrs	1.4 - 8.0	0.5 - 5.6
65 - 74 yrs	0.9 - 6.8	0.3 - 6.7
≥ 75 yrs	0.4 - 3.3	0.3 - 4.2

What should be considered in analyses?

To our knowledge there are no factors that require consideration in the analysis.

Distribution in *Understanding Society*

Figure 14 Distribution of DHEAS [µmol/l] by gender



Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded.
Understanding Society (Waves 2-3, sample weights employed)

Table 20 Dihydroepiandrosterone suphate (DHEAs) [µmol/l], by age and sex

Males							
ELSA, Wave 4 (2008-2009) (n= 2,803)							
Age group	50-54	55-59	60-64	65-69	70-74	75+	Total sample
Mean DHEAS	3.9 (0.11)	3.7 (0.09)	3.1 (0.07)	2.7 (0.08)	2.3 (0.07)	1.6 (0.06)	2.9 (0.04)
Understanding Society (England sub-sample; n= 2,744)							
Age group	50-54	55-59	60-64	65-69	70-74	75+	Total sample
Mean DHEAS	5.4 (0.25)	4.6 (0.16)	4.0 (0.15)	3.6 (0.15)	2.8 (0.12)	2.3 (0.11)	3.8 (0.06)
Females							
ELSA, Wave 4 (2008-2009) (n= 3,494)							
Age group	50-54	55-59	60-64	65-69	70-74	75+	Total sample
Mean DHEAS	2.8 (0.09)	2.3 (0.06)	1.9 (0.05)	1.7 (0.05)	1.4 (0.05)	1.2 (0.04)	1.8 (0.03)
Understanding Society (England sub-sample; n= 3,226)							
Age group	50-54	55-59	60-64	65-69	70-74	75+	Total sample
Mean DHEAS	3.5 (0.12)	2.9 (0.12)	2.6 (0.10)	2.2 (0.08)	2.1 (0.12)	1.7 (0.06)	2.5 (0.04)

Notes: Values belonging to the lowest/highest 0.5% of the distribution were excluded. Sample weights are employed for each dataset. Standard errors of the mean in parentheses.

Table 20 shows that mean DHEAs values from the England component of *Understanding Society* differ to those obtained from the ELSA study (2008-2009). However, the analyser employed by NUTH was changed between the ELSA analyses and those for *Understanding Society*, which may explain these differences.

DATAFILES, WEIGHTS ETC

The blood analyte data are released as part of the nurse health assessment dataset (SN 7251 at the UKDS). All of the analyte results are in the xlabblood_ns file, which combines both the GPS and BHPS samples.

The results of the physical measures, as well as information that might be required to adjust biomarker data such as medications are in the indresp_ns files (xindresp_ns, b_indresp_ns and c_indresp_ns) within the nurse health assessment data release. Individual BNF-coded medications are only available under special licence, while broad chapter codes and the key medications outlined above that need to be considered in the analysis of the biomarkers are in the EUL versions.

Some analysis weights have been prepared for the biomarker data to enable estimation samples to be representative of the general population. The general principles behind the weights and the weight naming conventions are consistent with those documented in the main *Understanding Society* User Guide.

The sample design involves stratification, clustering and weighting; since these design features affect standard errors they should be taken into account in analysis. A detailed discussion of how analyses might account for the complex sample design along with a description of the relevant variables and Stata syntax can be found in the main Understanding Society User Guide.

A cross-sectional weight variable for the combined GPS and BHPS sample (indbdub_xw) can be found in the xlabblood_ns datafile. This should be used for cross-sectional analysis combining both samples but using questionnaire data only from the same Wave in which the blood was collected. Note that this weight should not be used for analysis of either the GPS or BHPS sample separately, as the two samples have very different structures.

For longitudinal analysis – analysis in which for each respondent questionnaire data from at least two different Waves is used in conjunction with the blood data – two different longitudinal weights have been created for each of Waves 3 and 4 and will in future be created for each subsequent Wave. The first of these is for used with the combined GPS and BHPS sample, using questionnaire data from Waves 2 and 3 in conjunction with the blood data (c_indbdub_lw available in the c_indresp file), or for using questionnaire data from Waves 2, 3 and 4 with the blood data (d_indbdub_lw, available in the d_indresp file). The second is for using data from 1991 onwards for the BHPS sample in conjunction with the blood data: c_indbd91_lw (available in the c_indresp file) for 1991 to *Understanding Society* Wave 3 and d_indbd91_lw for 1991 (available in the d_indresp file) to *Understanding Society* Wave 4.

Technical details of how the weights were derived are presented in Appendix 2.

RELEASE OF THE DATA

The release of the biomarkers conforms to their classification for risk of disclosure, as agreed the *Understanding Society* Data Access Committee, and hence they are released in anonymised form through the UKDS via the EUL.

We request that researchers using the data notify us about errors, inconsistencies, and other problems with the data identified during their use of the data. We make use of this information in improving the data. Please raise any issues relating to data or data analysis with our user support service; https://www.understandingsociety.ac.uk/documentation/help.

We would also be very pleased to receive copies of publications using the data.

We will communicate information via Frequently Asked Questions on the *Understanding Society* web page about the data: https://www.understandingsociety.ac.uk/documentation/faq

CITATIONS AND ACKNOWLEDGEMENTS

Users should cite the data set in any publication using these data as below. They should also include an acknowledgement to the UK Data Service, study funders and Institute for Social and Economic Research. The suggested citation is:

> *The biomarker data from Understanding Society were collected by NatCen on behalf of the Institute for Social and Economic Research and funded by the Economic and Social Research Council. They are made available through the UKDS (SN 7251).*

Citation of the data:

University of Essex. Institute for Social and Economic Research and National Centre for Social Research, *Understanding Society: Waves 2 and 3 Nurse Health Assessment, 2010-2012* [computer file]. Colchester, Essex: UK Data Service [distributor], April 2013. SN:7251, http://dx.doi.org/

Citation of the User Manual and Glossary:

Benzeval, Michaela, Davillas, Apostolos, Kumari, Meena, Lynn, Peter. (2014). *Understanding Society:*UK Household Longitudinal Study: Biomarker User Guide and Glossary. Colchester: University of Essex.

APPENDIX 1 QUALITY CONTROL DATA

Brief details of the internal and external quality control information are provided here. A more detailed report is being prepared and will be made available on the *Understanding Society* website in due course.

Quality control processes are generally conducted internally and externally. Internal quality control measures how well measurements compare across time within a laboratory (ie does a sample measure the same if the measurement is made on day 'a' as it does on day 'a'+1). These quality control measures are sometimes called intra-assay coefficient of variation. External quality control reflects how the laboratory compares to other laboratories measuring the same analyte. A number of quality assessment schemes, which provide sample standards – often at more than one concentration - for comparison, are available. The NUTH uses the Welsh External Quality Assessment Scheme (WEQAS) for some analytes and the United Kingdom National External Quality Assessment Service (UK NEQAS) for others. These quality control measures are sometimes referred to as inter-assay coefficient of variation.

INTERNAL QUALITY CONTROL

Internal quality control measurements allow laboratories to assess how the assay varies from day-to-day. Samples with known values (one standard with a low level and one with a high level) are measured and the variation in actual measurement is expressed as a co-efficient of variation. Results for each biomarker are shown in Table A1.1, which shows that these coefficients are generally less than 5% and therefore well within acceptable limits.

EXTERNAL QUALITY CONTROL

External quality control methods use standards across laboratories and results are compared. The majority of the biomarkers in the *Understanding Society* dataset were analysed on a single analyser: the Roche P module analyser The NUTH laboratory participates in the WEQAS on a routine basis. We report the standard deviation index (SDI) in Table A1.2, where the SDI is an index of total error, including components of inaccuracy and imprecision, so lower values suggest more accurate measures. It is calculated as:

> (laboratory result – target value) / (WEQAS standard deviation * CF) where CF is a method-specific comparability factor. This adjustment ensures that each laboratory can compare their results with others using their own method, the peer reference method, and the overall mean of all groups.

A score below 1 SDI is good, and between 1-2 SDI is acceptable. The majority of monthly SDI figures for the biomarkers analysed on this machine had good EQA, with a few being acceptable only.

Steroid hormones (testosterone and DHEAs, IGF-1) and, ferritin and fibrinogen were part of external QA by the UK NEQAS. Samples are sent to a range of laboratories, the individuals laboratories perform the relevant assays and return the results to NEQAS. The results are analysed and a consensus result derived from all laboratory results to provide comparison with the overall median and peer group median i.e. laboratories using similar assay methodology. All results for analytes in *Understanding Society* met acceptability criteria.

Table A1.1: Internal quality control data from the Department of Clinical Biochemistry, NUTH.

	Level 1 low standard (mean)	**CV %**	**Level 2 high standard (mean)**	**CV%**
Total Cholesterol (mmol/L)	2.6	1.4	6.6	1.3
HDL Cholesterol (mmol/L)	0.7	6.0	1.7	4.3
Triglycerides (mmol/L)	0.87	2.6	1.98	1.7
C-Reactive Protein (mg/L)	7.0	0.1	53	1.8
Ferritin (ug/L)	33	2.4	388	1.8
ALP (U/L)	29	4.5	233	3.4
ALT (U/L)	23	4.3	172	1.6
AST (U/L)	39	3.2	240	0.6
GGT (U/L)	27	2.2	138	1.3
Albumin (g/L)	26	2.4	45	1.3
Urea (mmol/L)	5.4	2.4	23.1	1.3
Creatinine (umol/L)	55	2.6	586	3.2
Testosterone (nmol/L)	2.8	3.0	34.4	2.3
CMV IgG	1.3	1.4	23.8	2.1
CMV IgM	0.2	3.9	2.2	3.6
DHEAS (umol/L)	3.4	3.3	18.1	3.5
IGF-1 (nmol/L)	3.9	13.6	32.6	7.4
HbA1c (mmol/mol)	34	3.88	86	1.72
Hb (g/L)	62	0.4	164	0.8
Fibrinogen (g/L)	1.82	6.91	2.93	3.48

Table A1.2: External quality control data from the Department of Clinical Biochemistry, NUTH for analytes measured using the Roche P module analyser. Standard deviation index (SDI) data per month that samples were analysed.

Date	**Cholesterol** (mmol/L)	**HDL** (mmol/L)	**Triglycerides** (mmol/L)	**HbA1c** (mmol/mol)	**ALT** (U/L)	**AST** (U/L)	**ALP** (U/L)	**GGT** (U/L)	**Albumin** (g/L)	**Creatinine** (umol/L)	**Urea** (mmol/L)
Dec 13	0.35	0.45	0.24	0.60	0.38	0.54	0.14	0.17	1.33	1.04	0.92
Jan 14	0.25	0.62	0.15	0.19	0.53	0.25	0.63	0.37	0.72	0.48	0.35
Feb 14	0.39	0.36	0.26	0.47	0.42	0.32	1.05	0.42	0.95	0.94	0.38
March 14	0.21	0.36	0.22	0.41	0.42	0.06	1.11	0.19	1.10	0.34	0.85
April 14	0.89	0.49	0.26	0.66	0.52	0.12	0.54	0.24	1.13	0.37	0.70
May 14	0.31	0.58	0.27	1.12	0.47	0.06	1.09	0.70	0.37	1.22	1.03
June 14	0.86	1.24	0.30	0.41	0.16	0.23	0.70	0.26	0.26	0.95	0.81
July 14	0.15	0.64	0.23	0.07	0.21	0.20	0.37	1.40	1.06	1.02	0.43

Notes: A score below 1 SDI is good, between 1-2 SDI is acceptable.

APPENDIX 2: DERIVATION OF ANALYSIS WEIGHTS

This appendix describes how the various analysis weights for use with the blood data were derived.

The Wave 3 longitudinal blood weight for the combined sample (c_indbdub_lw) is based upon the equivalent nurse weight (c_indnsub_lw), with an additional adjustment for non-response to the blood measures. [Derivation of the nurse weights is described in the separate User Guide to the nurse health assessment data[41], McFall et al, 2014.] The adjustment comes from a logistic regression model predicting the presence of blood data and for which the base is all sample members who responded to the individual interview at both Waves 2 and 3 and the nurse visit, minus those who were ineligible for the collection of blood (those who reported a clotting or bleeding disorder, taking anticoagulant drugs, or having ever had a fit). Predictor variables in the model are a range of social, demographic and economic indicators from the Wave 3 household and individual questionnaires. The adjustment factor is the reciprocal of the model-predicted propensity for blood measures to be present. Thus, the adjustment is designed to deal simultaneously with drop-out at each stage in the process of obtaining blood measures, viz. consent to give blood, successfully taking blood, and successfully extracting analytes from the blood sample. The weighted sample should be representative of the population net of those who would have been ineligible for the nurse visit (pregnant, or inadequate English to complete a survey interview) or for the blood sample (clotting or bleeding disorder, taking anticoagulant drugs, or having had a fit).

[41] McFall, Stephanie L., Petersen, Jakob, Kaminska, Olena, Lynn, Peter. (2014). *Understanding Society* –UK Household Longitudinal Study: Waves 2 and 3 Nurse Health Assessment, 2010-2012, Guide to Nurse Health Assessment. Colchester: University of Essex. https://www.understandingsociety.ac.uk/documentation/health-assessment

The Wave 3 longitudinal blood weight for the BHPS sample (c_indbd91_lw) is derived in an analogous way. The only differences from the above description are that the base for the model is all eligible BHPS sample members who responded to the individual interview at Waves 2 and 3 and the nurse visit and that the adjustment is applied to c_indns91_lw.

The cross-sectional weight (c_indbdub_xw) is derived from the longitudinal weight c_indbdub_lw by applying the weight share method[42] within households. By this method, sample members who do not have a value of c_indbdub_lw by virtue of not having completed the interview at both Waves 2 and 3 are assigned a value of c_indbdub_xw based on an assumption of equal conditional response propensities within households.

Wave 4 longitudinal blood weights are based on the equivalent Wave 3 longitudinal blood weights, with an additional adjustment for conditional non-response to the Wave 4 individual interview. The adjustment factor is the reciprocal of the model-predicted propensity from a logistic regression model of response to the Wave 4 interview conditional on having responded at Waves 2 and 3 and having blood measures. Note that in the case of the BHPS sample there were no significant predictors in the model (the conditional response rate was over 96%), and hence no adjustment, so d_indbd91_lw is proportional to c_indbd91_lw.

For each weight variable the final step was to scale the weights to a mean of 1.00.

[42] Lavallée, Pierre (2007) *Indirect Sampling*. New York: Springer