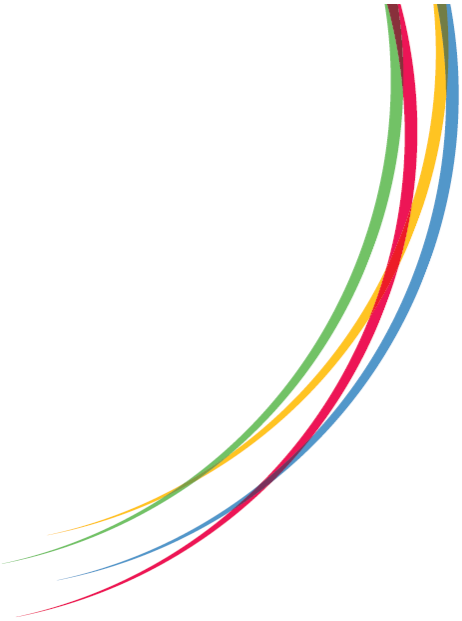


UNDERSTANDING SOCIETY

The UK Household Longitudinal Study

Waves 1-5

Quality Profile

Edited by
Peter Lynn & Gundi Knies
Institute for Social and Economic Research
University of Essex
Colchester
Essex

Version 1.2, February 2016

Understanding Society
THE UK HOUSEHOLD LONGITUDINAL STUDY

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

FOREWORD

The purpose of this Quality Profile is to provide an overview of all the aspects of the design and implementation of *Understanding Society* that bear upon the quality of the data that are released as a research resource.

We hope that the Quality Profile will be an invaluable resource for all data users (actual or potential), users of research based on *Understanding Society* data, and funders, as well as for anyone who is engaged in the design or implementation of similar studies.

This Quality Profile relates to the *Understanding Society* Main Survey which collects information from the UK General Population Sample (GPS) and the Ethnic Minority Boost Sample (EMBS). From the second round of interviews the main survey also includes information collected from continuing participants of the British Household Panel Survey (BHPS), a household panel survey of around 8,000 households in the UK, which has completed 18 annual waves of data collection and has been run by the Institute for Social and Economic Research (ISER) since it began in 1991. It also covers additional data collections such as biomarkers and linked administrative records which make *Understanding Society* an invaluable resource for research into life in the UK in the 21st Century.

The Quality Profile starts with an overview of the background of the Study, which includes its vision and mission, how the Study is funded and governed, and how the Study content is decided. It then takes the reader through the many processes and tasks involved in collecting the data and making them available to analysts, with an emphasis on how quality is controlled at each stage and what we know about quality-relevant outcomes.

With a study of such complexity, the amount of information that could be included in a publication such as this is enormous. We have attempted to focus on information that we judge likely to be of central interest to a range of potential readers. We have of course also been restricted by the constraints of time and resources. Consequently, the content is biased towards information that was readily available and we have been selective in the additional analysis that we have undertaken. We would welcome feedback and we envisage expanding the content of future editions of the Quality Profile in response to reader comments and to reflect the emerging findings of ongoing methodological research that will no doubt shed light on other aspects of survey quality.

Peter Lynn (plynn@essex.ac.uk)
Gundi Knies (gknies@essex.ac.uk)

Colchester, February 2016

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ARF	Address Record Form
BHPS	British Household Panel Survey
CAPI	Computer aided personal interview
CASI	Computer aided self-interview
CASCOT	Computer assisted structured coding tool
DAC	Data Access Committee
DWP	Department for Work and Pensions
ESRC	Economic and Social Research Council
ELSA	English Longitudinal Study of Ageing
EMBS	Ethnic Minority Boost Sample
ESAC	Ethnicity Strand Advisory Committee
ECHP	European Community Household Panel
FACS	Family and Children Study
GHQ	General Health Questionnaire
GOR	General Office Region
GPCS	General Population Comparison Sample
GPS	General Population Sample
SOEP	German Socio-economic Panel Survey
HBAC	Health and Biomarker Advisory Committee
HRS	Health and Retirement Survey
HSE	Health Survey for England
HILDA	Household, Income and Labour Dynamics in Australia
IP	Innovation Panel
ISER	Institute for Social and Economic Research
HIS	Integrated Household Survey
HRP	Household Reference Person
IWM	Inter-wave mailing
LFS	Labour Force Survey
LSOA	Lower Layer Super Output Area
MAC	Methodological Advisory Committee
MCS	Millennium Cohort Study
NatCen	National Centre for Social Research
NISRA	Northern Ireland Statistics and Research Agency
ONS	Office for National Statistics
SIC	ONS Standard Industry Code
SOC	ONS Standard Occupational Classification
OSM	Original Sample Member
PMDB	Panel Maintenance Database
PAPI	Paper aided personal interview
PSM	Permanent Sample Member
PSU	Primary Sampling Unit
SAC	Scientific Advisory Committee
SLT	Scientific Leadership Team
SDQ	Strengths and Difficulty Questionnaire
TSM	Temporary Sample Member
UKDS	UK Data Service
UKHLS	UK Household Longitudinal Study (official acronym for *Understanding Society*)

1. BACKGROUND: *UNDERSTANDING SOCIETY*: THE UK HOUSEHOLD LONGITUDINAL STUDY

Understanding Society is a longitudinal survey following members of approximately 40,000 households (at Wave 1) in the United Kingdom (England, Scotland, Wales and Northern Ireland). All persons in initially recruited households are re-visited at annual intervals thereafter to collect information on changes to their household and individual circumstances. Over the first five waves, interviews have been predominantly carried out face-to-face in respondents' homes by trained interviewers. Data collection began in January 2009 and has been carried out continuously since then.

The overall purpose of *Understanding Society* is to provide high quality longitudinal data about subjects such as health, work, education, income, family, and social life to help understand the middle term and long term effects of social and economic change, as well as policy interventions designed to impact upon the general wellbeing of the UK population. To this end, the Study collects both objective and subjective indicators and offers opportunities for research within and across multiple disciplines such as sociology and economics, geography, psychology and health sciences. The Study also provides a platform for additional data collections. Through a combination of asking a number of questions that also feature on other national or international (longitudinal and cross-sectional) surveys and the addition of innovative data collection methods such as data linkage, the Study does not only deliver breadth in content but also depth.

The key strengths of the Study are:

- Longitudinal and long-term: By repeatedly collecting data from the same individuals over many years, changes in people's lives can be understood and causality can be better identified than with cross-sectional survey data.
- Household focus: The collection of data from every adult and child aged 10 or older in the household of each sample member each year gives high quality and continuous information on the family and household environment in which individual dynamics takes place.
- Statistical precision: The large sample size permits analysis of small sub groups such as teenage parents and allows analysis at regional and country level.
- Ethnic minorities: the sample includes an over representation of ethnic minority groups to provide a more detailed and meaningful analysis of the make-up of UK society.
- Breadth of topics and interdisciplinarity: as well as core demographic and economic topics, the Study collects data on politics, health, attitudes, environment and more. In particular, the collection of biomarkers provides the opportunity for cutting-edge investigation that crosses the boundaries of social and medical research.

Understanding Society is part of an international family of national household panel studies. Opportunities exist for powerful and insightful cross-national research,

thanks to ongoing dialogue and collaboration between the Principal Investigators and Co-Investigators of these studies, which has led to a considerable degree of harmonisation of concepts, questions and key design features. From time to time this dialogue has taken the shape of more formal organisation, as funding and opportunities have allowed. One outcome has been the production of a five-country harmonised dataset known as the Cross-National Equivalent File (Frick et al. 2007; Lynn and Kaminska 2010). This is currently being updated to include *Understanding Society*. *Understanding Society* investigators initiated and organised the first International Panel Survey Methods Workshop at Essex in 2008. The workshop has taken place biennially since then, bringing together most of the world's household panel studies along with other longitudinal surveys that share similar methodological concerns. These events – and consequent networks and collaborations – have surely contributed to survey quality through the sharing of knowledge, best practice and innovative ideas.

Understanding Society is primarily funded by the UK Economic and Social Research Council (ESRC), with additional funding from multiple government departments (the Department for Work and Pensions (DWP), the Department for Education, the Department for Transport, the Department for Culture, Media and Sport, the Department for Communities and Local Government, the Department of Health, the Scottish Government, the Welsh Assembly Government, the Northern Ireland Executive, the Department for Environment, Food and Rural Affairs, and the Food Standards Agency).

2. GOVERNANCE AND MANAGEMENT STRUCTURES

Understanding Society is led by the *Scientific Leadership Team* (SLT) which is based primarily at the Institute for Social and Economic Research (ISER) at the University of Essex, UK, though the team has included one or two external members at various times. Professor Nick Buck was the Director of *Understanding Society* from its inception until July 2015, when Professor Michaela Benzeval took over the role. The Director is supported by an *Executive Team* which is collectively responsible for all aspects of management of the Study and strategic direction, including obtaining funding. The Executive Team consists of the Director and the Associate Directors responsible for each of the major work streams (methodology, surveys, data, innovation, outreach and policy) and the Study's project manager. It meets monthly, with additional business conducted online between meetings.

Data collection and associated tasks are subcontracted to a survey agency by means of competitive tender. The survey work is retendered every three years. The first two data collection contracts (for Waves 1-2 and 3-5 respectively) were both awarded to NatCen (known as the National Centre for Social Research at the time of the start of the *Understanding Society* contract), while the contract for Waves 6-8 was awarded to TNS BMRB. Fieldwork for all of the data collection reported in this release of the Quality Profile was therefore the responsibility of NatCen. NatCen subcontracted data collection in Northern Ireland to the Central Survey Unit of the Northern Ireland Statistics and Research Agency (NISRA).

The survey agency is an important strategic partner in the delivery of *Understanding Society* and collaborates closely with the SLT at ISER in many aspects of the design and implementation of the survey. While ISER has the primary responsibility for design work, the survey agency manages fieldwork, editing, coding and data entry and advises on the design of all research instruments. ISER plays a major role in quality control through detailed specification of fieldwork practices; survey materials; editing and coding requirements; and monitoring and analysing weekly fieldwork progress reports. The working relationship between ISER and the agency is reinforced by an agreed set of survey-specific procedures to ensure adequate response and effective data quality. Full details of technical aspects of the data collection and fieldwork, coding, and data processing are found in the *Technical Reports,* published each wave on the *Understanding Society* website.[1]

The SLT receives advice from four advisory committees: the *Scientific Advisory Committee* (SAC) and three specialist advisory committees, the *Methodological Advisory Committee* (MAC), *Ethnicity Strand Advisory Committee* (ESAC), and the *Health and Biomarker Advisory Committee* (HBAC). The SAC has responsibility for the oversight and stewardship of *Understanding Society* content and data collection methods. It advises on matters of content, design, sampling, new research and other activities affecting the Study. It also promotes the continuation, development and use of *Understanding Society* in research and teaching, as well as appreciation of the

[1] Technical reports are published at https://www.understandingsociety.ac.uk/documentation/mainstage/technical-reports.

contribution the survey makes to a better understanding of today's society. The committee also has a role in informing the development of public policy, advances scientific agendas and enhances methodological development. The Chairs of the three specialist advisory committees are also members of the SAC. In 2014, a Foresight Steering Group was formed, consisting of policy makers, representatives of promoting sector agencies and think tanks. This group advises the Study in general on developing policy impact, but is particularly focused on supporting the role of the *Understanding Society* Policy Unit.

The MAC provides advice – primarily from the operational and survey methodological community - about all matters regarding methodology, technology and innovation. The ESAC shapes the ethnicity strand of *Understanding Society* to ensure it meets the needs of the research and user communities concerned with questions of ethnicity. The work of this committee ranges from helping the ethnicity team engage users and potential users to discussing changing research priorities and the dissemination of results. The HBAC aims to ensure that the health and biomarker strand of *Understanding Society* meets the needs of the research and user communities. This committee advises on preferred measures of physical and mental health and biomarkers, on ethical issues in biomarker and health indicator data collection, and on engaging users.

Representatives of co-funding government departments meet with the Study team quarterly to discuss their engagement with *Understanding Society*.

The long-term development of *Understanding Society* is the responsibility of the *Governing Board*, which is responsible to ESRC, and includes internationally recognised academics. The Chair of SAC and the Co-Funders Group are members of the Governing Board. The Governing Board also ensures the financial resources required for the future of the Study are in place and has the power to act as the arbitrating authority should there be any major divergence of views over the general development of the Study. ESRC has also established a Management Board which reviews the general progress of the Study and consists of representatives of ESRC, the current fieldwork agency and the SLT.

Until 2015, access to data from the Study was overseen by a Data Access Committee (DAC), convened by the ESRC. This group developed the data access strategy and considered applications to use more disclosive data (Special Licence and Secure Data Access applications via the UK Data Service (UKDS)). In 2015, responsibility for consideration of all applications for use of data from the survey was delegated to the SLT, and the DAC was abolished. At the same time a cross-study access committee (METADAC) was established by the research councils to oversee access to survey data linked with genetics data and biological samples.

The current membership of all the advisory committees and the Governing Board is listed on the *Understanding Society* website[2].

[2] See https://www.understandingsociety.ac.uk/about/people.

3. STUDY DESIGN IN A NUTSHELL

Understanding Society is a panel survey of individuals in their household context. The initial sample consisted of all persons within a representative probability sample of households in the UK. The sample has three components. The largest part is the *General Population Sample* (GPS). In addition there is an *Ethnic Minority Boost Sample* (EMBS) and there is the sample that previously constituted the *British Household Panel Survey* (BHPS). Furthermore, a subset of the GPS is designated the *General Population Comparison Sample* (GPCS), to provide a comparator group to the EMBS. Members of the GPCS are asked additional questions that are otherwise only asked of ethnic minorities. A subgroup of the GPS and BHPS sample was eligible for the Wave 2 & Wave 3 *Nurse Health Assessment*, providing direct health measures and biomarkers. There is an extended discussion of sample design in section 4.

Data are collected primarily through annual interviews with each sample member aged 16 or older. At Waves 1-5 most of these interviews were carried out face-to-face in respondents' homes, though a small number were carried out by telephone. Sample members aged 10 to 15 are asked to complete a paper self-completion questionnaire each year. Once a sample member reaches the age of 16, they become eligible for a full interview. In the year of their first full interview, such sample members are referred to as "Rising 16s". Data collection for a single wave is scheduled across 24 months, so each successive pair of waves has a 50% temporal overlap. The sample is divided into 24 monthly samples. Data collection procedures are discussed in detail in section 6.

The survey questionnaires were developed following a period of extensive consultation over content. A considerable number of questions are designated as part of the "annual core" and are therefore asked in identical form at every wave. Other questions form part of "rotating modules" that are administered every two, three or four waves. At each wave, identical questionnaires are administered to all sample members, with the exception that an "Extra 5 minutes" of questions are asked of all ethnic minorities and of the GPCS. In Waves 2 and 3, *Understanding Society* augmented survey questions with direct health assessments and the collection of blood samples. Some segments of the sample were excluded from this component of the Study. At each wave, addresses and consents to link to a range of administrative records are collected, which opens up additional avenues for innovative research. The design and content of all the survey instruments are described in section 5.

In addition to the Main Survey, there is a separate survey run in parallel, the Innovation Panel (IP), which is designed for methodological development and testing. The IP is fielded in the year before the Main Survey and has been used to test a range of survey design and implementation features that might affect participation, costs, or measurement. The design of the IP is broadly similar to that of the main study, but fieldwork is constrained to one calendar quarter each year, and the instruments are somewhat different from the Main Survey. The IP is not included in this Quality Profile; details presented here relate to the main study.

Understanding Society furthermore provides a platform for Associated Studies, providing researchers with an opportunity to collect additional data using a wide range of methods and approaches, such as follow-up quantitative or qualitative studies of particular sub-groups within the sample.[3] To date ten Associated Studies have been approved by the SLT; they all draw on the IP sample and are therefore not included in this Quality Profile.

[3] For further information on Associated Studies, see https://www.understandingsociety.ac.uk/research/get-involved/associated-studies.

4. SAMPLE DESIGN AND SAMPLE STATUS

This section outlines the sample design for the core components of the *Understanding Society* main study sample – the GPS, EMBS, BHPS sample and the GPCS – as well as the nurse health assessment. The design of the core components is described in more detail in an *Understanding Society* working paper, see Berthoud et al. (2009).

4.1 GENERAL POPULATION SAMPLE

The GPS is based upon two separate samples of residential addresses, one covering England, Scotland and Wales and a separate sample for Northern Ireland.

The sample for England, Scotland and Wales is a proportionately stratified (equal probability), clustered sample of addresses that was selected in two stages. The first stage was to select a sample of postcode sectors as the primary sampling units (PSUs). The second stage was to select addresses within each sampled sector. Prior to selection, any postcode sector with fewer than 500 residential addresses was grouped with an adjacent sector and thereafter treated as a single sector. The list of all sectors was then sorted into twelve geographical strata, consisting of ten regions in England plus Scotland and Wales as separate strata. Within each of the twelve strata, sectors were sorted into three sub-strata based upon the proportion of household reference persons classified as non-manual workers, from 2001 Census data. Within each of the 36 sub-strata, sectors were then sorted into three further sub-divisions based on population density (households per hectare) and within each of the 108 resultant sub-divisions, sectors were listed in order of ethnic minority density. From the sorted list, a systematic random sample of 2,640 sectors was selected, with probability proportional to the number of residential addresses in the sector. These sectors were then allocated systematically to 24 monthly samples, with 110 sectors in each monthly sample. Within each postal sector, 18 addresses were selected from the Postcode Address File using systematic random sampling. The England, Scotland and Wales GPS sample is therefore based upon an initial sample of 47,520 addresses.

Northern Ireland has an unclustered systematic random sample of addresses: 2,395 addresses were selected in a single stage from the Land and Property Services Agency list of domestic addresses. In total, the GPS is therefore based upon a sample of 49,915 addresses.

At each address, the final stage of sampling was carried out by field interviewers. This consisted of identifying persons to be defined as sample members. All persons resident at each sample address at the time the interviewer made contact were deemed to be a sample member, with the exception of the small proportion of addresses that contained more than three dwellings or households. In those cases, three dwellings or households were sub-sampled at random.

4.2 General Population Comparison Sample

The GPCS consists of one sampled address in each of 40% of the selected postal sectors in GPS component for Great Britain. In other words, of the 2,640 general population sectors, 1,056 of them contain a GPCS address. The people residents at those addresses at Wave 1 are designated as members of the GPCS, regardless of ethnic group membership. Members of the GPCS are a random subsample of the GPS component and they form part of the GPS.

4.3 Ethnic Minority Boost Sample

The EMBS was designed to provide interviews with at least 1,000 adults from each of five groups: Indian, Pakistani, Bangladeshi, Caribbean, and African.

The initial step was to identify postal sectors with relatively high proportions of relevant ethnic minority groups, based upon 2001 Census data and more recent Annual Population Survey data. The set of 3,145 sectors constituted approximately 35% of the sectors in Great Britain and covered between 82% and 93% of the population of the five ethnic minority groups.

The 3,145 sectors were sorted into four strata based on the expected number of ethnic minority households that would be identified by the sampling and screening procedures. See Berthoud et al. (2009) for details. All sectors were included for the stratum where a yield of three or more households was expected. In the other three strata, sectors were sub-sampled at rates of 1 in 4, 1 in 8, or 1 in 16 respectively. This was done to constrain the number of sectors that might have just one or two eligible sample households (or even none). The total number of postal sectors selected for inclusion in the EMBS was 771. Of these 6 were in Scotland, 7 were in Wales, and the remaining 758 were in England, with a concentration in London (412 sectors).

The number of addresses selected per postal sector ranged from 15 to 103. Sampling fractions varied across the sectors in a way designed to deliver target numbers of respondents in each target ethnic minority group with adequate statistical efficiency. See Berthoud et al. (2009) for more details. In sectors selected for both the GPS component and the EMBS, a single systematic sample of the required total number of addresses was selected and allocated in a systematic way to the two sample components, thus ensuring that both sample components are spread throughout the whole sector.

The final stage of sampling was done by the interviewers. The steps are described in the *Project Instructions for Interviewers*[4]. At addresses containing more than three dwellings or households, the procedures to sub-select dwellings or households were as described above for the GPS component. Within each household, rather than all resident persons becoming sample members, there were three additional steps:

- A “screen” was carried out to identify whether there were any persons from target ethnic groups in the household.

[4] These are available at https://www.understandingsociety.ac.uk/documentation/mainstage/fieldwork-documents.

- A random mechanism was applied to certain target groups identified by the screen in order to select only a desired proportion into the sample (non-mixed Indian, Pakistani, non-mixed Caribbean, African, Far Eastern, Middle Eastern). For other target groups, all resident persons were included in the sample (mixed Indian, Bangladeshi, mixed Caribbean, Sri Lankan, Chinese, Turkish).
- In households included in the sample in the previous two steps, all members of target ethnic groups were deemed to be members of the EMBS (including children). All persons of other ethnic groups are not EMBS members. They will be interviewed as temporary sample members for so long as they remain co-resident with at least one EMBS member.

The overall sampling fractions combine a) the probability of sampling the sector, b) the fraction of addresses selected within the sector, and c) the probability of a household being retained following the application of the random selection.

4.4 FORMER BHPS SAMPLE

The BHPS sample was integrated into *Understanding Society* at Wave 2. The sample issued at Wave 2 consisted of all members from the BHPS sample who were still active at Wave 18 of the BHPS and who had not refused consent to be issued as part of the *Understanding Society* sample. It should be noted that the BHPS sample contains different components, including the original sample (first selected in1991), boost samples in Scotland and Wales (first selected in 1999), and a Northern Ireland sample (selected in 2001). For further details of the BHPS sample design, see sections 3 and 4 of the BHPS Quality Profile[5]. The BHPS sample members were randomly allocated to months 1 to 12 of the *Understanding Society* sample, so that data collection with the ex-BHPS sample always takes place in the first year of each wave. As Wave 18 of the BHPS had taken place in September-December 2008, this meant that the interval between the last BHPS interview and the first *Understanding Society* interview for this sample ranged between approximately 13 months and 27 months.

4.5 WAVE 2 & WAVE 3 NURSE HEALTH ASSESSMENT

An additional visit was made to each sample household during 2010-12 by a nurse in order to administer a health assessment interview, and to collect physical measures and blood samples. These visits were carried out by professional nurses, typically around five months after the survey interview. The visits for GPS members took place following their Wave 2 interview (between May 2010 and July 2012) while the visits for BHPS sample members followed their Wave 3 interview (between June 2011 and July 2012). The EMBS was not included in the health assessment. Within the GPS and BHPS samples, eligibility for the nurse health assessment was restricted to sample members who completed the personal interview at the relevant wave, lived in England, Scotland or Wales (the Northern Ireland sample was therefore excluded), completed their interview in English, and were not pregnant. Additionally, in year 2 sub-sampling of the GPS took place in England (though not in

[5] Available online at https://www.iser.essex.ac.uk/bhps/quality-profile.

Wales or Scotland), with a random sample of 81% of the PSUs included in the health assessment.

4.6 SAMPLE STATUS AND FOLLOWING RULES

Persons from whom, and about whom, data are collected have one of three possible sample statuses: Original Sample Member (OSM), Temporary Sample Member (TSM), or Permanent Sample Member (PSM). The term OSM is arguably something of a misnomer as it simply indicates that a person is a full sample member. The other two statuses indicate people who are not themselves sample members but from whom data should be collected by virtue of their relationship to one or more sample members. They are therefore members of the sample for fieldwork purposes but not members of the statistical sample for longitudinal analysis[6]. Rather, the data they provide can be considered an attribute of one of more sample members. The definitions are as follows:

4.6.1 ORIGINAL SAMPLE MEMBERS (OSMS)

All members of GPS households enumerated at Wave 1, including absent household members and those living in institutions (such as halls of residence or boarding schools), who would otherwise be resident, are OSMs. All ethnic minority members of an enumerated household eligible in the EMBS are OSMs. In both of these samples, any child born to an OSM mother after Wave 1 and observed to be co-resident with the mother at the survey wave following the child's birth is an OSM. In the former BHPS sample, OSMs are those who were enumerated at the first wave of the sample from which they come (Wave 1 for the original sample, wave 9 for the Scotland and Wales boost samples, Wave 11 for Northern Ireland) or who were subsequently born to an OSM mother or father (or both). From Wave 2 onwards of *Understanding Society*, in the former BHPS sample, as for the rest of the *Understanding Society* sample, only children born to an OSM mother will themselves become an OSM.

OSMs, of all ages, are followed for interview and remain eligible for survey participation as long as they are resident within the UK - potentially for the life of the survey. If an OSM moves house, they are followed to their new address and those living with the OSM become eligible for interview as TSMs (see section 4.6.2). If the OSM moves into an institution, just the OSM would be enumerated and interviewed and not other residents of the institution.

The case may arise where the only OSM in the household is a child. Other household members are then TSMs so long as they are co-resident with the child, and therefore eligible for interview, even if the child is not yet old enough to be eligible for interview. The child OSM is an eligible sample member and should be enumerated at each wave, even if they are not (yet) eligible for interview because of their age.

The sample of OSMs is a representative sample of the resident population of the UK (after weighting to correct for differences in selection probabilities) and would remain

[6] TSMs and PSMs can be included in cross-sectional analyses using cross-sectional weights that have been derived using a model-based assumption to estimate inclusion probabilities.

so over time if the population were 'closed'. However, new immigrants to country since the time of initial sample selection have no possibility to become an OSM.

4.6.2 TEMPORARY SAMPLE MEMBERS (TSMS)

At each survey wave, all members of the household of an OSM who are not themselves OSMs are designated TSMs. In the GPS there are therefore no TSMs at Wave 1, but TSMs appear from Wave 2 onwards as household composition begins to change. Any child born after Wave 1 to an OSM father, but whose mother is not an OSM, will be a TSM if they are observed to be co-resident with the father (or any other OSM) at the survey wave following the child's birth. TSMs remain eligible for enumeration and interview as long as they are resident in a household that includes at least one OSM or PSM (see below). When a TSM is no longer co-resident with an OSM or PSM, they are not followed and become ineligible for interview. TSMs are identified as "re-joiners" if they are subsequently found in an OSM/PSM household and then become eligible for interview again.

In the EMBS, any members of an enumerated household eligible for inclusion at Wave 1 who are not from a qualifying ethnic minority are TSMs at Wave 1. This was the only category of TSM at Wave 1.

4.6.3 PERMANENT SAMPLE MEMBERS (PSMS)

In certain circumstances the status of someone who would otherwise be a TSM may change to PSM, indicating that attempts should continue to be made to enumerate and interview them, even if and when they no longer live with an OSM. This change of status is made for substantive research reasons because of the additional contextual information these people may provide for the analysis of OSMs. At present, there is only one category of PSM, though in principle others could be defined in the future. Any TSM father of an OSM child born after Wave 1 and observed to be co-resident with the child at the survey wave following the child's birth becomes a PSM. Note that some PSMs will have been enumerated at previous waves with the status of TSM, while others will be enumerated for the first time as a PSM. PSMs remain potentially eligible for enumeration and interview for the life of survey.

5. INSTRUMENT DESIGN

5.1 PRINCIPLES AND PROCEDURES

The design of the *Understanding Society* questionnaires was influenced by a number of distinctive features and opportunities of longitudinal research design. The most distinctive feature of longitudinal research is that it allows an analysis of units - here: individuals - over time. In comparison, research on successive cross-sectional surveys focuses on change at the population level rather than individual-level dynamics. The research instrument is designed to allow the construction of continuous measures, of, for example, income, employment histories and labour market participation, household structure and residential mobility over the life course. This information is collected much more reliably in prospective annual panels than in long term retrospective history surveys. Hence, many questions in *Understanding Society* are concerned with events in the twelve months between interviews, rather than with the current situation at the time of interview. A further use of panel data collection is to compare expectations about change in the subsequent year with change that actually takes place. Important areas in this context are occupational change and residential mobility. For research of this kind, it is important that expectations are measured before the subsequent relevant events have taken place, to avoid post-hoc rationalisation and the contamination of memory. A panel design is therefore necessary rather than relying on retrospective recall.

A second important design feature is the collection of data from and about all household members at each wave. This provides detailed information on the household context of each sample member. This is important for many substantive areas of research in which dependencies between household members are important (e.g., income, poverty, material wellbeing, social capital, attitudes and values, genetics, etc.).

Given its breadth and scope, a great number of questions were specifically developed for *Understanding Society*. However, many were sourced from the BHPS and, to a lesser extent, from national and international longitudinal studies such as the Millennium Cohort Study (MCS), Family and Children Study (FACS), English Longitudinal Study of Ageing (ELSA), Health and Retirement Survey (HRS), German Socio-economic Panel Survey (SOEP), and the Household, Income and Labour Dynamics in Australia (HILDA). Some questions were sourced from cross-sectional surveys such as the Health Survey for England (HSE), Integrated Household Survey (IHS), Labour Force Survey (LFS), Workplace Employment Relations Study (WERS), Skills Survey, Citizenship Survey, while some are standard measures such as the SF-12 instrument. Inclusion of questions from other studies facilitates comparative research and triangulation of research findings, potentially adding more depth in areas where *Understanding Society* necessarily provides less detail. All questions included have a clear focus on characteristics, behaviour or values that are either expected to be subject to change, or are significant factors affecting the likelihood of change. Some questions such as the name of the school the children in

the sample attend or car registration details are included because they permit linkage to external data from public administrations.

5.1.1 CONTINUITY OF MEASUREMENT

It is a basic principle of *Understanding Society* to maximise continuity of measurement. This means that, unless there is a very strong case for change,

a) questions that are repeated on different waves are asked in an identical form, with identical routing, and in as similar a context as possible;
b) each survey instrument remains fixed for the entire fieldwork period of a wave of data collection.

However, under certain circumstances a strong case for change has been made and changes have been made either to the questionnaire within a wave or to questions between waves. Examples of these circumstances have been:

- at the end of the first six months of data collection in Wave 1, some questions were dropped for budgetary reasons after it was found that the interview was taking longer than predicted;
- questions about state benefits (welfare payments) have had to be altered when new benefits were introduced or names of benefits were changed;
- routing for certain questions was changed between waves when a routing error became apparent only after data collection for a particular wave was largely completed.

The switch in mode for the self-completion questions from paper to Computer Aided Self-interview (CASI) at Wave 3 meant that for some questions the response options were presented differently between waves. For example, where response options were arrayed horizontally in the paper self-completion, they were presented vertically in CASI. There is some evidence that this change in the way in which the response options were presented may have affected how some people responded to the question (Burton 2012).

All such changes have been documented in the variable view of the online documentation system.[7] Also see section 5.2.2 below.

5.1.2 TRANSLATION INTO MULTIPLE LANGUAGES

The survey materials are translated into multiple languages to reflect the multilingual composition of the sample and to minimise the extent of non-response due to language barriers. The chosen languages for translation were Welsh (required under the Welsh Language Act), Bengali and Punjabi in Urdu and Gurmukhi scripts, Arabic, Somali, Cantonese, Urdu and Gujarati.

Translated instruments were developed using independent translator and checker, and adjudication. Translated interviews were carried out using the translation by either accredited bilingual interviewers or by accredited translators alongside the interviewer. They were implemented in CAPI software, obviating the need for separate, paper-based versions of the translation, which are associated with much longer interview times. This meant that bilingual interviewers could simply switch to the language of choice (or between that and English), while non-bilingual

[7] See https://www.understandingsociety.ac.uk/documentation.

interviewers could work alongside a translator, helping them to enter the response directly into the CAPI laptop. A translation pilot was carried out on selected languages and found that this process largely worked well. At Wave 1, out of 50,994 interviews, 456 translated individual (adult) questionnaires were used (i.e., around 0.8%).

Where respondents spoke languages not available in these translations, practice resorted to the use of bilingual interviewers or translators or household members translating 'on the fly', for the respondent.

For further information about translations, read the designated section about this in the *Understanding Society Ethnicity Research User Guide*, see McFall, Nandi and Platt (2014). Translated instruments can also be obtained from the Study team.[8]

5.1.3 CONSULTATION OVER RESEARCH PRIORITIES AND CONTENT

As a longitudinal panel study, maintaining continuity of measures is a central component to allow longitudinal measurement of change and stability. Nevertheless, it is important to reflect emerging research agendas in *Understanding Society* content where we are able to do so. Given this, it is important to consult the user community periodically to ensure the questionnaire reflects scientific and policy imperatives.

The initial focus of the *Understanding Society* study content lay on

1. Standard of living measures
2. Family, social networks and interactions, local context, social support, technology and social contacts
3. Education, human capital and work
4. Lifestyle, social, political, religious and other participation, identity and related practices, dimensions of life satisfaction/happiness
5. Attitudes and behaviours related to environmental issues
6. Health outcomes and health-related behaviour
7. Psychological attributes, cognitive abilities and behaviour
8. Preferences, beliefs, attitudes, expectations
9. Illicit and risky behaviours (crime, drug use, antisocial behaviours etc.)
10. Initial conditions, life history data

There were a number of cross-cutting design and content considerations, around

11. Ethnicity research
12. Linkage to information from other sources[9]
13. Continuity with the BHPS. See Laurie (2010) for a summary of decisions taken with respect to balancing continuity of observation and incorporation of innovation.

[8] The email address is info@understandingsociety.ac.uk.

[9] Data linkage to routine administrative records is an area in which we consult widely and on an ongoing basis with key stakeholders. We share expertise across the longitudinal studies represented in the Cohort & Longitudinal Studies Enhancement Resources (CLOSER) to help implement data linkages, develop data sharing protocols and champion data linkage further in government and society. We inform about planned and ongoing data linkage projects via a designated website. See https://www.understandingsociety.ac.uk/about/data-linkage.

A process of extensive consultation was carried out within the British academic and policy research communities over the data requirements for each area, priority measures critical for longitudinal analysis, and the appropriate balance between the different substantive areas within the questionnaire.

The consultation process culminated in a consultation conference held at the Royal Statistical Society in June, 2007. The purpose of the conference was to review and discuss the conclusions and recommendations of all of the advisory groups and make final recommendations for the content and design of the *Understanding Society* questionnaires. Following the conference, the design work on the questionnaires began, taking forward the recommendations received. A number of designated scientific advisory committees were established who met regularly to oversee the content and conduct of the panel.

In 2013, as part of plans to implement Waves 6-8 of the Study, a further consultation with users on the content of the Study was conducted. Following input from a wide range of Government, third sector and academic users of the data, and discussions with the Study's SAC and Governing Board, five priority topic areas were agreed:

(1) income, wealth, consumption and expenditure,
(2) health wellbeing and health behaviours,
(3) employment,
(4) education,
(5) family.

Over the next few years the content of the Study in these areas will be reviewed and improved where appropriate. Topic Champions, experts in these priority areas of the Study, will be designated to consult academic and policy users in their areas of expertise to ensure content – through the questionnaire, data linkage and new forms of data collection - in each area keeps abreast of emerging agendas as well as effectively addressing key longitudinal research questions as new data sources and techniques make new approaches possible. We also plan to expand the kinds of data we collect and alter the frequency of data collection for these key areas. This may mean reducing content in other topic areas.

5.1.4 RESEARCH-INFORMED INSTRUMENT DEVELOPMENT

Alongside our wide consultations over content, we have undertaken a number of methodological studies to develop new questionnaire content (such as measures of ethnic identity Nandi and Platt (2009), to road-test modules that have not previously been implemented in a longitudinal household panel study context (such as the cognitive functioning measures in Wave 3, see Gray et al. (2011)), or to evaluate the effects on participation and/ or measurement of different data collection options that have been identified as priorities, see, e.g., Vannieuwenhuyze and Lynn (2014) and Lynn, Uhrig and Burton (2010).

Extensive research also went into the design and implementation of the *Understanding Society* Nurse Health Assessment (McFall et al. 2012; McFall, Conolly and Burton 2012), and into the design and longitudinal implementation of data linkage (Knies and Burton 2014; Knies, Burton and Sala 2012; Sala, Burton and Knies 2012; Sala, Knies and Burton 2014). A programme of research is currently being developed on how best to improve capturing household finance information and we are investigating ways of improving information families when they separate.

In 2015, we also started to review the questionnaire to reflect the decision to move to a mixed mode (face-to-face and web) interview. Some types of questions are not answered in a comparable way across modes and we will need to consider how best to address this.

5.2 SUMMARY OF INSTRUMENTS

Reflecting the interdisciplinary and multipurpose nature of *Understanding Society*, data are collected using a wide range of data collection instruments. Data collection already starts before the first interview is being conducted, and in addition to the core instruments described below we record paradata such as date the interview took place and time stamps for when a module was started.

An Address Record Form (ARF) is filled in by the interviewer at the time of calling at a sampled address. It contains observations about the type of neighbourhood and accommodation which assist in constructing unit non-response adjustments but may also be used as contexts in substantive analyses. The ARF also contains the call records.

A household coversheet is usually administered at the interviewer's first contact with an adult member of the household. It involves a listing of all household members together with brief summary data regarding sex, date of birth, marital status, employment status and relationship of household members to each other. It takes 5 minutes to complete on average.

A household questionnaire is administered to an adult in the household who is either the owner or main renter of the residential address. In the BHPS we referred to this person as the Household Reference Person (HRP) and a number of questions were worded with reference to this person, but this concept does not exist in *Understanding Society*. The household questionnaire takes 10 minutes to complete on average. It contains questions about the accommodation and tenure and some household level measures of consumption.

The key instrument is the individual questionnaire. It is administered to each member of the household aged 16 or over. Questions are arranged in modules that cover a broad range of topics, and it takes around 33 minutes to complete on average. There are also some questions about the interview and respondent for the interviewer to answer.

Respondents who participate in a face-to-face interview are also administered an adult self-completion questionnaire. It contains questions that are subjective or particularly vulnerable to the influence of other people's presence during completion, and potentially sensitive questions requiring additional privacy. In Waves 1 and 2 this was administered in a Paper Aided Personal Interview (PAPI) (and counted as a single topic module), and it is now administered in CASI. It takes around 7 minutes to complete on average.

Materials to link administrative records held by public administrations (e.g., in education, health, and employment) are administered at the end of the face-to-face interview with adults so as to not interrupt the flow of the interview. All data linkage requests include information leaflets detailing what is involved and interviewers are there to answer any additional questions the respondents may have. Some

administrations require consent in writing; in that case there are consent forms which need to be signed and a copy of which is kept by the respondent for future reference.

Sometimes it is difficult to conduct interviews with all members of the household. A 'proxy schedule' is used to collect information about household members absent throughout the field period or too old or infirm to complete the interview themselves. It is administered to another member of the household, with preference shown for the spouse or adult child. The questionnaire is a shortened version of the individual questionnaire, collecting some demographic, health, and employment details, as well as a summary income measure. It takes around 10 minutes to complete on average.

We also have a telephone questionnaire that is based upon the standard interview schedule with minor adjustments to reflect the interview mode (i.e., there are no showcards). This questionnaire is administered when all efforts to achieve a face-to-face interview have failed.

For our youngest panel members eligible for interview we use a youth self-completion questionnaire. This is administered to each person aged 10 to15 in a PAPI that is particularly designed to be enjoyed by young people and is typically completed in private whilst personal interviews with another member of the household are being undertaken. It takes 10 minutes to complete on average.

In Wave 2 and Wave 3 the survey instruments included a to date one-off nurse-administered interview, which collected a range of physical measures such as blood pressure, height, lung function, and, with consent, the collection of a blood sample for future analyses and DNA extraction. The procedures used for the Nurse Health Assessment are documented in the user guide for that component of the Study (McFall et al. 2014)[10]. Subsequently, blood samples were analysed to produce a range of blood analytes, which are outlined in a separate user guide (Benzeval et al. 2014)[11].

5.2.1 QUESTIONNAIRE STRUCTURE

The interview for each adult lasts 33 minutes on average, with an additional short household level questionnaire for one individual in the household. The design includes three main components:

- core questions repeated at each wave;
- rotating core questions repeated on a two to three year cycle;
- variable component questions.

Questions are arranged in topic modules and cover, among others, individual demographics, education and training, health and caring, current employment and earnings, values and opinions, environmental behaviours, transport, and parenting. About half of the questionnaire content is collected annually, with additional modules collected at different intervals, often every two to three years. Overall, the Wave 1 questionnaire included 36 topic modules, and the number has increased steadily to 67 in Wave 5.

The rotating core items covered topics where there was no expectation of rapid change and there is therefore no need to ask questions on them every year. This

[10] Available online at https://www.understandingsociety.ac.uk/documentation/health-assessment.
[11] See previous footnote.

also allowed a means to deal with competing demands for limited space within the questionnaire. The remainder at each wave included the variable component. The variable component was designed for:

- Questions which needed to be asked less frequently than core or rotating core items;
- New questions engendered by changing policy and research issues;
- Questions to elicit retrospective data on panel members' life history before the first interview.

The *Long-term Content Plan* summarizes the pattern that has been collected or planned.[12]

5.2.2 QUESTIONNAIRE CONTENT

In the following paragraphs we highlight the content for the first five waves of the Study.

Wave 1 collected important baseline data. Some Wave 1 measures are stable, that is, not time variant. In subsequent waves, we try to collect this type of information from individuals who are new entrants to the Study. These represent the strongest areas for examining annual change. See for example disability, caring, employment-related information, childcare, politics and income and benefits. These have been the focus of major longitudinal research in the BHPS and should also be a prominent focus with *Understanding Society*.

Wave 1 also saw the first rotating modules such as on parents and children, family networks, and environmental behaviour. Within the ethnicity strand – that is within the "Extra 5 minutes"-questions, there are modules about remittances, harassment, and discrimination. The module on parents and children has content about attitudes and behaviours related to education, activities and interaction with children, and parenting practices.

In Wave 2 and subsequent waves, the annual event history module is asked of persons previously interviewed. It asks about changing circumstances related to moves, marital status or cohabitation, new children including childbirth and pregnancy, new health conditions, educational experiences, and employment changes.

Wave 2 saw the introduction of a set of rotating modules related to health behaviours (nutrition, smoking, physical activity). There are modules about voluntary work and charitable giving and important modules about savings and personal pensions. Retirement planning is an age-triggered module that is taken up again in Wave 3. Within the self-completion questionnaire, there is content on alcohol consumption, dimensions of identity and gender role attitudes.

The "Extra 5 minutes"-questions have modules on political engagement and ethnic identity. Wave 2 also has a major module on work conditions that encompasses such topics as payment mechanisms, unions, pensions, work times, autonomy and security, and work stress.

[12] The long-term content plan can be found at https://www.understandingsociety.ac.uk/documentation/mainstage/long-term-content-plan.

There are multiple new modules for Wave 3 including those on local neighbourhoods, content on social networks in the Main Survey and the self-completion questionnaires, groups and organizations, use of news and media, and political self-efficacy. There is a major module on cognitive ability, see McFall (2013) for more detail about the concepts and measures of cognitive ability.

Important data related to family ties can be found in the parents and children and family networks modules, both of which are repeated from Wave 1. There is also data about child maintenance payments and relationships with children who do not live in the household. The self-completion modules have parents' reports on children including a version of the Strengths and Difficulties Questionnaire (SDQ), and parenting styles. The self-completion modules also include a Big 5 personality measure, sexual orientation, and several modules of questions for young adults which bring questions from the youth questionnaire into the 16-21 age group.

The "Extra 5 minutes"-questions have repeat modules on discrimination and harassment and a new module on Britishness.

Most modules for Wave 4 appeared in earlier waves. They include major modules on work conditions (covering, e.g., transport behaviour and job satisfaction) and modules on environmental behaviours and voluntary work. Wave 4 carried for the first time rotating modules on wealth and assets, financial attitudes and behaviours, and credit and debt. Another highlight of the Wave 4 questionnaire is a one-off module on leisure participation focussing on the Olympics 2012, which were held in and around London 27th July - 12th August 2012. The Wave 4 self-completion instrument for adults includes a focus on mental health and wellbeing and gender role attitudes.

The "Extra 5 minutes"-questions have repeat modules on remittances and on ethnic identity and Britishness.

Wave 5 repeated existing annual and repeating modules. In addition, there was a new module around cultural participation in the "Extra 5 minutes"-questions, and the Olympic module was carried for those interviewed between the start of Wave 5 and 26th July.

The adult self-completion included new sets of questions asking about delayed self-gratification, identity and self-efficacy. There was also a set of questions for those participants in Scotland, which asked about the Scottish Referendum. Those adults issued to the second year of Wave 5 (2014) and interviewed before 22nd August were asked whether they would be voting in the referendum, and if so how they were planning to vote. Those who were interviewed after 18th September were asked whether they had voted in the referendum, and if so, how they had voted. In addition, participants were asked about mode preference; asking how likely they would be to participate in an online survey.

5.2.2.1 Multi-item scales in the questionnaires

The *Understanding Society* questionnaire includes a number of validated and widely used scales developed to measure constructs such as personality, health, material deprivation and neighbourhood cohesion.

Table 1 lists the multi-item scales and provides a brief note on any adaptions undertaken in *Understanding Society*. Whilst these scales are widely used, unless

otherwise stated, to our knowledge no analyses of construct validity and reliability have as yet been undertaken in the *Understanding Society* context.

5.2.3 QUESTION NAMING CONVENTIONS

When specifying the survey instruments, it is necessary to assign a short name (typically no more than 10 digits) for each question asked so that data returned from the fieldwork can be merged into the existing survey database. In *Understanding Society*, we assign a name to a question when it is first designed and continue use that name in later rounds of data collection so long as the underlying question does not change substantially.[13] Question names directly correspond to variables in the data files made available to analysts. It was therefore paramount to follow some naming conventions that would make data analysis as straightforward as possible.

The most dominant principle for naming new study content was that the question names are mnemonic, i.e., the name attempts to give some information as to the content of the information. For example, information on the respondent's mother's educational qualifications is named "maedqf" with 'ma' for mother, 'ed' for education and 'qf' for qualifications. This variable naming convention contrasts, for instance, with that chosen by the German SOEP where the question name corresponds to the position of the question in the questionnaire. Whilst this makes it easy to find the question in the questionnaire, a typical task for users is to identify identical questions in different waves of the Study and assign a consistent variable name.

Questions that came from the BHPS are usually named the same as before (Nb. The BHPS also used mnemonic names), for the convenience of analysts, but this has not always been possible given overruling naming conventions in *Understanding Society*. In particular, to ease identification of groups of variables a number of additional general naming conventions have been applied. For instance, all information first collected in the PAPI self-completion interview with adults in Waves 1 and 2 starts with the prefix "sc"; information from the interview with youths starts with the prefix "yp". Similarly, we have attempted to include in the variable name the acronym of well-known instruments such as the SDQ or the General Health Questionnaire (GHQ). Pointers to other members in the household typically end on "pno" or "pid". The prefix "ff_" following the wave prefix shows variables that were fed forward from previous waves to route respondents appropriately in the script.

The aforementioned conventions have not been systematically applied for new study content from Wave 3 onwards, and there were no specific conventions for marking out other relevant content such as the "Extra 5 minutes"-questions. Given the link between question names assigned in the instruments and variables in the data, and given fieldwork for different waves of data collection it is difficult to change question names post-hoc.

[13] In the process of data preparation for release we also add a wave prefix to unambiguously define the temporal nature of the data.

6. DATA COLLECTION

6.1 FIELDWORK PROCEDURES

The sample for each wave of *Understanding Society* is issued to field as 24 monthly samples, with the exception of addresses in Northern Ireland and the former BHPS sample, each of which are restricted to the first 12 months of the wave. Table 2 shows the timing of the sample issue for Waves 1-4.

Most data collection takes place face-to-face via computer aided personal interview (CAPI). There are also self-completion instruments for youth and adults. The youth instruments are administered on paper. The adult self-completion questionnaire was administered on paper at Waves 1 and 2 and by CASI at Waves 3, 4 and 5. From Wave 3 onwards, there was also a telephone mop-up at the end of the fieldwork period for each sample month. Telephone interviews were also used instead of face-to-face for a small proportion of BHPS sample households who had previously been interviewed by this mode on BHPS.

A sample of addresses was selected for Wave 1, as described in section 4.1, with no information about the residents at each address. Interviewers mailed an introductory card from ISER to all sampled addresses (addressed to "The Occupier"), together with a small leaflet outlining the purpose of the survey. Then the interviewer visited within a week of the mailing to identify and interview the residents. The interviewer's task at Wave 1 was to identify all usual residents, and hence the sample members, to introduce the survey, to carry out household enumeration, household and individual interviews, and to encourage completion of self-completion questionnaires.

From Wave 2 onwards, interviewers were issued with a set of households containing OSMs. In most cases the households were as identified at the previous wave, though in some cases updated information had been received since that date. The interviewer's task at each wave is to locate each of the OSMs in this set of households, to identify all other current members of the household of each OSM (by administering the household grid), and to administer household and individual interviews, and to encourage completion of self-completion questionnaires.

At the end of each individual interview, respondents are asked to provide their contact details, including phone numbers and email addresses. They are also asked to provide contact details for at least one other person who would be likely to know where they are in the event of a move (these are referred to as "stable contact details"). Both the personal contact details and the stable contact details are subsequently transferred to the panel maintenance data base (see section 6.5.2) and provide an important means of tracking sample members in the event that an interviewer is unable to locate them in the field at a subsequent wave.

All participating households receive a brochure, giving further information about the survey and thanking respondents for participating.

Interviewers were instructed to make a minimum of six calls at each sampled address before it could be considered a non-contact. These calls had to include at least one at the weekend and at least one on a weekday evening (after 6pm).

Interviewers were encouraged to make further calls, if possible. If there was a potential for success, a special conversion letter was sent to households which had refused to participate or had not been contacted. Post interview quality control was carried out with a telephone recall on 10% of all completed interviews.

Interviewers uploaded their work daily, including information about all the calls they had made, and whether or not there was any response. This information was collated by the survey agency to construct a weekly field progress monitor report for ISER.

A telephone “mop-up” of remaining non-respondents at the end of the fieldwork period for each monthly sample was introduced from the start of Wave 4, but applied also to the Wave 3 year 2 sample (which was in the field at the same time as the Wave 4 year 1 sample). The aim of this mop-up was to contact adults who could not be contacted during the main fieldwork period (in participating households) and adults in households that were non-responding in the main fieldwork period, except those who had adamantly refused or were deemed to be mentally or physically incapable of participating. The trained and briefed telephone interviewers introduce themselves, remind the sample members of the survey, and ask whether they would be able to do the interview by telephone. The purpose of the mop-up was to increase participation among those who were hard to contact in person.

Analysis by NatCen indicates that the telephone mop-up increased the overall household response rate for that period by about three percentage points for the EMBS and by just less than two percentage points for the GPS. This mop-up was not conducted with the BHPS sample in Wave 3 since they are interviewed in the first year of each wave.

Towards the end of Wave 3, September 2012, a trial was conducted in two field areas in which an additional incentive was used at the re-issue stage. This was then rolled-out across the sample from October, and so covers the last quarter of Wave 3. In the implementation, non-responding households were reviewed by NatCen for re-issue and possible re-allocation to a different interviewer. Households which had refused to participate in the initial fieldwork period, but where the assessment was that this was a “soft” refusal, were sent a re-issue letter which mentioned an additional incentive if they participated during the re-issue fieldwork period. Other non-responding households were sent a normal re-issue letter, but the interviewers had discretion to offer the additional incentive on the doorstep if they felt that this would convert a non-responding household to a participating household.

In addition, during the latter quarter of Wave 3 fieldwork, more effort was made to increase interviewer continuity for households across waves, rather than prioritising interviewer efficiency. It is estimated that these two procedures, which were launched almost simultaneously, increased household response rates by around 4 percentage points for the EMBS and by around 2.5 percentage points in the GPS in Quarter 8. The procedures adopted in Wave 3 to maintain household response were continued in the fieldwork for Waves 4 and 5.

6.2 PRE-TESTING AND PILOTING

Prior to the first wave of the Main Survey, two small pilot studies and a dress rehearsal were carried out. A cognitive pilot of 70 individuals was conducted March –

April 2008 to test screening and other questions relevant to the ethnicity strand. A translation pilot was conducted in June 2008: 50 interviews were carried out using Bengali and Punjabi translations of the questionnaire to see if there were problems with the operation of the translation program or problems with interviewing with the translated instruments. A pilot of all data collection instruments and procedures in 100 households, called a dress rehearsal, took place August-September 2008.

A pilot for Wave 2 tested all instruments and data collection procedures. For this wave, the data collection also focused on assessing any problems with integrating members of the former BHPS sample component, which included a small segment conducted by telephone interviews. In all 237 households were issued. Of these, 91 were households interviewed in the Wave 1 pilot. The BHPS sample component was represented by households that were part of the BHPS between 1997 and 2001, the European Community Household Panel (ECHP). See Lynn (2006) for further detail on the ECHP sample in the BHPS. Households for which we had a telephone number were issued to telephone interview to test the telephone interview instruments and procedures. The Wave 2 pilot took place September-October 2009.

From Wave 3 onwards, dress rehearsal pilots took place September-November, and interviewers returned to the same sample that was interviewed at the previous waves' pilots.

6.3 RESPONDENT MAILINGS

Given the heterogeneity of the *Understanding Society* study sample members and of their participation in the Study, there are a number of ways in which we try to engage with them. Some of those methods involve posting printed materials, while others have an electronic form. Communications take place both during the fieldwork period while and in-between their annual interviews. The methods used to communicate with sample members are summarised in this section.

6.3.1 ADVANCE LETTERS

An introductory card from ISER to all sampled addresses (addressed to "The Occupier"), together with a small leaflet outlining the purpose of the survey was sent in advance of the fieldwork for Wave 1. At each subsequent wave, advance letters have been sent to each adult in participating households about a week before the start of the fieldwork for the sample month. These are sent by the fieldwork agency and include multiple variants which differ depending on:

a. the last wave household outcome (response or non- response);
b. last wave individual outcome (response or non- response) ;
c. incentive amount and type (conditional or unconditional) in last wave non-respondent households;
d. whether an individual is a Rising 16 (i.e., young person who was 15 last year and is now eligible for a full adult interview for the first time);
e. whether an individual is a new entrant to the household (though relatively few of these are identified in advance of fieldwork);

In addition to the letter, the advance mailing also includes a change of address card and an unconditional incentive in the form of a gift voucher.

Advance mailing has a number of objectives. The main one is to notify sample members of an upcoming annual interview. In order to maximise the chance of interviewing the individuals each year, the mailing also encourages them to notify the University of Essex about any changes in their address or other contact details. The letter also provides the Study contact details and encourages sample members to send us feedback on our communications with them. Finally, the letter thanks the participants for their continuing support, in this way hopefully motivating them and encouraging their loyalty to the Study.

6.3.2 INTER-WAVE MAILINGS

An inter-wave mailing (IWM) to adult sample members is used to help maintain contact with participants and update addresses between waves. The mailing also aims to report, raise awareness and engage sample members about recent findings and examples of how the Study has been used and its impact. The frequency of these mailings has varied over time. At first, adult sample members received one IWM a year but since the middle of 2013 this strategy changed to three IWMs a year. From 2016 the IWM strategy reverted back to one mailing per year.

The increase in the number of IWMs in 2013 was in response to findings from a qualitative study conducted by NatCen on the factors affecting participation in *Understanding Society*. The previous single IWM used to take place about 6 months after the interview. The three mailings were quarterly with different quarters of the sample receiving the mailing at different times: the quarter that was in the field at the time of the mailing would not receive that particular mailing at all, with the quarter that was about to go to the field receiving the IWM report as part of their advance mailing. So at any time the only sample quarters that received an independent IWM were the two for which that wave's fieldwork was already complete. The timetable of quarterly IWMs was as follows:

April IWM:
- Q3/Q7 (mailed mid-April as their IWM)
- Q4/Q8 (mailed mid-April as their IWM)
- Q2/Q6 (mailed in Apr, May, June as part of their advance mailing)
- Q1/Q5 (not mailed)

July IWM:
- Quarter1/QuarterQ5 (mailed mid-July as their IWM)
- Q4/Q8 (mailed mid-July as their IWM)
- Q3/Q7 (mailed in July, Aug, Sep as part of their advance mailing)
- Q2/Q6 (not mailed)

October IWM:
- Q1/Q5 (mailed mid-Oct as their IWM)
- Q2/Q6 (mailed mid-Oct as their IWM)
- Q4/Q8 (mailed in Oct, Nov, Dec as part of their advance mailing)
- (Q3/Q7 not mailed)

December IWM:
- Q2/Q6 (mailed mid-Dec as their IWM)
- Q3/Q7 (mailed mid-Dec as their IWM)

Q1/Q5 (mailed in Jan, Feb, March as part of their advance mailing)
(Q4/Q8 not mailed)

April, July and October mailings would have a 'bite-size' format of an A5 postcard and focus on various recent findings from the Study, while the December mailing always included a repurposed participant version of 'Insights', which was the more substantial report the sample members would get between their interviews. Examples of the IWM reports can be found on the Study website.[14] These include: 'Tell us a story', 'What's changed in 25 years', 'Make the right move', participants' version of the annual Study findings '*Insights'*, 'Closer to nature', 'Media round up', 'The story so far', 'GDP & Beyond', and ' Poverty'.

Each IWM also includes an address confirmation slip and materials to encourage registration with the participant website.[15]

From Wave 3 onwards, IWMs also included reminders of any data linkage consents the respondent had given. This gives the respondent the opportunity to review whether they are still happy for us to combine their survey responses with their administrative records for statistical analyses.

6.3.3 PARTICIPANT E-NEWSLETTER

To re-inforce the postal mailings, we also send out a quarterly e-newsletter which is sent to just under 29,000 adult sample members for whom we have an email address. The objectives of these email contacts include updating/signposting sample members to the IWM information, summarising other recent and relevant news and content on the website, providing additional contact with sample members between mailings and encouraging them to visit the participant website. The aim is to aid ongoing location of sample members, as well as to encourage co-operation at the next wave.

6.4 INCENTIVES

As a token of thanks, respondents receive High Street vouchers. There are two levels of incentive: £10 per adult, and £5 for 10-15 year olds. In most cases incentives for adults are unconditional and sent with advance letters. There are some cases where interviewers hand out incentives:

- When handing out the youth paper self-completion questionnaire,
- After an individual interview if the respondent has not received the advance letter and incentive
- After an individual interview if the respondent is a new sample entrant (see sections 4.6.2 and 4.6.3) and was not sent the advance mailing.

[14] See https://www.understandingsociety.ac.uk/participants/features
[15] The participant website can be seen at https://www.understandingsociety.ac.uk/participants

6.5 SAMPLE MANAGEMENT

6.5.1 PANEL MEMBERSHIP

The rules for following – i.e., attempting to collect data from/about – individuals over time are based upon the concept of sample status, as described in section 4.6. The sample potentially eligible for interview at any given wave includes all OSMs and PSMs who are still alive and living in a household in the UK. However, only a sub-set of this potentially eligible sample will be attempted or found at any given wave. From Wave 2 onwards, sample members who were known to have died or left the UK, whole households untraced at the previous wave, those in prison and those who had adamantly refused to take part at the previous wave were not issued to field. Other than those who had died, these sample members remained potentially eligible for inclusion in the survey and were interviewed if further information about their whereabouts became available (those who had been untraced), they moved back to the UK (those who had moved abroad), or they decided to take part in the survey despite a previous adamant refusal.

From Wave 4 onwards, sample members in households which had been classified as a whole household refusal in the previous two waves, households which were non-contact in the previous two waves, and households which had refused two waves earlier and been a non-contact at the previous wave were also not issued to interviewers. These households were issued to the survey agency as 'dormant' cases, along with prior-wave untraced movers. If a new address were obtained, or if a request to participate were received from sample members in these households, the fieldwork agency would 'activate' the case and issue it to an interviewer.

Furthermore, when the data for each sample quarter are delivered to ISER, all comments entered by interviewers are read and reviewed. These include reasons given for a non-responding case (household or individual). Where it is clear that the reason will be permanent (e.g., an adamant refusal, the on-set of dementia), the case is withdrawn. Where this is a household-level comment (e.g., adamant refusal), the household is coded appropriately and withdrawn. Where the comment only relates to a specific individual amongst others in the household, they are excluded from the active sample. The name and basic information of this person is still fed-forward at the next wave to be included in the enumeration grid, but they are flagged (ff_exclude = 1). This means that they are not sent an advance letter, invitation email or reminders, and that interviewers are not expected to approach them for an interview.

Prior to issuing the sample at each wave, sample members were categorised as belonging to either the active or inactive sample as follows.

6.5.1.1 Active sample

- the issued sample - all expected sample members at a given wave. This category includes all members of interviewed households from the previous wave as well as members of non-interviewed households being attempted again. Full details fed-forward and issued to field.
- the inaccessible sample - includes untraced movers from earlier waves, sample members who are living in an institution where they cannot be interviewed (prison), sample members who are too ill or elderly to be interviewed and no proxy possible, and sample members who are out-of-

scope (outside UK). Details are fed-forward, not issued to field, but are available to interviewers if required.
- the retiring sample - includes sample members who refused to take part any longer in the panel at the previous wave and are to be withdrawn from the active sample at the following wave. Details are fed-forward, not issued to field, but are available to interviewers if required.

6.5.1.2 Inactive sample

- the retired sample - includes the deceased, adamant refusals from earlier waves, TSMs no longer living with an OSM so ineligible for inclusion and untraced movers. The details of these sample members are not fed-forward but are archived and not available during fieldwork for interviewers even though they could be restored to the sample if required. It should be noted that at each wave some members of the active sample turned out to be ineligible for interview (e.g., because they have moved into an institution or out of UK), while some persons not in the active sample turned out to be eligible for interview (as OSMs or TSMs).

This method of managing the sample allowed details of respondents to be made available to interviewers as required while providing a systematic means of removing sample members from the active sample when required. Periodic checking of the inactive sample is carried out to establish their current status e.g., checking death registers, checking the whereabouts of out-of-scope members through relevant contacts (e.g., the 'stable contacts' provided in an earlier interview or ex household members, who may themselves still be active sample members)..

6.5.2 PANEL MAINTENANCE

A custom designed Panel Maintenance Database (PMDB) is used by ISER to maintain accurate address records and other between-wave information for panel members. The PMDB is maintained as a database of names and addresses of sample members held separately from the survey database containing the interview data for reasons of confidentiality and to comply with the UK Data Protection Act. The PMDB contains contact information collected during the survey interviews and is updated in the year between interview points if notification of a death or change of address is received. It includes indicators of new addresses, household splits and moves out of the country or into an institution. At each wave, sample members are issued to the most recently known address that may or may not be the address of interview at the previous wave. Maintaining contact with respondents is facilitated by:

- Providing a named contact person amongst the survey agency's field staff, a Freephone number (with answerphone), a Freepost mailing address, an email address and web-site for sample members
- The use of an IWM (see section 6.3.2), consisting of a letter, a short report of findings or examples of impact, and a change-of-address card with a Freepost return envelope
- Sending an email "Participant Update" four times a year to those sample members for whom we have an email address
- Sending a £5 gift voucher incentive to any person returning a change of address card between interview points
- Updating address details between interview points

- Maintenance of an historical record of all addresses ever occupied for each sample member

However, despite best efforts, there are still people who change their addresses and cannot be traced. When an interviewer identifies that someone has moved from the issued address, they start the tracing process. The first point of information would be the current residents, who may have a forwarding address. The interviewer also has additional contact information for the mover, such as telephone numbers (mobile, work numbers). We also collect and feed-forward stable contact information; this is the name and contact details of someone who is likely to know where the sample member has moved in case we cannot contact them. The interviewer will try to contact the stable contact, either in person (if they are local) or by telephone. Where we have an address for the stable contact, the interviewer can also send them a specific "stable contact letter". Interviewers can also try to talk to neighbours, to see if any of them have forwarding information for the mover. If the interviewer is not able to track the person in the field, the case is returned to the fieldwork agency. Each week a file of untraced movers is transferred to ISER, where cases are reviewed and securely exported to an online facility which checks name and previous address details against a set of databases, such as the Electoral Roll and the National Deceased Register. Untraced cases are then passed to members of staff who attempt to further contact the sample member and/or the stable contact by telephone and/or email. These members of staff work evenings and weekends and are able to make more contact attempts than would be efficient for interviewers. If ISER staff are able to trace a mover, and the fieldwork period for the sample is still open, the information is passed back to the survey agency who then re-issue it to an appropriate interviewer.

6.6 FIELDWORK DATES

The dates when interviewing was carried out with each quarterly sample at each wave are summarised in Table 3. It can be seen that interviewing usually began within a few days of the start of the calendar quarter and was usually 90% complete by the end of the quarter or very soon after. However, the final 10% of interviewing always took at least two further months to complete and sometimes the last interview was not completed until as much as eight months after the end of the quarter (though such cases were rare).

6.7 INTERVIEWERS

Understanding Society has aimed to use interviewers of above average levels of experience and ability because of the demanding nature of *Understanding Society*. The majority of interviewers in Northern Ireland had worked on the BHPS Northern Ireland component (the Northern Ireland Household Panel Survey), and were familiar with the design and operation of *Understanding Society*. Interviewers working on *Understanding Society* are required to have experience working on random probability projects. So in addition to the standard training that interviewers go through, provided by the survey agency, they are required to have successfully completed at least one random probability interviewing assignment. Interviewers are accompanied by supervisors when working at least twice each year as part of their performance management process.

In addition to general interviewer training, interviewers working on the Study attended a one day survey-specific briefing. Generally around 12-20 interviewers attended each briefing, along with two or three briefing managers or area managers. The briefings were led by at least one researcher from NatCen with the majority also attended by ISER staff. The briefings at Wave 1 took place across the UK: Belfast, Birmingham, Brentwood, Bristol, Derby, Edinburgh, Glasgow, Leeds, London and Manchester. Similar topics and locations were used for the Wave 2 briefings. At Wave 3, the Edinburgh briefing was dropped and two briefings were held in Glasgow. Additional briefings were added in Bury St. Edmonds, Liverpool and Gateshead.

The morning sessions were devoted to fieldwork procedures, for example the administrative forms to record contact information, and how to deal with the complexities of multiple dwelling units and multiple households. The afternoon was spent discussing the survey content and reviewing and working with the Blaise CAPI instrument. At Wave 3 there were two types of briefing; for interviewers experienced with the Study or for interviewers with experience who were new to the Study. The latter briefing went into more detail about the background of *Understanding Society*, early findings, the more technical details of the sample, and the task of enumerating the household.

At Wave 4, the style of briefing changed in Great Britain. Many interviewers had worked on the Study for three waves and were familiar with the mechanics of how to conduct the survey. For such interviewers, in Wave 4, the focus of the briefing switched from the survey procedures to motivating the interviewers and giving them information to enable them to motivate the sample members when making contact. Interviewers who were new to the survey still attended a standard briefing, as did interviewers in Northern Ireland. These standard briefings were held in Belfast (three times) and London (once).

Experienced interviewers attended ‘conference-style’ briefings. These briefings were much larger than standard briefings, with 150-250 interviewers attending each event. There were three such events held prior to the start of Wave 4; in Birmingham, Liverpool and London. During the breaks in day, there were stalls and displays of media coverage, research findings, and information about the Study, a Twitter stand and an area where interviewers could write questions on post-it notes for discussion later in the day. The content of the briefing consisted of ‘plenary’ sessions where an overview of progress on the Study “so far” was presented, along with researchers talking about how they used *Understanding Society* in their research, videos of the directors of NatCen and ISER were shown, and medals awarded to interviewers who had achieved 100% response rate in any of their allocations at Wave 3. Once during the morning, and once in the afternoon, there were a number of ‘break-out’ sessions with small groups of interviewers to share best practice and experience of (i) contact and co-operation and (ii) how to deal with household splits and allocating outcome codes. Overviews of these break-out sessions were then discussed at the plenary sessions.

The Wave 5 briefings were also conducted as ‘conference-style’ briefings, with four such briefings; two in London, one each in Glasgow and Liverpool. In addition, there were three standard briefings in Northern Ireland (all in Belfast). The structure was similar to that in Wave 4, with an introduction from the Chief Executive of NatCen followed by presentations covering what was new for Wave 5, results of qualitative

research carried with on *Understanding Society* sample members, former-sample members and some interviewers. Results from a recent IP were shared, with information on how this has impacted on the design of Wave 5. Also, an example of quantitative research, looking at religion using *Understanding Society*, was presented by a NatCen researcher.

Once interviewers are working on the Study, their work is validated twice a year – more frequently for less-experienced interviewers. Around 15% of assignments are selected for validation, with all addresses in the assignment attempted. The validation is generally carried out by telephone and involves asking about the interview process (e.g., time, date, length of the interview) and re-asking questions from the interview to verify the answers.

Interviewers are assigned to specific areas. For Wave 1, 911 interviewers were employed to cover 3,517 areas in the sample. The number of interviewers briefed in Wave 2 was 819 and 746 at Wave 3. At Wave 4, 692 interviewers worked on the Study. At Wave 5, 570 interviewers worked on the Study. There has been a considerable degree of interviewer continuity from wave to wave, from the perspective of sample members. For each pair of consecutive waves, a majority of sample members (out of those who were issued to field at both waves) have been issued to the same interviewer at each wave, and the proportion has increased steadily between Waves 2 and 5, ranging from 54.3% at Wave 2 to 81.9% at Wave 5 (Table 4).

7. PARTICIPATION

This section reports the extent to which the fieldwork operation has been successful in obtaining respondent participation in the survey. At each wave the objective has been to obtain a completed household grid and household questionnaire for each household containing at least one OSM, plus an individual interview with each person aged 16 or over in each such household and a youth self-completion questionnaire for each young person aged 10 to 15. Outcomes are therefore reported separately for each of these survey instruments, and also for the main sample components: GPS, EMBS and BHPS.

7.1 FIELD OUTCOMES AND UNIT RESPONSE RATES

7.1.1 WAVE 1

At Wave 1, 57.6% of eligible households in the GPS participated in the Study (Table 7). In the EMBS it is harder to estimate the household response rate as it is not known how many households were eligible (contained ethnic minority members) amongst those that did not respond to the screening questions. However, if we assume that the proportion of eligible households was similar amongst the households who responded and did not respond to the screen, then we obtain an estimate of around 50.4% (derived as 94.1% response to the screen (Table 6) multiplied by 53.6% response amongst households identified by the screen as eligible (Table 7)). Explicit refusals accounted for almost four in five of the non-responding GPS households, whereas non-contacts were relatively more prevalent amongst the EMBS (Table 7).

Interview data was obtained for 87.0% of all persons aged 16 or over enumerated in Wave 1 households in the GPS, and for 79.2% in the EMBS (Table 8). For both samples, the vast majority of these provided personal interviews, while for a minority who were unavailable or unable to be interviewed, proxy interviews were carried out (6.0% of GPS interviews and 8.7% of EMBS interviews). Of all those who provided a personal interview, 84.9% also completed the adult self-completion supplement, a proportion that was substantially higher in the GPS (87.4%) than in the EMBS (69.6%) (Table 9). Amongst young persons aged 10-15 in enumerated households, 77.0% completed the youth self-completion questionnaire in GPS, as did 62.9% in the EMBS (Table 10).

7.1.2 WAVE 2

At Wave 2, 77.3% of eligible GPS households, 78.2% of BHPS households and 68.0% of EMBS households participated (Table 12). As was the case at Wave 1, the non-contact rate was higher for the EMBS than for the other two sample components, while refusal rates were similar across the samples. Interview data was successfully obtained for 71.7% of persons aged 16 or over enumerated in the GPS (Table 13), including 79.3% of those who had been interviewed at Wave 1 (Table 16). Equivalent interview response rates were 72.7% for the BHPS sample (Table 14) and 60.0% for the EMBS (Table 15), including 69.4% of those who had been

interviewed at Wave 1 (Table 17). Of all those completing a Wave 2 individual interview, 87.0% also completed the self-completion questionnaire (Table 18), including 88.7% in the GPS (and 91.5% of those who had completed the self-completion questionnaire at Wave 1 - Table 19), 89.4% in the BHPS sample, and 69.8% in the EMBS (and 79.0% of those who had completed the self-completion questionnaire at Wave 1 - Table 20). The youth self-completion questionnaire was completed by 78.3% of eligible young persons in the GPS, 82.0% in the BHPS and 63.5% in the EMBS (Table 21).

7.1.3 WAVE 3

At Wave 3, 75.5% of eligible GPS households, 81.7% of BHPS households and 69.2% of EMBS households participated (Table 24). The non-contact rate was against highest for the EMBS, while household refusal rates were lower for the BHPS sample than for the other two sample components.

Interview data was successfully obtained for 69.9% of persons aged 16 or over enumerated in the GPS (Table 25), 61.1% of persons enumerated in the EMBS (Table 26) and 76.1% of those in the BHPS sample (Table 27). In all three samples, TSMs were considerably less likely than OSMs to give a full interview. A proxy interview was more likely for TSMs. In both GPS (Table 25) and BHPS (Table 27), a proxy interview was more than twice as likely for TSMs as for OSMs, while the differences were smaller for the EMBS (Table 26). Broadly similar patterns are observed at each wave.

Amongst OSMs who had been interviewed at Wave 2, interview data was obtained again at Wave 3 for 84.1% in the GPS (Table 28), 78.1% in the EMBS (Table 29) and 89.4% in the BHPS sample (Table 30).

Response rate to the adult self-completion – which was administered by CASI rather than on paper for the first time at Wave 3 – was 88.7% amongst all persons interviewed, including 89.7% in the GPS, 89.5% in the BHPS and 79.6% in the EMBS (Table 31). The self-completion response rate was even higher amongst those who had completed the self-completion at Wave 2, at 91.5% in the GPS (Table 32), 92.7% in the BHPS (Table 34) and 84.8% in the EMBS (Table 33). Overall response rate to the youth self-completion was 74.3% (Table 35). Amongst OSMs, the youth questionnaire response rate was 83.9% amongst those who had completed the youth questionnaire at Wave 2, 71.9% amongst 10 year-olds who were eligible for the youth questionnaire for the first time at Wave 3, and 48.6% amongst those who had been eligible at Wave 2 but had not completed the questionnaire (Table 36).

7.1.4 WAVE 4

At Wave 4, 80.7% of eligible GPS households, 83.5% of BHPS households and 72.6% of EMBS households participated (Table 40). These proportions were slightly higher than at earlier waves, though of course the denominator excludes any households that were no longer issued to the field due to persistent non-contact or adamant refusal (see section 0). The household non-contact rate was again higher for the EMBS than for other sample components.

Amongst enumerated OSMs aged 16 or over, interview data was obtained for 76.9% in the GPS (Table 41), 66.2% in the EMBS (Table 42) and 80.0% in the BHPS (Table 43). Amongst those interviewed at previous wave, the equivalent rates were

88.6% in the GPS (Table 44), 80.8% in the EMBS (Table 45) and 89.2% in the BHPS sample (Table 46).

Response rate to the adult self-completion amongst all eligible persons interviewed at Wave 4 was 90.1% overall, including 91.5% in the GPS, 90.6% in the BHPS and 79.8% in the EMBS (Table 47) – very similar proportions to those observed at the previous wave. Amongst those who had completed the self-completion component at Wave 3, the response rate was 95.6% in the GPS (Table 48), 96.2% in the BHPS (Table 50) and 88.5% in the EMBS (Table 49).

The youth questionnaire response rate for OSMs was 85.8% amongst those who had completed the youth questionnaire at Wave 3, 70.2% amongst 10 year-olds who were eligible for the youth questionnaire for the first time at Wave 4, and 49.0% amongst those who had been eligible but had not completed the questionnaire at Wave 3 (Table 52). Amongst all young persons eligible for the youth questionnaire at Wave 4, the response rate was 74.7% (Table 51).

7.1.5 WAVE 5

Analysis of participation rates at Wave 5 will appear in the next edition of the Quality Profile.

7.1.6 HEALTH ASSESSMENT

Of those eligible for the Nurse Health Assessment visit (see section 4.5), 57.9% of GPS members took part at Wave 2 (Table 37) and 56.6% of BHPS sample members took part at Wave 3 (Table 38). Of those who took part, 63.6% of the GPS and 66.9% of BHPS sample members also provided a valid blood sample.

7.2 ITEM NON-RESPONSE

Levels of item non-response, i.e., missing, refusal and don't know responses to particular questions, provide a measure of data quality.

Item non-response may be caused by the respondent's perception that the question is too personal or intrusive, resulting in a refusal to provide a response. Alternatively, respondents may give a "don't know" response which is more difficult to interpret (Beatty and Herrmann 2002). It may be the case that respondents genuinely do not know the correct answer, especially where an exact amount or specific date is being requested for example. On the other hand, a "don't know" response may be a polite refusal. The respondent does not want to give the information but rather than refusing outright, says they do not know. Missing responses are often due to interviewer error where the routing has been followed incorrectly and a question that should have been asked is not. Seeing as *Understanding Society* is administered in CAPI, missing data of this kind should be eliminated even though interviewers could still code a response incorrectly and, as a result, be routed incorrectly by the CAPI programme. The other source of missing data is when the respondent does not understand the question or see that it applies to their particular situation so is therefore unable to respond.

In a panel survey, item non-response, particularly to potentially sensitive questions such as income, might be expected to fall over time. As the respondent grows to trust the survey and build a rapport with the interviewer they might be expected to be

more co-operative and more inclined to provide valid responses. Items that require recall over an extended period of time, e.g., the start date of current job, will be subject to recall error, particularly where dates are further from the date of interview. In this case, the recall error would be expected to increase in later waves of the panel for respondents who have not changed their job and for whom the start date is receding into the past.

This section details the overall levels of item non-response on *Understanding Society* Waves 1 - 5. The levels of item non-response on some selected variables are then described. The missing value codes that are used in *Understanding Society* to differentiate between different sources of item non-response are described in section 11.4.

7.2.1 OVERALL ITEM NON-RESPONSE

The following tables show the mean levels of item non-response across all variables collected in *Understanding Society* for each wave.

Table 53 shows that the overall level of item non-response in the individual questionnaire was 2.8% in Wave 1 and this rate has declined to 2.2% in Wave 5. As expected, rates are higher for the "Extra 5 minutes" samples and those who do not provide a self-completion interview. There was an increase to 3.9% in Wave 2 which is partly explained by the incorporation of the BHPS sample for whom a great deal of questions asked about events that had occurred up to two years ago (i.e., since the previous interview). The Wave 2 questionnaire furthermore contained a number of cognitively challenging questions such as the identity module in the self-completion questionnaire and questions about dates when first diagnosed with listed health conditions. The shift from PAPI to CASI in Wave 3 appears to have led to much lower item non-response rates in the self-completion instrument.

Table 54 reports the level of item non-response in the household questionnaire. The overall rate amounted to 1.2% in Wave 1 and reduced to 0.8% in Waves 2 and 3. Wave 4 included a module asking for information about all cars in the household. The information was difficult to retrieve, in particular for those not driving the car themselves (e.g., the engine size). Wave 5 included a new module to confirm the relationships between household members which has high levels of item-non-response (around 3%), in particular for larger households.

7.2.2 ITEM NON-RESPONSE FOR SPECIFIC VARIABLES

Table 55 provides the mean level of item non-response for selected variables. The level is consistently low for items such as marital status and health status, suggesting that these are items which respondents do not perceive as sensitive in any way and are items that they can respond to without difficulty. We have also included in the table some of the items with higher levels of item non-response where the main reason for item non-response lies in failure to code free text information to official codes.

We also included in the table a number of items which are considered sensitive by the respondents and tend to have high rates of item non-response in surveys. These include, notably, measures of income, savings, housing costs and political allegiance. It is clear that item non-response was higher at Wave 1 (and Wave 2, which was the first wave as part of *Understanding Society* for the BHPS sample) for many of these variables and in subsequent waves has tended to fall.

Item non-response for items such as exact dates, which tend to be difficult to recall, follow a different trend. Whilst less than 10% of respondents in Wave 1 had difficulties to recall the exact date their current employment had started, the same was true for 28.5% in Wave 5. This is, however, remarkably low compared to previous studies: the respective figures in the BHPS were 41.8% and 47.2%, respectively.

8. DATA PROCESSING AND CLEANING

The schedule for delivering data from the fieldwork agency to ISER has changed over the course of the Study. The current schedule is that ISER receives non-validated data for the first sample month (i.e., January year 1) about five months after fieldwork starts, and validated data for that sample month six weeks after the end of fieldwork (i.e., about six months after fieldwork for the new wave has started). Validated data for the entire first quarter of the wave (i.e., year 1 Q1) are sent to ISER six weeks after Q1 finishes. Data for the following sample months are sent to ISER in quarterly batches. This allows time for interview re-issue, coding, and data entry from paper documents, e.g., the youth self-completion or consent forms. Data is delivered as SPSS system files, which are then exported to triple-S data exchange format and imported into a SIR database.

Quality control processes include extensive data checking to ensure that the data conform to the expected structure and to the routing and range constraints defined by the questionnaire specifications. Data anomalies are investigated to determine whether they are related to:

- the invalid specification of the questionnaire;
- the incorrect scripting of the questionnaire;
- a failure to specify that a particular constraint should be included in the questionnaire;
- an incorrect implementation of the check, or;
- a problem in exporting and/or delivering the data.

After investigation, steps may include correcting the specification, data editing, reporting the error to the fieldwork agency to be fixed in a subsequent delivery and/or a quality feedback report suggesting changes to the questionnaire or field practice in subsequent waves.

Some variables are also routinely checked in the process of creating added-value content. These include the respondent's age and sex, based on the information collected across all waves, see section 8.1.1 below, or pointers to specific other members in the household such as a biological parent, see section 8.1.2 below.

Batch-specific databases are merged into a single database, from which anonymised data is exported for the creation of public use files, see section 11.8 below.

Data distributions are also checked for theoretical and statistical plausibility. This checking is done through direct scrutiny and by analyses which 'road-test' the data.

Data processing and cleaning is an ongoing process.

8.1 CONSISTENCY CHECKS OF CORE VARIABLES

8.1.1 CONSISTENCY OF REPORTED AGE AND SEX

Basic demographic information on sample members is collected in a number of different places and at different points in time, which means there is scope for inconsistencies, both within and across time.

Information on age and gender is first collected from the respondent providing information for the household enumeration grid. In addition to the respondent not knowing the exact date of birth from which the age can be computed, there may be recall errors and data entry errors. If the respondent does not know the exact date of birth of other household members they are asked for an estimate. The respondent's age is then computed (see w_dvage) from this information for the current wave. Anybody participating in a personal interview and agreeing to do the self-completion interview in Wave 1 and 2 was furthermore asked to provide their own date of birth (and sex). Although respondents can be assumed to know their own date of birth, there may have been issues with the interpretation of the respondent's handwriting when the information was scanned and translated to data.

From the second wave onwards stable characteristics are typically only asked of new entrants each wave during the household enumeration and initial conditions questionnaire modules. Likewise, data about continuing respondents are checked and corrected each wave. Before releasing the data we compute cross-wave corrected variables for gender (see w_sex_cr), and age (see w_age_cr) based on data from the latest enumeration of each sample member. In this way the data quality of the demographics is assumed to increase over time. Hypothetically, data entry errors could creep in at a later wave, but would again be subject to further checks the following wave and so on. Only values that deviate from the computed age by more than a year are used for the corrected age variable. Both the computed and derived variables are subsequently released.

Table 56 reports the prevalence of cross-wave inconsistencies in reported age and gender which are corrected in the respective wave-specific age and gender variables for the Wave 5 release. It can be seen that less than 1% of the sample members' sex and age reports are inconsistent, and that the number of inconsistencies diminishes significantly over time as the information gets confirmed or corrected by the respondents in a later interview.

8.1.2 CONSISTENCY OF RELATIONSHIPS BETWEEN HOUSEHOLD MEMBERS

Relationships between household members are collected via a full household grid and using a coding schema that covers 30 different categories (see Table 57).

The information is provided by the person answering the household grid. The number of questions asked depends on the household size. A two-person household is asked only one question: "What is your relationship to <<name of other person>>?" Because relationships are reflexive there is no need to ask the reverse relationships. A five-person household is asked a total of six questions; the respondent (PNO1) would be asked first about their relationship to PNO2, PNO3 through to PNO5, then about PNO2's relationships to PNO3 through to PNO5, and so on until, at last, PNO1 is asked for PNO4's relationship to PNO5.

Whilst the full household grid is easier to collect than, for instance, the relationships to a reference person - because the respondent and interviewer do not have to remember who the reference person is whilst working out the relationships, and the number of questions that need to be asked to get a complete picture of the relationships is smaller - there is nevertheless potential for measurement issues. Firstly, a relationship may have been reported with error. For example, the respondent may not have made the intended distinction between, biological, half- and step- siblings, the relationship 'partner/cohabitee' may have been interpreted as 'live-in business partner', family friends may have been reported as 'cousin' or 'uncle', and the tenant/landlord or adopted parent/child relationships are not necessarily exclusive from biological relationships. Secondly, there may have been data entry errors such as entering the wrong relationship code. Any such errors will translate into the reflexive relationship also being computed incorrectly.

Table 58 reports how common any of nine potential problems are in Waves 1- 5. The universe is relationship pairs recorded in the respective wave's w_egoalt data file. Any pair in a household with a potentially problematic pair is flagged up. Note that this error checking relies on information on respondents' age (w_dvage) which may be recorded with error (see section 8.1 above).

It can be seen that 5.5% of relationship pairs were flagged up for further inspection in Wave 1, and this proportion was fairly stable over time amounting to 4.6% in Wave 5. In practice, a number of problems cumulate (in larger households and households in the EMBS in particular), and some flagged relationships – such as the presence of siblings who are greater 20 years apart in age - turn out to be consistent upon case-by-case inspection. The pattern of prominence of potential issues changes somewhat over time. Whilst exceptionally large age differences between siblings continue to be the most frequent potential issue flagged up (as is expected given this relationship, if correct, will persist over time), it may be the case that recording errors (rather than reporting errors) became more prominent in Waves 4 and 5.

Whilst it is straightforward to identify these potential conflicts in the data, it is not straightforward to correct the information: not all relationships are mutually exclusive and over time, relationships and perceptions may change. Whilst we provide a cleaner edited version of the relationship variable, we do not test it for cross-wave consistency.

Table 59 reports the relationship codes in Waves 2-5 for respondents who were coded as a natural child in Wave 1 and whose relationship pair was observed at least once more in Waves 2-5.

It can be seen that there is considerable stability over time in this relationship status. In the absence of change we would expect 100% of the relationships to remain the same: 99.2% of natural children in Wave 1 were observed again as natural children when the original relationship variable is considered, and the respective figure is 99.7% when the edited information is considered. Whilst the edited information is somewhat more longitudinally consistent both versions include relationship codes that are implausible over time. For example, some of the Wave 1 natural children were recorded as their parent's husband/wife/ or cohabitee, or as natural siblings in at least one later wave.

Note that the relationship grid is used during the interview to create a number of pointers to significant others which are then used to route questions (for example,

about the partnership or parent-child-relationships). If there is an error in the relationship grid this will therefore feed forward to substantive information. Whilst we provide an edited relationship variable (w_relationship_dv) we do not currently edit substantive information collected in the interview to make it consistent with our edited household grid.

Information from relationship checks included in the script will be used more efficiently in the next rounds of interviews to help minimise relationship inconsistencies within and across waves. We will then also review the procedure regarding editing substantive information to missing or inapplicable.

8.2 CODING

Understanding Society collects free text information on respondents' job titles and the industry of the job held. Industry descriptions are coded to ONS Standard Industry Code (SIC) 2007, SIC 2007. Job titles are coded to the ONS Standard Occupational Classification (SOC) 2000, or SOC 2000. Coding is undertaken using the Computer Assisted Structured Coding Tool (CASCOT) system. From Wave 3 onward job titles are coded to SOC 2010 in the first instance and look-ups are used to code to SOC 2000. We also provide SOC 1990 codes using look-ups between SOC 2000, SOC 2010 and SOC 1990. It should be noted, however, that there are currently some gaps in this coding for the respondents' current and last jobs in the first waves of interviews (because SOC 2010 and look-ups to SOC 1990 were not yet available).

Several questions, e.g., country of birth, religion, political party, national identity, and citizenship had an "other, please specify" option. These responses were coded using an automated process.

Coding was also done for an open-ended question which read "We've asked you a lot of questions but we also want to know what has happened in your own life that has been especially important to you. Can you please tell me anything that has happened to you, or your family, over the past year that has stood out as important?" The respondent could give up to four answers. The answers were recorded verbatim and manually coded for type of event and its subject.

8.3 DERIVING VARIABLES

Derived variables are variables that are computed from one or more variables. Some are computed during the interview to control the routing within the questionnaire, whilst others are computed post-field for the purpose of analysis. Information about how the derived variable is produced is shown in the notes for derived variables in the detailed variable view of the online dataset documentation.[16] The view provides descriptive statistics and, in the 'Origin' field, lists of the variables used in the computation of the derived variable. For variables that were computed during the interview, additional information is available in the questionnaires.

It is anticipated that the number of derived variables will increase over time, and we are generally open to suggestions for inclusion of derived variables that are of high

[16] See https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation.

quality, are widely used, and can be produced routinely on the basis of existing *Understanding Society* data.

One area where we have already established a routine for deriving further variables is in socio-economic classifications. Here, we use look-up files between SOC 2000 and other classifications to derive additional occupational classifications.[17] We provide the following classifications International Standard Classification of Occupations (ISCO88), Registrar General Social Class (RGSC), National Statistics Socio-economic Classification (NS-SEC), Employment Status (ES), and Socio-economic Group (SEG). These are computed for the respondent's current job and last jobs only. SOC 2010 is provided but SOC 1990 is not available for the current and last job.

8.4 DEPENDENT INTERVIEWING

From Wave 2 onward, proactive dependent interviewing was used to increase efficiency of data collection and lessen respondent burden. Specifically, information reported at an earlier time is fed forward to the respondent to personalize the question. So rather than ask a question about current occupation with its complex probing by interviewers, the status check question might say, "the last time you were interviewed you said you were <specific occupation> are you still <specific occupation>? If the respondent confirms the previous occupation, the previous information is used. If the respondent does not confirm the previous occupation, they are asked the independent question instead. In the data, information collected using dependent interviewing is combined with the respective information collected using independent interviewing; the status check variable is also included to allow disentangling the different questioning routes.

Feed-forward variables are used at both the household and individual levels. For example b_ff_hhsize feeds forward the household size from the previous wave (Wave 1). The variable b_ff_plbornc is the country of birth of the respondent fed forward from the previous wave. Some of the fed-forward variables were not used in the wording of a question but were used by the CAPI script to route respondents appropriately based on information from the previous wave.

[17] For further information, see www.cf.ac.uk/socsi/CAMSIS/occunits/distribution.html#UK

9. WEIGHTING PROCEDURES

Analysis weights are provided on the public use data files. These are designed to correct for differences between the sample and population distributions caused by the sample design (design weights) and for differences between the responding sample and the selected sample caused by non-response (non-response weights). As the level and nature of non-response can be different for different survey instruments (household enumeration grid, household questionnaire, individual questionnaire, self-completion questionnaire, youth questionnaire, nurse health assessment etc.) and for different waves of data collection (see section 7.1), several different sets of analysis weights are provided, corresponding to different combinations of instruments and waves that may be appropriate for different analyses.

9.1 RANGE OF WEIGHTS PROVIDED

The choice of weights to provide for data users is an important decision for a household panel survey. An extremely large number of potential analysis bases are possible and in principle the optimum set of weights would be different for each analysis base. For either the total population or any subset of it (e.g., women, unemployed persons, or persons in Scotland) an analysis base is typically defined by a combination of waves and instruments. After five waves of Understanding Society, there are 31 possible combinations of waves (5 individual waves, 10 pairs of two waves, 10 triplets of three waves, 5 sets of four waves, plus the set of all five waves). There are at least four major combinations of instruments that are of interest to analysts (household grid/questionnaire, individual interview, individual or proxy interview, individual interview and self-completion)[18]. Thus, the total number of combinations of waves and instruments is already at least 124 and will continue to grow exponentially as more waves of data are collected. Additionally, there are combinations of instruments that include the nurse health assessment and/or the blood analysis. The provision of a separate set of weights for each of these combinations would in principle support the majority of possible analyses, but there are at least two arguments against producing all possible weights:

a) The resource implications of deriving and documenting so many weights are considerable;
b) There would be an increased risk of confusion on the part of the analyst about which weight to use for a particular analysis (and a consequent greater risk of inappropriate weights being used).

On *Understanding Society*, a decision was therefore made to limit the sets of weights produced. The sets of weights that have been provided were chosen based on

[18] The survey procedures do not permit certain combinations, such as an individual interview without a household enumeration grid.

knowledge of the data structure and of forms of analysis that are most likely to be of interest to users. Weights have *not* been produced for:

- combinations of instruments and waves that occur only rarely (for example, household grid without household questionnaire);
- combinations of instruments and waves that are unlikely to be of interest to many analysts (for example, youth self-completion questionnaire in Waves 1 and 5 [as only young people aged 10 or 11 at the time of Wave 1 would still be eligible for the youth self-completion questionnaire at Wave 5]).

The *Understanding Society* data can also support analysis for some bases that are not defined by combinations of waves and instruments. Examples would include analysis of calendar years or financial years. These non-standard analyses are handled not by producing extra sets of weights but instead by providing guidance on how the existing weights can be used in such analyses.

With the first five waves of data, a total of 92 sets of analysis weights have been provided to users, in addition to eight sets of design weights for specialist use by advanced users. The purpose of each set of weights is described in Knies (2015), where guidance is also given on identifying the appropriate weight for any particular analysis.

9.2 METHODS FOR DERIVING WEIGHTS

Each set of weights has been developed using best practice in survey weighting methodology and drawing upon the widest possible set of auxiliary variables to inform the weighting models. At Wave 1, auxiliary variables were taken from linked Census Small Area Statistics, official Neighbourhood Statistics and interviewer observations. At all subsequent waves a broad range of survey items from previous waves were also used. The methods used for each set of weights are documented in the User Guide, see Knies (2015).[19] The complexity of the design of *Understanding Society* has sometimes resulted in the need to extend standard methods. This has led to original research into weighting methods for complex longitudinal surveys (e.g., Kaminska and Lynn 2012a; Kaminska and Lynn 2012b; Lynn and Kaminska 2010; Sadig 2014a; Sadig 2014b).

[19] Available online at https://www.understandingsociety.ac.uk/documentation/mainstage.

10. IMPUTATION PROCEDURES

10.1 VARIABLES FOR WHICH IMPUTATION IS IMPLEMENTED

Imputed values are provided for income variables, but not for any other variables. The reason for this is that income measures are particularly important to many data users, but missing data rates are relatively high for income variables compared to most other types of variables on *Understanding Society*. There are a large number of income variables that contribute to computed measures of total personal and household income. Imputation is carried out for each such variable in order to be able to compute total personal and household income for all individuals belonging to responding households at each wave. Income variables are not imputed for individuals in non-responding households.

For individuals who respond to the individual questionnaire but do not provide answers to all income questions (item non-response), the following personal income variables are imputed if missing: wages, self-employment earnings, second job earnings, interests and dividends, pensions, benefits and other income sources. For individuals for whom a proxy interview was carried out, total earnings and total income are imputed whenever missing (the proxy interview does not include questions on the more detailed components of personal income). For individuals in responding households for whom neither the personal nor the proxy interview was completed, only total personal income is imputed. This imputed value is not included in the dataset, but is used in the computation of total household income.

10.2 METHODS FOR IMPUTATION

Both cross-sectional and longitudinal imputation methods are used to replace missing data. For each survey wave, cross-sectional imputation methods are applied first, using linear regression (for continuous variables), interval regression (for continuous censored variables), logistic regression (for binary variables), ordered logistic regression (for ordered variables), multinomial logistic regression (for non-ordered categorical variables), as well as predictive mean matching and hot-deck imputation. For certain sets of related variables, the imputed values are predicted jointly using the method of imputation by chained equations (ICE). Longitudinal imputation then imputes values taking advantage of values reported or imputed in previous waves, using either the carryover method or the Little and Su method (Little and Su (1989)). The details of the methods used and the variables to which they are applied are described in section 3.8.2 of the User Guide, see Knies (2015).[20]

For each income variable with imputed values, a corresponding ‘flag’ variable indicates whether the amount was reported or imputed.

[20] Available online at https://www.understandingsociety.ac.uk/documentation/mainstage.

11. DATA

11.1 DATA STRUCTURE

Given the design of *Understanding Society* (individuals within households, longitudinal, new entrants at each wave, multiple instruments), it is inevitable that the data will have a relatively complex structure. Moreover, the range of different research uses focusing on different combinations of units of analysis, waves and instruments means that researchers require flexibility in the way that data can be combined and structured. Reducing the released data into a simple rectangular format was therefore unlikely to be either feasible or efficient. Instead, data are organised in a series of files for each wave, with each file typically corresponding to a single questionnaire instrument, or to a module of related items with a particular structure. For example, one file contains information relating to childcare and related issues, with one record per child, although the information was asked in the interview of a parent/carer, with each interviewee potentially providing information about multiple children.

Analysts are provided with a number of potential linking variables which make it easy to merge data from different files. Prime amongst these is the unique person identification number, pidp, which appears on all data records relating to that person and remains unchanged over the life of the Study. The User Guide (Knies 2015)[21] provides guidance and sample code for merging files in a variety of ways that are commonly required by users.

Simple, informative and consistent file naming conventions are used to make data analysis less demanding and less error-prone. File names begin with a prefix designating the wave of data collection ("a_" for the first wave, "b_" for the second wave, and so on). The core part of the filename has the same meaning at each wave. For example, individual level data collected from interviews with responding adults at wave *w* is stored in the file *w*_indresp (w = a, b, c, d, e) and household level data collected in the household interview at wave *w* is stored in the file *w*_hhresp. There are also files which contain no substantive data, but whose purpose is to facilitate linkage of data relating to the same individual at separate waves.

Table 60 lists the core data files that are produced for each wave of *Understanding Society*. First, it lists files that contain substantive information collected in interviews with responding households and individuals. They are the most commonly-used data files. Next are data files that are available for all enumerated households and individuals and cross-wave files that include information for everyone who has ever been observed in the Study. For example, the file xwavedat contains stable characteristics of individuals, such as ethnicity and age of leaving full-time education, which is typically collected only once in the lifetime of the Study. Last but not least the table list files that contain paradata, i.e., additional data collected about the interview process such as call records and time stamps.

[21] Available online at https://www.understandingsociety.ac.uk/documentation/mainstage.

For ease of use, some information stored in higher level data files is copied over to other data files. Examples include the region variable which is available for all sampled households and copied over to all data files as it is an important context variable. Additionally, all files include derived variables, weights and imputed incomes, as relevant for analyses of the respective data. The complete list of files and their descriptors can be seen in the online documentation system.[22]

11.2 IDENTIFIERS

In addition to the cross-wave personal identity number pidp, already mentioned above, the *Understanding Society* data includes wave-specific key variables which can be used to link together information from different levels within one wave (but not to connect information across waves). These variables uniquely identify:

(1) a household at wave *w*; *w*_hidp;
(2) an individual's number within the household at a given wave, *w*_pno. This number is used to identify various relationships between people within a household.

The variables *w*_hidp, *w*_pno and pidp are included in all relevant files.

Additionally, for members of the ex-BHPS sample, their previous BHPS cross-wave personal identity (pid) is included on *Understanding Society* data files in order to facilitate linkage to BHPS data from earlier years prior to *Understanding Society*.

11.3 VARIABLE NAMES

As with data files types, all variable names (except those used in cross-wave operations) begin with a single character wave identifier followed by an underscore. Variable names correspond to questions specified in the interview script, also see section 5.2.3 above. Questions that have multiple items are typically presented with the question name followed by (a) the category number (e.g., w_trqual1-w_trqual31) or (b) letters running from a to z (e.g., w_scghqa-w_scghql). Where question wording remains identical or substantially the same across waves the variable name suffix will remain the same. The documentation draws attention to minor wording differences, and for example to a small number of cases where there are changes in the coding response categories. Where it has been judged that the question wording, or the range of response categories has changed sufficiently that the responses might be treated as equivalent rather than identical, a new variable suffix is used. This is identified in the documentation, and users can also use the cross-wave subject index to identify other potentially comparable variables.

11.4 MISSING VALUES

Standard conventions have been consistently applied to represent the variety of situations where respondents did not provide data in response to questions, or where a variable could not be computed. These conventions distinguish the following:

[22] See www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation.

0 represents “not mentioned” or “none” (unless it has some other meaning in the coding frame). Thus, where respondents are asked which of a list of items apply to them, those not selected will be coded 0.

[-1] represents a respondent response of “don’t know” – these include both pre-codes, and interviewer write-in.

[-2] represents a respondent refusal – these include both pre-codes, and interviewer write-in.

[-7] is used on individual respondent records to indicate that the respondent was interviewed by proxy or by telephone and the relevant question was not asked, or the derived variable could not be computed. It is also used on the household record to indicate that the only household contact was a telephone interview, so that the household schedule was not completed.

[-8] is used to indicate that a question was not applicable according to the instrument routing.

[-9] represents data that are missing in error, with no other explanation, including implausible values that failed hard edit checks and derived variables which could not be computed due to missing components. This is the default missing value. It is assigned also to information such as coded variables when the correct applicability has not yet been determined.

These conventions are used consistently across all waves.

11.5 RESPONSE DATA

There is a complete accounting for the response status of all active sample members at each wave. One data file (xwaveid) contains household and individual response status for each sample member at all waves. This permits identification of both response status, and reason for non-response (distinguishing non-contact, refusal at household or individual level, out of scope and deceased). This information is also carried on individual wave data files, including a household sample data file (w_hhsamp), giving response status of all households issued, and also further data about reasons for non-response. Another record type (xwlsten) gives sample status, and wave at which this last changed for all sample members.

11.6 VARIABLES OF PARTICULAR INTEREST

11.6.1 SURVEY DESIGN VARIABLES

Standard statistical data analysis programmes will treat survey data as if they were from a simple random sample and calculate estimates accordingly, unless the analyst specifies otherwise. As described in section 4, with the exception of the Northern Ireland sample this is not the case in *Understanding Society* (and even in Northern Ireland, the sample is subject to non-response, for which corrective weights are supplied – see section 9 above). In order for analysts to be able to adjust their analysis for the complex survey design we include information about stratification and primary sampling units in our data products.

The variable w_psu is an indicator of the PSU to which the sample member belongs. Values of the variable *w*_psu are further described in

Table 61. The variable w_ strata this indicates the sampling stratum from which the sample member was selected. The range of values on *w*_strata is listed and explained in Table 62.

The values of *w*_psu and *w*_strata do not change between waves, but for new sample entrants they are only defined from the wave at which they enter the sample. These variables are included in all core *Understanding Society* data files.

11.6.2 WEIGHTS

Given the complexity and multi-purpose nature of the *Understanding Society* design we provide multiple weights to meet the different needs of users, as outlined in section 9.1 above. Naming conventions for weights are intended to help users to pick the correct weight and to quickly locate any particular weight. The name of each weight reflects the wave for which the weight is calculated, level of analysis, data source and its nature (design weight, cross-sectional analysis weight or longitudinal analysis weight). The rules are described in Table 63. The *Understanding Society User Guide* provides guidance to users on applying weights in their analysis, as well as technical details of how each weight was derived.

The User Guide also contains a list of all the available weight variables, broken into separate tables corresponding to different analysis bases. This enables the user to rapidly identify the weight relevant to any particular analysis. Further guidance on weighting is also provided on the online *User Support Forum* (see section 12.1).

11.6.3 DERIVED VARIABLES

Users are provided with a suite of derived variables to simplify the analysis task and to ensure a degree of consistency between users. (The original variables from which these were derived are of course also included, so that users can derive their own alternative measures if they wish to.) The derived variables include multiple-item scales such as GHQ and SDQ, flags for whether or not a certain characteristic is true for a study member (e.g., *w*_jbft_dv is a flag for whether or not a respondent has a full-time job), counts of the number of people in the household for whom a certain characteristic is true (e.g., *w*_nemp_dv is the number of employed people in the household), and pointers to significant others in the household (e.g., mnpid records the cross-wave person identifier of the respondent's biological mother).

As a rule of thumb, variables derived post-field end with the suffix '_dv', and pointers to others in the household end on 'pno' or 'pid'; they can, therefore, easily be identified in the data. Derived variables that are created post-field are added last to the data files and can therefore also easily be identified by their position in the files.

By contrast, variables derived during the interview by the CAPI script cannot easily be distinguished from other questionnaire items in the data. These variables can, however, be identified using the search term 'Compute' in the questionnaires, and many of them have derived variable notes in the online dataset documentation.

Note that a data file may offer alternative versions of a derived variable. This is particularly true for derived variables that point to others in the household: One version is computed post-field after the information collected in the household grid

has undergone extensive data cleaning, whilst the other version is the unedited original. The 'Associated variables' field in these variables' variable view provides references to such related variables.

11.6.4 INCOME VARIABLES

A range of derived net income variables is provided for users. At the individual level monthly income, net of tax and national insurance, is provided, along with each of six main components of net income, which are defined in a way that is intended to be comprehensive and mutually exclusive. These components are: labour income, private benefit income, investment income, pension income, social benefit income and miscellaneous income. The sub-components of each of these six income components, as reported in the questionnaire, are documented in the User Guide. As well as providing these net income variables, to facilitate analysis and promote coherence, the sub-component source variables as reported by respondents are also provided so users can derive their own measures should they so wish. Similarly, a measure of gross monthly income is also provided, along with its three major components, gross labour income, gross income from savings and investment, and gross income from benefits and other sources. Table 64 lists all net income variables provided with the *Understanding Society* data.

As these derived income measures are reliant on a large number of reported values of specific components, there is a high proportion of respondents for whom at least one of the components is missing. To deal with this, a set of imputation procedures is used for income variables prior to deriving the summary measures of net and gross income. These procedures are described in section 10 above.

11.7 PROFILE OF RESPONDING HOUSEHOLDS AND INDIVIDUALS

Understanding Society is designed to be a multi-topic multi-purpose household longitudinal study. Its sample is designed to assure that it is possible to analyse the data at multiple levels and over time, including for groups and phenomena that are less prevalent in society.

Table 65 provides information about sample sizes for a number of characteristics of households in *Understanding Society*. It can be seen that *Understanding Society* includes substantial numbers of households in all but a few of the characteristics. For example, in all waves of the Study, the responding households include:

- more than 2,200 households with five or more people,
- around 1,000 male (2,500 female) pensioners living by themselves,
- more than 650 singles living with one child, and roughly the same number of singles living with more than one child,
- more than 300 households formed by three or more adults excluding any couples, some with children, and some without.
- more than 170 households that live in accommodation supplied by the employer,
- more than 3,400 households who live in crowded accommodation (defined here as there being less than one room per person in the accommodation),

- more than 2,800 households living in poverty[23],
- more than 1,000 households who are behind with paying some bills, such as rent or mortgage payments, Council Tax or other bills.

Table 66 provides information about sample sizes for a number of characteristics of adult respondents in *Understanding Society*. It can be seen that *Understanding Society* includes substantial numbers of individuals in all but a few of the characteristics. For example, in all waves of the Study, the responding adults include:

- more than 20,000 people in paid employment, around 10,000 people in retirement, but also sizeable numbers of individuals in less frequent employment statuses such as 229-365 adults on maternity leave;
- 478-699 men who have fathered a child, and 427-619 women who gave birth since the last interview;
- more than 20,000 adults who were either mostly or completely satisfied with their life, as well as more than 3,000 adults who were either completely or mostly dissatisfied with their life.

The key benefit arising from a longitudinal study is that it permits the analysis of change at the individual level.

Table 67 reports a range of indicators of wave-on-wave individual change. Absolute frequencies are reported separately for each wave, pooled for all waves on an unbalanced sample, and pooled using a balanced sample of only those individuals who provided a full adult interview in all waves.

Overall, there were 228,608 full adult interviews from Waves 1-5, and 22,285 adults provided an adult interview in all five waves, yielding 111,400 person-year observations for a balanced panel analysis. In each wave of the survey the incidence of change recorded in the Study is sizeable. At each wave, the Study has captured, for example:

- more than 250 incidences of adults living in a different region of the UK than in the last wave, yielding a total of 1,256 such transitions (balanced sample: 719) over the first five waves;
- more than 1,800 transitions into employment, more than 600 transitions into self-employment, and more than 1,600 transitions into unemployment from one wave to the next;
- 2,200-4,000 exits from poverty each wave resulting in more than 11,000 exits (balanced sample: 6,406) over the first five waves; likewise, 7,875 entries into poverty from one wave to the next have been recorded (balanced sample: 4,339);
- more than 800 incidences of adults gaining a new qualification, yielding a total of 3,614 skills improvements (balanced sample: 1,359) over the first five waves.

[23] Defined here as individuals living in households with less than 60% of the median equivalent household income in the UK in the respective year.

11.8 DATA RELEASE

Table 68 lists all *Understanding Society* data products that are deposited with the UK Data Service. The core product is SN 6614; it is the standard end-user licence version of the *Understanding Society* data for all waves for which data collection and processing has been concluded. In the process of testing and cleaning the new data we also update, correct and add to information for previous waves. We therefore recommend that users analyse the latest release data, and that they cite the release version of the data so the analysis can be replicated.

Typically, data for new waves are ready for deposit with the UKDS mid October (i.e., five months after the last interviews of that wave have been collected and were transmitted by the fieldwork agency to ISER), and the UKDS takes another month to check the data and update their documentation systems. Special License (SN 6931) and Secure Data Access (SN 6676) versions that include additional sensitive variables are made available at the same time, as described in the User Guide.

Look-up files between official geographical identifiers and *Understanding Society* household identifiers (see products marked by the Study type 'geo-link-id' in Table 68 are deposited with the UKDS in the first quarter of the following year; producing these files requires using a clean version of the respondent's postcode as recorded in the Study administration system with the respective geographical identifier on the National Statistics Postcode Lookup.

There are also a number of products which are currently deposited as a one-off. The Nurse Health Assessment (SN 7251 and SN 7587), for example, is released as one set of files relating to data collected in Waves 2 and 3 only; likewise, school codes (SN 7182) and linked education records (SN 6742), and the Interviewer Survey 2014 (SN 7615) data currently relate to the first wave of the Study. Genetics data from the genome wide scan is available by application, which can be found on the Study website.

In accordance with the UK research councils' open access policy we are happy to liaise with researchers about the possibility to distribute added-value data from their research projects as part of the *Understanding Society* data series. The harmonised net income, fertility and employment history files on the BHPS, and the *Understanding Society* Geographical Accessibility (SN 7533) products are examples of this collaboration between the Study team and individual researchers. It is anticipated that there will be further *Understanding Society* data products of this nature in the future. Researchers who wish to make their project data available through this route should get in touch with the *Understanding Society* team.[24]

11.9 DATA USAGE

Understanding Society has a multi-topic multilevel research design and offers opportunities for research within and across disciplinary boundaries. Household longitudinal studies have been very much rooted in the social sciences, more specifically in economics and quantitative sociology, and their focus was predominantly on collecting socio-demographic and socio-economic data with some

[24] The Study email address is info@understandingsociety.ac.uk.

additions of more subjective markers of quality of life, including in the domain of health and wellbeing, at later stages in the life of the studies. Parallel to this, investments have been made into longitudinal cohort studies which are much more rooted in the health sciences. Whilst data from the BHPS have been used by analysts from a broad range of disciplines, they were most heavily used by economists and quantitative sociologists, see Lynn (2006). The aim for *Understanding Society* is explicitly to be a high-quality resource for research across the health and social sciences.

11.9.1 REGISTERED USERS AND PROFILE

Since its first release, in 2010, the number of *Understanding Society* data users has constantly increased.

Table 69 reports the annual number of unique downloads (i.e., count of *Understanding Society* data downloads from the UKDS by a unique user – only counted once per day regardless of the number of daily downloads completed) for the standard end user licence data and its special licence and secure access versions from 2010-2015. Overall, access to the end user licence data has steadily increased over the years, starting from just under 450 in 2011, and amounting to 1,427 in 2015. Quarterly figures (Figure 1) indicate that new downloads occur with greater frequency in the first and fourth quarters of the year, i.e., when a new wave of data and revisions to previous waves have been released.

There has also been steady demand for special licence and secure data products. Each year since its release, a small but steady number of new users have been granted access to the special licence version of the core *Understanding Society* data. In 2013, this number was a bit higher than usual which may be explained by the increased demand for innovative secondary data products in that year (prompted by the ESRC's Secondary Data Analysis Initiative).

Table 70 shows that the *Understanding Society* data user group comprised of 4,289 unique users and presents numbers separately by sector. The majority of users are based in the UK Higher Education sector (90%). The second largest user group is based in UK Government (6%).

Figure 2 shows that use of *Understanding Society* data is on the rise across disciplines. Whilst increases in all disciplines have been steady, there has been a somewhat more marked increase in the number of users from the economics and econometrics, business and finance, health sciences, sociology and social policy disciplines in 2014. This may be driven in part by the release of the Nurse Health Assessment data and the inclusion of net household incomes for the first time in the *Understanding Society* main study in that year; these products are particularly salient in these disciplines.

11.9.3 RESEARCH OUTPUTS

Information on research outputs based on *Understanding Society* is much less complete than information on usage, because of the difficulties of tracking down all publications from a very large number of users.[25]

We use various methods to track down publications that use *Understanding Society* data, depending on the publication type and the resource discovery and delivery options available. In terms of journal articles, alerts have been set up through many publishers, databases, hosting services, and web search engines, to search for the phrase 'Understanding Society', 'UKHLS', or 'United Kingdom Household Longitudinal' anywhere within the full text of articles. This includes amongst others, publishers such as SAGE, Wiley and BioMed Central and a range of databases across disciplines, including those within the EBSCOHost service (Business Source Complete, CINAHL Complete, MEDLINE and others), NCBI (PubMed), and Web of Science®, hosting services such as HighWire and the Google Scholar web engine. Any results are sent to us immediately. Alerts have also been set up with the Public Information Online database to alert us to Parliamentary papers and other official publications that use the data. Where alerts are not possible, we perform periodic manual searches. This includes quarterly or biannual searches of databases such as RePEc: Research Papers in Economics, Social Science Research Network, British Library EThOS, and websites such as those of the Joseph Rowntree Foundation, Gov.co.uk, ONS, and the Google Books search engine.

All results are evaluated to verify their use of the data (if possible through analysis of the full text) and then catalogued and indexed. Links to all publications are also provided on the *Understanding Society* publications web page.[26] In addition to the automated alerts and searches, the *Understanding Society* Communications and Web team actively seek the authors of publications that use the data to inform us of the relevant details, through calls and reminders on several pages of the *Understanding Society* website, and through various engagement activities, including social media.

The figures in Table 71 show that there already have been 178 academic papers based wholly or partially on *Understanding Society* data in the period 2008-2015.[27] Of these, 84 are published journal articles and 94 are research papers presented at conferences or published in working paper series such as the ISER Working Paper Series or *Understanding Society* Working Paper Series. The data have also been used heavily in reports (N=74). We note that the data have already been used by five

[25] Albeit, any publication, whether printed, electronic or broadcast, based wholly or in part on the *Understanding Society* data collection provided by the UKDS must be accompanied by the correct citation and acknowledge ISER as the data provider and the UKDS as the data distributor. The acknowledgement, which gives credit to sponsors or distributors, is not a replacement for a proper citation. We recommend the following wording: "*Understanding Society* is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by the National Centre for Social Research and TNS BMRB. The research data are distributed by the UK Data Service."

[26] https://www.understandingsociety.ac.uk/research/search

[27] For comparison, the BHPS Quality Profile has reported 1,042 publications in the period 1991-2005, see Lynn, Peter (Ed.). 2006. *Quality Profile: British Household Panel Survey*. Colchester: University of Essex., p.124. Note, however, that the definition of publications has changed in light of the increasing importance of, e.g., media contributions.

junior researchers in their published doctoral theses, indicating the value of the Study for innovative research projects (counts not included).

12. USER SUPPORT

12.1 ONLINE USER SUPPORT FORUM

An online user support forum was set up in February 2012 to help data users to get the most out the resources we offer and to provide a facility for reporting potential data issues. The forum is powered by Redmine© 2006-2014 (Jean-Philippe Lang). We aim to give short, precise answers within 10 working days. In some cases it can take longer if the nature of the query mandates input from topic specialists.

Users are advised to make use of the documentation section of the website and check the forum for similar questions before posting their own question. They are also advised to offer as much detail as possible about the variables, files, waves, release date, and software that gives rise to the question.

Information on questions posted in the forum as of 10th November 2015 is provided in Table 72 and Table 73. Overall, there were 402 questions. The largest number of them was classified as 'Data documentation' issues (23%), followed by 'Data analysis' questions (19%), and questions about specific aspects of the Study, i.e., 'Weights' (11%), and 'Income' (7%). About half the questions asked have been classified as relating to all waves of *Understanding Society* (i.e., 189 out of 397, or 48%). With respect to questions relating to specific waves of the Study, there were more wave-specific questions relating to Waves 1-3; this is expected seeing as the early wave releases had not undergone as much testing, and important new study components such as interviews with the continuing BHPS sample and issues around how to analyse the data longitudinally became salient to analysts.

In fact, a sizeable number of questions were categorized as relating to the BHPS (20%). Many of them focus on how to analyse information from that sample in terms of data management, but there are also a great deal of queries relating to continuity and comparability of measures across the two studies.

Whilst many of the questions posted in the user forum required pointing the users to more specific sections of the existing study documentation and training materials available online, they have also prompted revisions to these materials. Examples include revisions of the sections on derived variables, income and weights for the latest *Understanding Society* User Guide. The User Guide now also includes look-up tables for which variables are in the self-completion and "Extra 5 minutes"-questionnaires to make it easier for users to make an informed decision about which weight is (most) appropriate for their analysis (Knies 2015).

We have also produced topic-specific user guides for ethnicity research (McFall, Nandi and Platt 2014), cognitive development (McFall 2013), the Nurse Health Assessment (McFall et al. 2014) and biomarkers (Benzeval et al. 2014) and geographical analyses (Knies and Menon 2014). Further guides to weighting and income are forthcoming.

12.3 USER TRAINING

In addition to add-hoc provision of support via the online user support, we launched *Understanding Society* training courses in 2009. The courses aim at familiarising researchers as well as policy analysts with the Study content as well as the complex data structure and analysis techniques.

To date, more than 500 participants from across Europe have benefited from attending *Understanding Society*-based user workshops, amongst them more than 230 have benefitted from attending the 2-day hands-on introductory workshop "Introduction to *Understanding Society* using Stata". These hands-on courses take place 2-3 times are year, and a number of spin-off courses have been developed on its basis so as to train specific audiences, including analysts in government departments, survey methodologists, SPSS and SAS users.[28]

Courses are open to all researchers (although some background knowledge in using the statistical data analysis program is strongly advised).

The majority of *Understanding Society* courses take place at the University of Essex where the Study team has access to computer labs, all *Understanding Society* data, and administrative support. However, we have also responded to requests from the research and policy communities and deliver customised courses in different locations, such as London, Sheffield and Edinburgh. We encourage user groups to get in touch with us to discuss the possibility of delivering the training course at their location.

Whilst we have provided all course materials online since the launch of the hands-on training courses, this was not presented in the most transparent way. In March 2015 we have therefore launched a number of online training courses in Moodle. The materials from the 2-day hands-on workshops are presented there in a more user-friendly format that also allows us to refer analysts who posted questions on the user support forum (see section 12.1 above) to specific applied examples of the data management steps they need for their analysis.

Table 74 lists different types of *Understanding Society* training events alongside the number of participants. It should be noted that the number of participants is not an indicator of popularity of the respective event but rather it is determined by room capacity. It should also be noted that a number of events are not included in the listing. This is true, for example, for a number of 0.5 day introductory courses on *Understanding Society* design and content, which were held by *Understanding Society* team members at conferences and workshops organized by third parties because there are no reliable data on participant numbers available to us.

Table 75 indicates that attendance of training events has been particularly high in the disciplines of sociology and social policy, economics and econometrics and in the health sciences. It should be noted, however, that the majority of course attendants have not specified their disciplinary background.

We ask all participants at our training events to provide feedback on the lecture and IT facilities, training materials (including on separate elements of the courses), and

[28] For further information on *Understanding Society* training courses, see https://www.understandingsociety.ac.uk/documentation/training.

trainers. The feedback has been overwhelmingly positive. There was some indication that the two day workshop was not long enough to cover all aspects of the Study (which has prompted us to make the training materials more accessible online) and users have expressed they would like us to offer more courses on longitudinal data analysis. Options to deliver this are being explored, for example, via webinars.

Table 76 provides statistics on users of the online training courses we have launched in March 2015. It can be seen that 143 users have signed up for the core 'Introduction to *Understanding Society* using Stata' training course and Figure 3 shows that the number of registrations has steadily increased each month. For 42% of registered participants it is not possible to infer from the email where the participant is based, however, the vast majority of online training course participants are registered with a UK higher education email address (55%), and there are also more than five registered users in countries such as Germany and the United States of America. We also note that many of the UK higher education email addresses indicate that a sizeable number of users are based in regions of the UK from where it would be costly to travel to Essex to attend training courses in person.

13. ETHICS

Ethical requirements primarily involve ensuring that the respondent's participation in the Study is informed and voluntary, that the confidentiality of respondents' personal details is maintained and that individual respondents cannot be identified in publicly released data. Legislation in the form of the UK Data Protection Act 2000 governs how personal data are handled and the *Understanding Society* study is required to conform to this legislation.

Beyond these legal requirements there are a number of additional ethical considerations to be taken into account when conducting a household panel which will impact on survey and data collection procedures, see, e.g., Lessof (2009). In some respects there is an inherent tension between the needs of a panel design and the ethics underpinning survey practice. The survey design demands a low attrition rate to maintain the viability and data quality of the panel but this needs to be balanced with the ethics underpinning the implementation of survey procedures.

13.1 INFORMED CONSENT

Some of the key ethical issues are concerned with gaining informed consent and what that should comprise in the context of a panel survey. This is not only informed consent from the respondent taking part but also how one can collect contact details of people not in the survey for tracking and tracing purposes. In order to be valid, consent must be given by the respondent alone and there must not be any coercion. The respondent must be given full information about what their decision involves, including the benefits and risks, and they must have the capacity to understand the information provided to them.

13.1.1 VOLUNTARY PARTICIPATION

Whilst we make respondents aware of the importance of the survey, the voluntary nature of participation is always emphasised. The ESRC Research Ethics Framework section on informed consent reads as follows:

"In all cases of research, researchers should inform subjects of their right to refuse to participate or withdraw from the investigation whenever and for whatever reason they wish. There should be no coercion of research subjects to participate in the research. Consent has to be freely given in order to be valid."

Similarly, the SRA Ethical Guidelines state:

"Research work involving the active participation of human subjects shall be based as far as practicable on respondents' freely given informed consent. Respondents should be made fully aware that participation in the survey is voluntary. Information that would be likely to affect a respondent's willingness to participate should not be deliberately withheld. Respondents shall be told of their entitlement to refuse at any stage to give information, for whatever reason, and to withdraw data just supplied."

In practise, the *Understanding Society* study follows very closely the approach adopted by the BHPS. See Lynn (2006) for a discussion. The key elements are:

(1) The advance letter, which is sent at the start of fieldwork in each wave, gives respondents the opportunity to make enquiries or to withdraw before interviewers enter the field. Respondents are given the name of project's Survey Liaison Manager to contact in the case of queries with a Freephone telephone number and Freepost address as well as an email contact address.
(2) The panel survey is based on projections of long-term co-operation. While it is hoped that respondents will come to feel a sense of belonging to the project, participation at each approach is treated as a *single* commitment. No formal commitment to be re-interviewed is made or asked for at any stage.
(3) At the beginning of each year's interview respondents are advised that each element in the Study and each question within each element is completely voluntary. Verbal permission is obtained from the parent/legal guardian and from the child before giving a youth self-completion questionnaire to a child.
(4) At the end of each year's interview, interviewers are instructed to say that ISER will be approaching respondents about 12 months later for another interview. Information necessary to trace the respondent is then requested. Interviewers are required to give a full explanation of the reasons for asking for this information. Also see section 13.1.2.
(5) Respondents are kept fully informed about the Study through a variety of channels, see section 13.1.3.

13.1.2 COLLECTING STABLE CONTACT DETAILS

The *Understanding Society* study asks all those interviewed to provide a contact name in case they move in the coming year and we are unable to find them. In this case, it is not possible for us to speak to the named contact person to ask for their consent to hold their details. We therefore ask the respondent to tell the contact person so that they are aware we have their details and that they will only be used in the event we cannot find our respondent. On *Understanding Society* we regard this tracking data as an attribute of the individual respondent which is freely provided by them. It is not held in any way which would allow the details of the contact person to be retrieved other than via the indexing of our sample member.

13.1.3 KEEPING RESPONDENTS INFORMED

An aspect of the ethics of social research that can be problematic in any context is the imbalance of information and control of the survey process between respondents and those collecting the data. Respondents provide detailed information about their lives and the control and use of those data effectively passes to the data collector. While this imbalance is an inevitable outcome of the survey process, it is arguable that there is an ethical responsibility incumbent on the survey organisation to provide information to the respondent about how the data are being used and by whom. In a panel where the relationship with the respondent lasts over a period of years, this becomes more important as the aim is to build a relationship of trust and demonstrate the value of continuing within the panel to sample members. On *Understanding Society*, respondents are kept fully informed about the Study via IWMs which give some key findings from the survey and details of how the data are being used. A named individual contact person at ISER with a Freephone telephone number and Freepost address are included on all correspondence with respondents so that they have a means of contacting us directly if required. There also is a respondent website.

13.1.4 WITHDRAWAL OF PARTICIPATION IN THE STUDY

No interview will be sought with respondents in subsequent waves where the whole household refuses in the first wave. If co-operation is refused after a first interview, the name of the respondent will be retained in the Study files unless the respondent demands its removal. An attempt will be made to convert refusals at the time of refusal, but no further requests for an interview will normally be made after two refusals across consecutive waves.

If, after interview, respondents inform the interviewer, ISER or the fieldwork agency that they are unwilling to be interviewed in subsequent waves, this will be formally recorded. The following rules will then be applied:

A. If the respondent categorically rejects any further approach, ISER may write to the respondent before the next fieldwork period requesting reconsideration. No approach will be made by an interviewer unless the respondent replies accepting a further interview.

B. Other respondents who indicate reluctance to continue co-operation but who do not categorically reject any further contact will be approached under standard fieldwork procedures.

13.1.5 WITHDRAWAL OF PERSONAL DATA (NAMES AND ADDRESSES)

Personal data may be held, *inter alia*, in the following form of names and addresses of all household members and names and addresses of people nominated to help ISER trace respondents if contact is lost.

Under the subject access provisions of the Data Protection Act 1998, a copy may be provided of such data held in respect of that individual, either free-of-charge or for a nominal fee. No such information can be given on behalf of other household members unless jointly requested.

Requests for amendment of names will be accepted. If the respondent demands *removal* of their name from the records, a letter will be sent asking for re-consideration. If permission to retain the name is not forthcoming, it will be removed from the active sample database within 28 days of the request. This applies only to the individual concerned unless joint applications are made. The withdrawal will apply to children who are not yet members of the sample in their own right where no responsible adult remains in the sample.

Withdrawal of a name means that the person's records will be known by ID number only. The respondent will be treated as a refusal in respect of future approaches. Neither rule (A) nor rule (B) will apply.

Names and addresses of stable contacts (see section 0) are covered by the Data Protection Act 1998. Requests (in writing) by respondents for removal of the information they have provided will be respected. The record will be destroyed or amended as necessary and the respondent will be informed in writing that this has taken place. If the *contact person* requests (in writing) the removal or alteration of information held on him or her, this will take place automatically, and that person will be informed in writing when the change has been made.

When a respondent requires the removal of *personal* data from the active sample database, this will be kept on paper to clarify problems which might subsequently arise, as long as this is in accordance with current legislation. However, an

instruction to destroy a questionnaire containing personal data (for example, a cover sheet) will be respected.

13.1.6 WITHDRAWAL OF SUBSTANTIVE DATA

The Data Protection Act 1998 does not provide a right of subject access to non-personal data kept for social research purposes, nor therefore to its amendment or withdrawal.

If a respondent requests the withdrawal of data held on a questionnaire but not yet transferred to computer or released to the UKDS, ISER will write to the respondent asking for reconsideration. If permission is not forthcoming the respondent's request will be complied with. This will apply to individuals only, unless joint applications are made. The outcome will be recorded as a refusal. In this case a request for destruction of a questionnaire or data record will be respected.

Withdrawal of data will not be undertaken after final preparation for deposit of the data in the UKDS (28 days before actual deposit).

If a proxy subject requests the withdrawal of data held on them, a letter will be sent explaining the nature of this information and requesting reconsideration. If permission is not granted, the withdrawal shall proceed within 28 days of the request, as long as the expiry of this period is not later than 28 days before deposit of the data in the UKDS. Individual data on the proxy subject will be removed from the dataset and the outcome recorded as a refusal.

13.1.7 INFORMED CONSENT TO RECORD LINKAGE

Ethical approval for both the way the consents are collected and the actual linkages are sought from the appropriate Research Ethics Committees (REC), including approval from the relevant statutory bodies for linkage to health databases.

Consents to data linkage do not expire. However, participants are free to withdraw their consent at any time and are provided with details of how to do this. If participants withdraw their consent and (a) if their survey data have not been linked to administrative records at that stage, their records will not be linked, (b) if their data have been linked already, no future data linkages will be made. Consenters are periodically reminded of any consent they have given, and consents to link to child records are confirmed with children once they take part in an adult interview.

It is important to note that no data from the Study is fed back to the providers of administrative data. Respondent’s dealings with the agencies holding the data are not affected by data linkage.

13.2 CONFIDENTIALITY

As noted earlier, *Understanding Society* is obliged to conform to UK legislation regarding the handling and use of personal data, in addition to which we voluntarily conform to professional codes of best practice. ISER has systems and procedures in place to ensure that any personal information such as names and addresses is held in strictly confidential conditions. This includes that all persons dealing with participants on a regular basis, whether by post or phone, must obviously take great care in recording or transmitting messages. Letters from participants are not to be left on desks when not being processed, not to be left open in pigeonholes, and

should be filed as soon as possible. When conveying refusal information from panel maintenance to fieldwork staff, always ensure that this gets to the right person and quickly. Email communications with participants should not contain personal information other than the participant's name and address. If a participant sends personal information other than his/her name and address in an email message, any reply to it should delete that information and reply in general terms that does not make direct reference to it or other personal information. Personal information should be communicated to individual participants only.

ISER's Code of Ethics forms part of our contract with our survey fieldwork agency. In addition, members of ISER staff handling personal data sign an ethical undertaking to ensure these data are treated correctly. The design of the panel where all household members are interviewed does raise some additional issues as the confidentiality of respondents must be respected within the responding household. Interviewers must not, even unwittingly, tell other members of a household what a given respondent has said. Partners, spouses and parents in particular may not always recognise this and feel they have a right to know what their partner or child has answered. In practice, preserving confidentiality within the household is often difficult especially where other members are present in the room where the interview is taking place. This in turn raises methodological questions about potential contamination effects and how responses to particular questions might vary depending on who is present at the time.

In preparing data for general release, we take steps to maintain the confidentiality of responses. Certain variables are excluded from the general release data, such as full date of birth and the most detailed occupation (SOC) and industry (SIC) codes. Household income has been top coded. Open or narrative text, e.g., names of schools or employers, is not released since it may indirectly identify individuals. Geographical identifiers below the level of regions (GOR) are also not included in the general release.

It is possible, however, to access some restricted resources upon application. The Study has a Data Access Strategy which sets out principles decisions on applications requesting access to restricted electronic data and applications to use the genetics data with survey information or the stored biological samples *Understanding Society* are considered by the cross research council body METADAC. The aim of our approach is to allow important research to proceed while minimising risks, particularly to study participants.

13.3 ETHICAL REVIEW OF STUDY PROTOCOLS

The *Understanding Society* study protocols and research programme are scrutinised by a number of research ethics committees to assure that ethical and legal obligations are respected at all times. Table 77 provides information on the various committees which have provided ethical approval of the *Understanding Society* study and its components as appropriate. It can be seen that, prior to signing off materials for the next wave of data collection, in addition to ethical review of the entire study by the University of Essex Ethics Committee, a number of components of the Study require additional ethical review. Namely, this applies to ethical review by a Medical Research Ethics Committee for components that involve collection and storage of

biomedical data as well as adding data from administrative records to the Study responses.

14. ACKNOWLEDGMENTS

People who have contributed to this first edition of the *Understanding Society* Quality Profile include, in alphabetical order, Vasiliki-Maria Agalioti-Sgompou, Michaela Benzeval, Nicolas Buck, Jonathan Burton, Alexandru Cernat, Glenn Dunn, Jane Ismail, Gundi Knies, Peter Lynn, Alita Nandi and Jakob Petersen. Lynn (2006) and Knies (2015) provided invaluable templates for this Quality Profile. Analysis presented in this Quality Profile is based on either the *Understanding Society* (2015), Wave 1-5 standard end-user licence data release[29] or the *Understanding Society* (2015), Wave 1-5 special licence data release.[30]

A very big "Thank you!" goes to the many people have contributed and continue to contribute to the unrelenting success and timely delivery of *Understanding Society*. They not only include the Study team and scientific advisors (see https://www.understandingsociety.ac.uk/about/people/), but also the many interviewers who collect the data on behalf of ISER and the fieldwork agencies, and most of all the Study members who provide their information.

[29] The format for bibliographic references is as follows: University of Essex. Institute for Social and Economic Research and National Centre for Social Research, *Understanding Society*: Wave 1-5, 2009-2014 [computer file]. 7th Edition. Colchester, Essex: UK Data Service [distributor], December 2015. SN: 6614, http://dx.doi.org/10.5255/UKDA-SN-6614-7.

[30] The format for bibliographic references is: University of Essex. Institute for Social and Economic Research and National Centre for Social Research, *Understanding Society*: Wave 1-5, 2009-2014 [computer file]. 5th Edition. Colchester, Essex: UK Data Service [distributor], December 2015. SN: 6931, http://dx.doi.org/10.5255/UKDA-SN-6931-5.

15. REFERENCES

Beatty, Paul, and Douglas Herrmann. 2002. "To answer or not to answer: Decision process related to survey item nonresponse." Pp. 71–85 in *Survey Nonresponse*, edited by Robert M. Groves, Donald A. Dillman, John L. Eltinge, and Roderick J. A. Little. New York: Wiley.

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

Berthoud, Richard, Laura Fumagalli, Peter Lynn, and Lucinda Platt. 2009. "Design of the Understanding Society Ethnic Minority Boost Sample." *Understanding Society Working Paper* 2009-02.

Buckner, John C. 1988. "The development of an Instrument to Measure Neighborhood Cohesion." *American Journal of Community Psychology* 16(6):771-91.

Burton, Jonathan. 2012. "Understanding Society Innovation Panel Wave 4: Results from Methodological Experiments." *Understanding Society Working Paper* 2012-06.

Buysse, Daniel J., Charles F. Reynolds Iii, Timothy H. Monk, Susan R. Berman, and David J. Kupfer. 1989. "The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research." *Psychiatry Research* 28(2):193-213.

Frick, Joachim R., Stephen P. Jenkins, Dean R. Lillard, Oliver Lipps, and Mark Wooden. 2007. "The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies." *Schmollers Jahrbuch* 127(4):627-54.

Gray, Michelle, Joanna d'Ardenne, Meera Balarajan, and S. C. Noah Uhrig. 2011. "Cognitive testing of Wave 3 Understanding Society questions." *Understanding Society Working Paper* 2011-03.

John, Oliver P., and Sanjay Srivastava. 1999. "The Big Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives." in *Handbook of Personality: Theory and Research*, edited by Oliver P. John and Lawrence A. Pervin. New York: Guilford Press.

Kaminska, Olena, and Peter Lynn. 2012a. "Combining refreshment or boost samples with an existing panel sample: challenges and solutions." *International Panel Survey Methods Workshop, Melbourne, Australia, July*.

—. 2012b. "Taking into account unknown eligibility in nonresponse correction." *International Workshop on Household Survey Nonresponse, Ottawa, Canada, September*.

Knies, Gundi. 2015. *Understanding Society –UK Household Longitudinal Study: Wave 1-5, 2009-2014, User Manual*. Colchester: University of Essex.

Knies, Gundi, and Jonathan Burton. 2014. "Analysis of Four Studies in a Comparative Framework Reveals: Health Linkage Consent Rates on British Cohort Studies Higher than on UK Household Panel Surveys." *BMC Medical Research Methodology* 14(125):1-12.

Knies, Gundi, Jonathan Burton, and Emanuela Sala. 2012. "Consenting to health record linkage: evidence from a multi-purpose longitudinal survey of a general population." *BMC Health Services Research* 12(1):52.

Knies, Gundi, and Seetha Menon. 2014. "Understanding Society: Waves 1-3, 2009-2012: Special Licence Access, Geographical Accessibility. The UKHLS-Accessibility Data File User Guide " *http://doc.ukdataservice.ac.uk/doc/7533/mrdoc/pdf/7533_ukhls_accessibility_userguide.pdf*.

Laurie, Heather. 2010. "Continuity and innovation in the design of Understanding Society: the UK Household Longitudinal Study." *Understanding Society Working Paper* 2012-12.

Lessof, Carli. 2009. "Ethical issues in longitudinal surveys." in *Methodology of Longitudinal Surveys*, edited by Peter Lynn. Chichester: Wiley.

Little, Roderick J. A., and Hong-Lin Su. 1989. "Item nonresponse in panel surveys." *Panel Surveys (ed.s) D. Kasprzyk, G. J. Duncan, G. Kalton and M. P. Singh). New York: Wiley*.

Lynn, Peter (Ed.). 2006. *Quality Profile: British Household Panel Survey*. Colchester: University of Essex.

Lynn, Peter, and Olena Kaminska. 2010. "Criteria for developing non-response weight adjustments for secondary users of complex longitudinal surveys." *International Workshop on Household Survey Nonresponse, Nürnberg, Germany, September*.

Lynn, Peter, S.C. Noah Uhrig, and Jonathan Burton. 2010. "Lessons from a randomised experiment with mixed-mode designs for a household panel survey." *Understanding Society Working Paper* 2010-03.

McFall, Stephanie L. 2013. "Understanding Society- UK Household Longitudinal Study: Cognitive Ability Measures." *Understanding Society User Manual*.

McFall, Stephanie L., Cara L. Booker, Jonathan Burton, and Anne Conolly. 2012. "Implementing the biosocial component of Understanding Society – nurse collection of biomeasures." *Understanding Society Working Paper* 2012-04.

McFall, Stephanie L., Anne Conolly, and Jonathan Burton. 2012. "Collecting biomarkers using trained interviewers. Lessons learned from a pilot study." *Understanding Society Working Paper* 2012-05.

McFall, Stephanie L., Alita Nandi, and Lucinda Platt. 2014. *Understanding Society: UK Household Longitudinal Study: User Guide to Ethnicity Research*. Colchester: University of Essex.

McFall, Stephanie L., Jakob Petersen, Olena Kaminska, and Peter Lynn. 2014. *Understanding Society: The UK Household Longitudinal Study. Waves 2 and 3 Nurse Health Assessment, 2010 – 2012. Guide to Nurse Health Assessment*. Colchester: University of Essex.

McKay, Steven, and Sharon Collard. 2003. "Developing deprivation questions for the Family Resources Survey." *Personal Finance Research Centre Working Paper* 13.

Nandi, Alita, and Lucinda Platt. 2009. "Developing ethnic identity questions for Understanding Society, the UK Household Longitudinal Study." *Understanding Society Working Paper* 2009-03.

Sadig, Husam E. 2014a. "Non-response subgroup-tailored weighting: the choice of variables and the set of respondents used to estimate the weighting model." *ISER Working Paper* 2014(36).

—. 2014b. "Unknown eligibility whilst weighting for non-response: the puzzle of who has died and who is still alive?" *ISER Working Paper* 2014(35).

Sala, Emanuela, Jonathan Burton, and Gundi Knies. 2012. "Correlates of obtaining informed consent to data linkage: Respondent, interview and interviewer characteristics." *Sociological Methods & Research* 41(3):414-39.

Sala, Emanuela, Gundi Knies, and Jonathan Burton. 2014. "Propensity to consent to data linkage: experimental evidence on the role of three survey design features in a UK longitudinal panel." *International Journal of Social Research Methodology* 17(5):455-73.

Schmiade, Nicole , C. Katharina Spiess, and Wolfgang Tietze. 2008. "Zur Erhebung des adaptiven Verhaltens von zwei- und dreijährigen Kindern im Sozio-oekonomischen Panel (SOEP) " *SOEPpapers on Multidisciplinary Panel Data Research* 116.

Spritzer, K. 2004. "SAS Code for Scoring 12-Item Health Survey Version 2.0." *http://gim.med.ucla.edu/FacultyPages/Hays/utils/sf12v2-1.sas*.

Vannieuwenhuyze, Jorre T.A., and Peter Lynn. 2014. "Measurement effects between CAPI and Web questionnaires in the UK Household Longitudinal Study." *Understanding Society Working Paper* 2014-01.

Warr, Peter. 1990. "The measurement of well-being and other aspects of mental health." *Journal of Occupational Psychology* 63:193-210.

16. TABLES AND FIGURES

Table 1: Overview of multi-item scales

Scale name[1]	Description/Notes (also see variable notes in the online documentation)	Variable stem name	Occurrence (wave)
Adult material deprivation	Specialised surveys such as the Poverty and Social Exclusion in Britain (PSE) have pioneered direct measurement of poverty on the basis of deprivation indicators. *Understanding Society* includes a short 8-item set of questions recommended by DWP for inclusion in the FRS after extensive analysis of existing longer multi-item scales, see McKay and Collard (2003). Summary scores not computed.	w_matdep*	1,2,4
Big Five Inventory (adapted)	Big Five Inventory is a 44-item scale measuring agreeableness, conscientiousness, extraversion, neuroticism and openness developed by John and Srivastava (1999). The version carried in *Understanding Society* includes 15 items (i.e., 3 for each personality trait) asking for respondents views about themselves. Response categories run from [1] "does not apply to me at all" to [7] "applies to me perfectly". Response categories in the original version run from [1] "disagree strongly" to [5] "agree strongly". Summary scores are computed.	w_big5a_dv w_big5c_dv w_big5e_dv w_big5n_dv w_big5o_dv	3
Body Mass Index (BMI)	The Body Mass Index (BMI) is defined as body weight in kg divided by the square of the respondent's height in metres. Summary scores computed for adult respondents (in Wave 1) and for participants in the Nurse visit (Wave 2 and 3). Also collected as part of the youth questionnaire (in Wave 2 and 4).	w_bmi_dv w_bmi w_ypbmi_dv	1 2,3 2,4
Buckner's Neighbourhood Cohesion Scale (adapted)	Scale measuring psychological sense of community, attraction to neighbourhood, and social interaction within neighbourhood. Eight items (2,4,7,11-14, and 17) out of 18 are included, see Buckner (1988, p. 783). Summary scores not computed.	w_scopngbh*	1,3
Child material deprivation	Direct measurement of child poverty on the basis of deprivation indicators is relatively new. *Understanding Society* includes a short 9-item set of questions recommended by DWP for inclusion in the FRS after extensive analysis of the few existing multi-item scales, see McKay and Collard (2003). Summary scores not computed.	w_cdephave* w_cdepdo* w_cplay w_cdeptrp w_cdelply	1,2,4
Cognitive ability	The cognitive development module was sourced from the US American Health and Retirement Study. Summary scores for practical numerical	w_cgna_dv w_cgns1sc10_dv	3

	knowledge, fluid reasoning, working memory, verbal fluency, delayed and immediate memory were produced following precedents on the HRS (cf. McFall 2013).	w_cgns2sc10_dv w_cgs7ca_dv w_cgvfw_dv w_cgwrd_dv w_cgwri_dv	
General Health Questionnaire (GHQ)	The GHQ is a screening questionnaire aimed at detecting those with a diagnosable psychiatric disorder. For further information, see http://www.gl-assessment.co.uk/products/general-health-questionnaire-0. *Understanding Society* includes the 12-item short form. Summaries are computed using Likert scoring and caseness scoring.	w_scghq1_dv w_scghq2_dv	1,2,3,4,5
Pittsburgh Sleep Quality Index (PSQI)	This is a self-report of sleep quality within the past month developed by (Buysse et al. 1989). For items and numbering see http://www.brandeis.edu/roybal/docs/PSQI_website_PDF.pdf *Understanding Society* uses items 5a, 5b and 5e adding the response category "more than once each night", plus items 4, 6, 7 and the rating of sleep quality overall (without alterations). Collected as part of the adult self-completion in Wave 1, moved to main interview iwht adults in Wave 4. Summary scores not computed.	*hrs_slph *tslp_30m *tslp_wak *tslp_cgh *med_slp *tsta_awk *slp_qual	1,4
Short Form 12 (SF-12) Health Questionnaire	SF-12 was developed for the Medical Outcomes Study (MOS), a multi-year study of patients with chronic mental and physical health conditions. Summary scores for physical and mental health are computed using scoring instructions provided by Spritzer (2004).	w_sf12pcs_dv w_sf12mcs_dv	1,2,3,4,5
Strengths and Difficulties Questionnaire (SDQ)	The SDQ is a brief behavioural screening questionnaire for children. It is based on 25 items measuring positive and negative behaviours in children. We compute five subscales, measuring emotional symptoms, conduct problems, hyperactivity, peer relationship problems, and prosocial behaviour. An overall Total Difficulties Score is calculated over the first four scales. Items included bi-annually in the youth questionnaire and, from Wave 3 onwards, in the child development module about children aged 5 and 8.	w_ypsdqes_dv w_ypsdqcp_dv w_ypsdqha_dv w_ypsdqpp_dv w_ypsdqps_dv w_ypsdqtd_dv w_chsdqes_dv w_chsdqcp_dv w_chsdqha_dv w_chsdqpp_dv w_chsdqps_dv w_chsdqtd_dv	1,3,5 3,4,5

Vineland Adaptive Behaviour Scale (VABS)	VABS is based on 20-items designed to measure child competencies in the areas of language, general skills, motoric skills, and social interactions. The items carried *Understanding Society* as part of the child development module for children aged 3 were sourced from the German SOEP, where the instrument achieved acceptable levels of construct validity when measured in children aged 2-3, see Schmiade, Spiess and Tietze (2008). Summary scores not computed.	w_cdvl*	3,4,5
Warr's job-related wellbeing scale (Skills Survey)	Two job-related wellbeing scales, originally devised by Warr (1990), were included in the Skills Survey to help monitor the wellbeing of the working population. The *Understanding Society* questionnaire sourced the items measuring work-related anxiety (from Warr's "Anxiety-Contentment" scale) and work-related depression (from Warr's Depression-Enthusiasm scale). Summary scores for these subscales are computed.	w_jwbs1_dv w_jwbs2_dv	2,4
Warwick and Edinburg Mental Wellbeing Scale (WEMWBS)	WEMWBS is a scale of 14 positively worded items, with five response categories, for assessing a population´s mental wellbeing. *Understanding Society* carries a short 7-item version, the (SWEMWBS), which has been shown to provide greater construct validity. For further information, see http://www.healthscotland.com/scotlands-health/population/Measuring-positive-mental-health.aspx	w_swemwbs_dv	1, 4

Notes: 1 Listed in alphabetical order. Listing may be incomplete.

Table 2: Timing of data collection start

Year	Quarter	Survey wave/component
2008	Q1	
	Q2	
	Q3	BHPS W18
	Q4	
2009	Q1	W1, Y1*
	Q2	
	Q3	
	Q4	
2010	Q1	W1, Y2
	Q2	
	Q3	W2, Y1**
	Q4	
2011	Q1	W2, Y2
	Q2	
	Q3	Health Assessment GPS
	Q4	
2012	Q1	W3, Y1*
	Q2	
	Q3	Health Assess't BHPS
	Q4	
2013	Q1	W3, Y2
	Q2	
	Q3	W4, Y1*
	Q4	
2014	Q1	W4, Y2
	Q2	
	Q3	W5, Y1*
	Q4	W5, Y2

Notes:
* Northern Ireland sample interviewed in year 1 of each wave.
** BHPS becomes sample component in Wave 2, year 1.

Table 3: Fieldwork dates

Sample	Interviewing started	90% of interviewing completed	Interviewing completed
Wave 1 Q1	08-01-2009	30-03-2009	16-11-2009
Wave 1 Q2	01-04-2009	30-06-2009	28-01-2010
Wave 1 Q3	01-07-2009	05-10-2009	13-03-2010
Wave 1 Q4	01-10-2009	18-01-2010	23-03-2010
Wave 1 Q5	02-01-2010	10-04-2010	23-07-2010
Wave 1 Q6	07-04-2010	12-07-2010	14-10-2010
Wave 1 Q7	02-07-2010	11-10-2010	09-01-2011
Wave 1 Q8	20-09-2010	27-01-2011	10-03-2011
Wave 2 Q1	10-01-2010	06-04-2010	05-09-2010
Wave 2 Q2	28-03-2010	08-07-2010	27-01-2011
Wave 2 Q3	26-06-2010	09-10-2010	30-01-2011
Wave 2 Q4	29-09-2010	21-01-2011	06-04-2011
Wave 2 Q5	08-01-2011	06-04-2011	10-08-2011
Wave 2 Q6	07-04-2011	06-07-2011	12-10-2011
Wave 2 Q7	08-07-2011	10-10-2011	04-01-2012
Wave 2 Q8	06-10-2011	11-01-2012	27-03-2012
Wave 3 Q1	02-01-2011	11-04-2011	11-08-2011
Wave 3 Q2	31-03-2011	09-07-2011	07-10-2011
Wave 3 Q3	27-06-2011	12-10-2011	12-01-2012
Wave 3 Q4	29-09-2011	12-01-2012	21-05-2012
Wave 3 Q5	09-01-2012	16-04-2012	14-09-2012
Wave 3 Q6	05-04-2012	14-07-2012	30-11-2012
Wave 3 Q7	29-06-2012	13-10-2012	29-01-2013
Wave 3 Q8	06-10-2012	15-01-2013	22-04-2013
Wave 4 Q1	03-01-2012	19-04-2012	10-09-2012
Wave 4 Q2	02-04-2012	11-07-2012	19-12-2012
Wave 4 Q3	28-06-2012	17-10-2012	03-02-2013
Wave 4 Q4	02-10-2012	16-01-2013	18-04-2012
Wave 4 Q5	08-01-2013	24-04-2013	16-07-2013
Wave 4 Q6	06-04-2013	23-07-2013	20-10-2013
Wave 4 Q7	08-07-2013	25-10-2013	07-02-2014
Wave 4 Q8	05-10-2013	25-01-2014	19-05-2014
Wave 5 Q1	08-01-2013	21-04-2013	24-07-2013
Wave 5 Q2	07-04-2013	22-07-2013	19-10-2013
Wave 5 Q3	06-07-2013	25-10-2013	13-02-2014
Wave 5 Q4	05-10-2013	24-01-2014	22-05-2014
Wave 5 Q5	06-01-2014	24-04-2014	12-08-2014
Wave 5 Q6	04-04-2014	28-07-2014	07-11-2014
Wave 5 Q7	06-07-2014	30-10-2014	17-02-2015
Wave 5 Q8	06-10-2014	26-01-2015	13-04-2015

Note: Dates in this table refer to the completion of household interviews.

Table 4: Wave-to-wave interviewer continuity

Waves	Proportion issued to same interviewer at both waves	Base (number of persons issued to field at both waves)
1 and 2	54.3%	76,793
2 and 3	61.3%	87,219
3 and 4	69.0%	77,005
4 and 5	81.9%	70,975

Table 5: Household eligibility status in Wave 1, by sample

	Number	Percent
GPS		
Eligible	45325	89.7
Demolished/derelict/in construction	358	0.7
Institution/not private	601	1.2
No resident household/Other	4255	8.4
Total	50539	100.0
EMBS		
Eligible	40748	91.0
Demolished/derelict/in construction	380	0.8
Institution/not private	806	1.8
No resident household/Other	2835	6.3
Total	44769	100.0
Total		
Eligible	86073	90.3
Demolished/derelict/in construction	738	0.8
Institution/not private	1407	1.5
No resident household/Other	7090	7.4
Total	95308	100.0

Table 6: EMBS: Household screen outcome, Wave 1

	Number	Percent
Eligible	7616	18.7
Not eligible	30734	75.4
Screen refused	1057	2.6
Screen non-contact	1140	2.8
Other screen non-response	201	0.5
Total	40748	100.0

Table 7: Household outcome if eligible, Wave 1 by sample

	Number	Percent
GPS		
Complete household	19128	42.2
Partial household	6879	15.2
Household grid	82	0.2
Refusal	15250	33.6
Non-contact	3645	8.0
Other non-response	341	0.8
Total	45325	100.0
EMBS		
Complete household	2566	33.7
Partial household	1459	19.2
Household grid	55	0.7
Refusal	1971	25.9
Non-contact	1478	19.4
Other non-response	87	1.1
Total	7616	100.0
Total		
Complete household	21694	41.0
Partial household	8338	15.7
Household grid	137	0.3
Refusal	17221	32.5
Non-contact	5123	9.7
Other non-response	428	0.8
Total	52941	100.0

Table 8: Individual response outcome, Wave 1 by sample (persons over 15 in enumerated households)

	Number	Percent
GPS		
Full interview	41047	81.8
Proxy interview	2627	5.2
Refusal	3352	6.7
Other non-interview	3173	6.3
Total	50199	100.0
EMBS		
Full interview	6685	72.4
Proxy interview	635	6.9
Refusal	803	8.7
Other non-interview	1114	12.0
Total	9234	100.0
Total		
Full interview	47732	80.3
Proxy interview	3262	5.5
Refusal	4155	7.0
Other non-interview	4287	7.2
Total	59436	100.0

Table 9: Adult self-completion outcome, Wave 1 by sample (conditional on individual response)

	Number	Percent
GPS		
Adult self-completion	35857	87.4
No self-completion	5190	12.6
Total	41047	100.0
EMBS		
Adult self-completion	4656	69.6
No self-completion	2029	30.4
Total	6685	100.0
Total		
Adult self-completion	40513	84.9
No self-completion	7219	15.1
Total	47732	100.0

672 respondents completed self-completion but not CAPI interview.

Table 10: Youth response outcome, Wave 1 by sample (persons aged 10-15 in enumerated households)

	Number	Percent
GPS		
Youth interview	3995	77.0
Youth non-response	1195	23.0
Total	5190	100.0
EMBS		
Youth interview	904	62.9
Youth non-response	533	37.1
Total	1437	100.0
Total		
Youth interview	4899	73.9
Youth non-response	1728	26.1
Total	6627	100.0

Table 11: Household eligibility status by sample, Wave 2

	Number	Percent
GPS		
Eligible	27209	98.5
Whole household deceased	116	0.4
Other ineligible	297	1.1
Total	27622	100.0
EMBS		
Eligible	4107	96.0
Whole household deceased	4	0.1
Other ineligible	168	3.9
Total	4279	100.0
BHPS sample		
Eligible	8558	95.2
Whole household deceased	104	1.2
Other ineligible	330	3.7
Total	8992	100.0
Total		
Eligible	39874	97.5
Whole household deceased	224	0.5
Other ineligible	795	1.9
Total	40893	100.0

Table 12: Household outcome by sample if eligible, Wave 2

	Number	Percent
GPS		
Complete household	16058	59.0
Partial household	4950	18.2
Household grid	19	0.1
Refusal	3129	11.5
Non-contact	3020	11.1
Other non-response	33	0.1
Total	27209	100.0
EMBS		
Complete household	1807	44.0
Partial household	976	23.8
Household grid	9	0.2
Refusal	514	12.5
Non-contact	787	19.2
Other non-response	14	0.3
Total	4107	100.0
BHPS sample		
Complete household	5430	63.4
Partial household	1246	14.6
Household grid	16	0.2
Refusal	1060	12.4
Non-contact	800	9.3
Other non-response	6	0.1
Total	8558	100.0
Total		
Complete household	23295	58.4
Partial household	7172	18.0
Household grid	44	0.1
Refusal	4703	11.8
Non-contact	4607	11.6
Other non-response	53	0.1
Total	39874	100.0

Table 13: GPS: Individual response, Wave 2 by type of sample member (persons over 15 in enumerated households)

	Number	Percent
OSM		
Full interview	33509	66.6
Proxy interview	2636	5.2
Refusal	8135	16.2
Other non-interview	6011	12.0
Total	50291	100.0
PSM		
Full interview	23	54.8
Proxy interview	7	16.7
Refusal	6	14.3
Other non-interview	6	14.3
Total	42	100.0
TSM		
Full interview	619	50.1
Proxy interview	169	13.7
Refusal	274	22.2
Other non-interview	173	14.0
Total	1235	100.0
Total		
Full interview	34151	66.2
Proxy interview	2812	5.5
Refusal	8415	16.3
Other non-interview	6190	12.0
Total	51568	100.0

Table 14: BHPS sample: Individual response, Wave 2 by type of sample member (persons over 15 in enumerated households)

	Number	Percent
-9		
Full interview	0	0.0
Proxy interview	0	0.0
Refusal	1916	65.3
Other non-interview	1017	34.7
Total	2933	100.0
OSM		
Full interview	9726	87.4
Proxy interview	286	2.6
Refusal	606	5.4
Other non-interview	504	4.5
Total	11122	100.0
PSM		
Full interview	734	81.5
Proxy interview	59	6.5
Refusal	56	6.2
Other non-interview	52	5.8
Total	901	100.0
TSM		
Full interview	1126	70.2
Proxy interview	105	6.6
Refusal	221	13.8
Other non-interview	151	9.4
Total	1603	100.0
Total		
Full interview	11586	70.0
Proxy interview	450	2.7
Refusal	2799	16.9
Other non-interview	1724	10.4
Total	16559	100.0

Table 15: EMBS: Individual response, Wave 2 by type of sample member (persons over 15 in enumerated households)

	Number	Percent
OSM		
Full interview	4495	53.4
Proxy interview	558	6.6
Refusal	1568	18.6
Other non-interview	1793	21.3
Total	8414	100.0
PSM		
Full interview	1	12.5
Proxy interview	2	25.0
Refusal	4	50.0
Other non-interview	1	12.5
Total	8	100.0
TSM		
Full interview	482	53.4
Proxy interview	60	6.7
Refusal	170	18.8
Other non-interview	190	21.1
Total	902	100.0
Total		
Full interview	4978	53.4
Proxy interview	620	6.6
Refusal	1742	18.7
Other non-interview	1984	21.3
Total	9324	100.0

Table 16: GPS: Individual response, Wave 2 by outcome in Wave 1 (OSM aged over 15 in enumerated households)

	Number	Percent
Adult interview Wave 1		
Full interview	31288	77.3
Proxy interview	794	2.0
Refusal	4611	11.4
Other non-interview	3765	9.3
Total	40458	100.0
Proxy interview Wave 1		
Full interview	606	23.6
Proxy interview	982	38.2
Refusal	565	22.0
Other non-interview	417	16.2
Total	2570	100.0
Youth interview Wave 1		
Full interview	447	67.8
Proxy interview	48	7.3
Refusal	88	13.4
Other non-interview	76	11.5
Total	659	100.0
Interview non-response Wave 1		
Full interview	1166	17.7
Proxy interview	811	12.3
Refusal	2869	43.5
Other non-interview	1743	26.5
Total	6589	100.0
Total		
Full interview	33507	66.6
Proxy interview	2635	5.2
Refusal	8133	16.2
Other non-interview	6001	11.9
Total	50276	100.0

10 respondents who completed the child questionnaire in Wave 1 were excluded

Table 17: EMBS: Individual response, Wave 2 by outcome in Wave 1 (OSM aged over 15 in enumerated households)

	Number	Percent
Adult interview Wave 1		
Full interview	3919	66.3
Proxy interview	187	3.2
Refusal	763	12.9
Other non-interview	1046	17.7
Total	5915	100.0
Proxy interview Wave 1		
Full interview	150	25.1
Proxy interview	162	27.1
Refusal	163	27.3
Other non-interview	122	20.4
Total	597	100.0
Youth interview Wave 1		
Full interview	91	59.5
Proxy interview	17	11.1
Refusal	26	17.0
Other non-interview	19	12.4
Total	153	100.0
Interview non-response Wave 1		
Full interview	334	19.1
Proxy interview	192	11.0
Refusal	616	35.2
Other non-interview	606	34.7
Total	1748	100.0
Total		
Full interview	4494	53.4
Proxy interview	558	6.6
Refusal	1568	18.6
Other non-interview	1793	21.3
Total	8413	100.0

One respondent classified as a child in Wave 1 was excluded

Table 18: Adult self-completion outcome, Wave 2 by sample (conditional on Wave 2 individual response)

	Number	Percent
GPS		
Adult self-completion	30298	88.7
No self-completion	3853	11.3
Total	34151	100.0
EMBS		
Adult self-completion	3476	69.8
No self-completion	1502	30.2
Total	4978	100.0
BHPS sample		
Adult self-completion	10359	89.4
No self-completion	1227	10.6
Total	11586	100.0
Total		
Adult self-completion	44133	87.0
No self-completion	6582	13.0
Total	50715	100.0

672 respondents completed self-completion but not CAPI interview.

Table 19: GPS: Self-completion outcome, Wave 2 by Wave 1 self-completion outcome (conditional on being an OSM)

	Number	Percent
Adult self-completion Wave 1		
Adult self-completion	25760	91.5
No self-completion	2400	8.5
Total	28160	100.0
No self-completion Wave 1		
Adult self-completion	3577	73.0
No self-completion	1325	27.0
Total	4902	100.0
Youth self-completion Wave 1		
Adult self-completion	418	93.5
No self-completion	29	6.5
Total	447	100.0
Total		
Adult self-completion	29755	88.8
No self-completion	3754	11.2
Total	33509	100.0

672 respondents completed self-completion but not CAPI interview.

Table 20: EMBS: Self-completion outcome, Wave 2 by Wave 1 self-completion outcome (conditional on being an OSM)

	Number	Percent
Adult self-completion Wave 1		
Adult self-completion	2255	79.0
No self-completion	598	21.0
Total	2853	100.0
No self-completion Wave 1		
Adult self-completion	777	50.1
No self-completion	774	49.9
Total	1551	100.0
Youth self-completion Wave 1		
Adult self-completion	80	87.9
No self-completion	11	12.1
Total	91	100.0
Total		
Adult self-completion	3112	69.2
No self-completion	1383	30.8
Total	4495	100.0

672 respondents completed self-completion but not CAPI interview.

Table 21: Youth interview outcome, Wave 2 by type of sample member (persons aged 10-15 in enumerated households Wave 2)

	Number	Percent
GPS		
Youth interview	3238	78.3
Youth non-response	900	21.7
Total	4138	100.0
EMBS		
Youth interview	664	63.5
Youth non-response	382	36.5
Total	1046	100.0
BHPS sample		
Youth interview	1117	82.0
Youth non-response	245	18.0
Total	1362	100.0
Total		
Youth interview	5019	76.7
Youth non-response	1527	23.3
Total	6546	100.0

Table 22: Youth interview outcome, Wave 2 by youth outcome in Wave 1 (OSM aged 10-15 in enumerated households Wave 2)

	Number	Percent
Youth interview Wave 1		
Youth interview	2775	83.4
Youth non-response	551	16.6
Total	3326	100.0
Youth non-response Wave 1		
Youth interview	494	52.9
Youth non-response	440	47.1
Total	934	100.0
Child under 10 Wave 1		
Youth interview	557	70.6
Youth non-response	232	29.4
Total	789	100.0
Total		
Youth interview	3826	75.8
Youth non-response	1223	24.2
Total	5049	100.0

Table 23: Household eligibility status by sample, Wave 3

	Number	Percent
GPS		
Eligible	25130	97.2
Whole household deceased	110	0.4
Other ineligible	622	2.4
Total	25862	100.0
EMBS		
Eligible	3664	94.9
Whole household deceased	6	0.2
Other ineligible	189	4.9
Total	3859	100.0
BHPS sample		
Eligible	7685	95.8
Whole household deceased	44	0.5
Other ineligible	295	3.7
Total	8024	100.0
Total		
Eligible	36479	96.6
Whole household deceased	160	0.4
Other ineligible	1106	2.9
Total	37745	100.0

Table 24: Household outcome by sample if eligible, Wave 3

	Number	Percent
GPS		
Complete household	14304	56.9
Partial household	4631	18.4
Household grid	34	0.1
Refusal	3190	12.7
Non-contact	2692	10.7
Other non-response	279	1.1
Total	25130	100.0
EMBS		
Complete household	1571	42.9
Partial household	954	26.0
Household grid	11	0.3
Refusal	461	12.6
Non-contact	631	17.2
Other non-response	36	1.0
Total	3664	100.0
BHPS sample		
Complete household	4985	64.9
Partial household	1270	16.5
Household grid	22	0.3
Refusal	670	8.7
Non-contact	677	8.8
Other non-response	61	0.8
Total	7685	100.0
Total		
Complete household	20860	57.2
Partial household	6855	18.8
Household grid	67	0.2
Refusal	4321	11.8
Non-contact	4000	11.0
Other non-response	376	1.0
Total	36479	100.0

Table 25: GPS: Individual response, Wave 3 by type of sample member (persons over 15 in enumerated households)

	Number	Percent
OSM		
Full interview	29728	65.4
Proxy interview	2309	5.1
Refusal	8319	18.3
Other non-interview	5103	11.2
Total	45459	100.0
PSM		
Full interview	57	50.9
Proxy interview	24	21.4
Refusal	21	18.8
Other non-interview	10	8.9
Total	112	100.0
TSM		
Full interview	890	44.8
Proxy interview	251	12.6
Refusal	492	24.7
Other non-interview	355	17.9
Total	1988	100.0
Total		
Full interview	30675	64.5
Proxy interview	2584	5.4
Refusal	8832	18.6
Other non-interview	5468	11.5
Total	47559	100.0

Table 26: EMBS: Individual response, Wave 3 by type of sample member (persons over 15 in enumerated households)

	Number	Percent
OSM		
Full interview	3977	53.4
Proxy interview	567	7.6
Refusal	1451	19.5
Other non-interview	1449	19.5
Total	7444	100.0
PSM		
Full interview	7	43.8
Proxy interview	3	18.8
Refusal	3	18.8
Other non-interview	3	18.8
Total	16	100.0
TSM		
Full interview	458	50.0
Proxy interview	103	11.2
Refusal	181	19.8
Other non-interview	174	19.0
Total	916	100.0
Total		
Full interview	4442	53.0
Proxy interview	673	8.0
Refusal	1635	19.5
Other non-interview	1626	19.4
Total	8376	100.0

Table 27: BHPS sample: Individual response, Wave 3 by type of sample member (persons over 15 in enumerated households)

	Number	Percent
OSM		
Full interview	8998	74.6
Proxy interview	367	3.0
Refusal	1550	12.8
Other non-interview	1151	9.5
Total	12066	100.0
PSM		
Full interview	696	69.5
Proxy interview	58	5.8
Refusal	128	12.8
Other non-interview	119	11.9
Total	1001	100.0
TSM		
Full interview	1089	58.7
Proxy interview	154	8.3
Refusal	378	20.4
Other non-interview	233	12.6
Total	1854	100.0
Total		
Full interview	10783	72.3
Proxy interview	579	3.9
Refusal	2056	13.8
Other non-interview	1503	10.1
Total	14921	100.0

Table 28: GPS: Individual response, Wave 3 by outcome in Wave 2 (OSM over 15 in enumerated households)

	Number	Percent
Adult interview Wave 2		
Full interview	26993	81.9
Proxy interview	711	2.2
Refusal	3276	9.9
Other non-interview	1969	6.0
Total	32949	100.0
Proxy interview Wave 2		
Full interview	558	21.7
Proxy interview	1017	39.6
Refusal	671	26.1
Other non-interview	324	12.6
Total	2570	100.0
Youth interview Wave 2		
Full interview	358	70.6
Proxy interview	45	8.9
Refusal	65	12.8
Other non-interview	39	7.7
Total	507	100.0
Interview non-response Wave 2		
Full interview	1817	19.3
Proxy interview	535	5.7
Refusal	4302	45.7
Other non-interview	2766	29.4
Total	9420	100.0
Total		
Full interview	29727	65.4
Proxy interview	2308	5.1
Refusal	8316	18.3
Other non-interview	5099	11.2
Total	45450	100.0

Table 29: EMBS: Individual response, Wave 3 by outcome in Wave 2 (persons over 15 in enumerated households)

	Number	Percent
Adult interview Wave 2		
Full interview	3526	73.8
Proxy interview	205	4.3
Refusal	494	10.3
Other non-interview	552	11.6
Total	4777	100.0
Proxy interview Wave 2		
Full interview	149	25.1
Proxy interview	212	35.8
Refusal	110	18.5
Other non-interview	122	20.6
Total	593	100.0
Youth interview Wave 2		
Full interview	65	70.7
Proxy interview	9	9.8
Refusal	8	8.7
Other non-interview	10	10.9
Total	92	100.0
Interview non-response Wave 2		
Full interview	592	22.3
Proxy interview	197	7.4
Refusal	971	36.6
Other non-interview	890	33.6
Total	2650	100.0
Total		
Full interview	4332	53.4
Proxy interview	623	7.7
Refusal	1583	19.5
Other non-interview	1574	19.4
Total	8112	100.0

Table 30: BHPS sample: Individual response, Wave 3 by outcome in Wave 2 (OSM over 15 in enumerated households)

	Number	Percent
Adult interview Wave 2		
Full interview	8397	87.8
Proxy interview	155	1.6
Refusal	536	5.6
Other non-interview	476	5.0
Total	9564	100.0
Proxy interview Wave 2		
Full interview	66	24.8
Proxy interview	113	42.5
Refusal	52	19.5
Other non-interview	35	13.2
Total	266	100.0
Youth interview Wave 2		
Full interview	136	75.6
Proxy interview	15	8.3
Refusal	15	8.3
Other non-interview	14	7.8
Total	180	100.0
Interview non-response Wave 2		
Full interview	395	19.3
Proxy interview	83	4.1
Refusal	944	46.1
Other non-interview	626	30.6
Total	2048	100.0
Total		
Full interview	8994	74.6
Proxy interview	366	3.0
Refusal	1547	12.8
Other non-interview	1151	9.5
Total	12058	100.0

Table 31: Adult self-completion outcome, Wave 3 by sample (conditional on Wave 3 individual response)

	Number	Percent
GPS		
Adult self-completion	27502	89.7
No self-completion	3173	10.3
Total	30675	100.0
EMBS		
Adult self-completion	3534	79.6
No self-completion	908	20.4
Total	4442	100.0
BHPS sample		
Adult self-completion	9655	89.5
No self-completion	1128	10.5
Total	10783	100.0
Total		
Adult self-completion	40691	88.7
No self-completion	5209	11.3
Total	45900	100.0

672 respondents completed self-completion but not CAPI interview

Table 32: GPS: Self-completion outcome, Wave 3 by Wave 2 self-completion outcome (conditional on being an OSM)

	Number	Percent
Adult self-completion Wave 2		
Adult self-completion	22266	91.5
No self-completion	2070	8.5
Total	24336	100.0
No self-completion Wave 2		
Adult self-completion	2942	79.1
No self-completion	779	20.9
Total	3721	100.0
Youth self-completion Wave 2		
Adult self-completion	339	94.7
No self-completion	19	5.3
Total	358	100.0
Total		
Adult self-completion	25547	89.9
No self-completion	2868	10.1
Total	28415	100.0

672 respondents completed self-completion but not CAPI interview.

Table 33: EMBS: Self-completion outcome, Wave 3 by Wave 2 self-completion outcome (conditional on being an OSM)

	Number	Percent
Adult self-completion Wave 2		
Adult self-completion	1983	84.8
No self-completion	355	15.2
Total	2338	100.0
No self-completion Wave 2		
Adult self-completion	847	69.9
No self-completion	364	30.1
Total	1211	100.0
Youth self-completion Wave 2		
Adult self-completion	59	92.2
No self-completion	5	7.8
Total	64	100.0
Total		
Adult self-completion	2889	80.0
No self-completion	724	20.0
Total	3613	100.0

672 respondents completed self-completion but not CAPI interview.

Table 34: BHPS sample: Self-completion outcome, Wave 3 by Wave 2 self-completion outcome (conditional on being an OSM)

	Number	Percent
Adult self-completion Wave 2		
Adult self-completion	7055	92.7
No self-completion	557	7.3
Total	7612	100.0
No self-completion Wave 2		
Adult self-completion	547	56.8
No self-completion	416	43.2
Total	963	100.0
Youth self-completion Wave 2		
Adult self-completion	129	94.9
No self-completion	7	5.1
Total	136	100.0
Total		
Adult self-completion	7731	88.7
No self-completion	980	11.3
Total	8711	100.0

672 respondents completed self-completion but not CAPI interview.

Table 35: Youth interview outcome, Wave 3 by type of sample member (persons aged 10-15 in enumerated households Wave 3)

	Number	Percent
GPS		
Youth interview	2799	75.6
Youth non-response	902	24.4
Total	3701	100.0
EMBS		
Youth interview	617	63.9
Youth non-response	348	36.1
Total	965	100.0
BHPS sample		
Youth interview	1011	78.4
Youth non-response	278	21.6
Total	1289	100.0
Total		
Youth interview	4427	74.3
Youth non-response	1528	25.7
Total	5955	100.0

Table 36: Youth interview outcome, Wave 3 by youth outcome in Wave 2 (OSM aged 10-15 in enumerated households Wave 3 and also enumerated Wave 2)

	Number	Percent
Youth interview Wave 2		
Youth interview	2975	83.9
Youth non-response	570	16.1
Total	3545	100.0
Youth non-response Wave 2		
Youth interview	418	48.6
Youth non-response	442	51.4
Total	860	100.0
Child under 10 Wave 2		
Youth interview	659	71.9
Youth non-response	257	28.1
Total	916	100.0
Total		
Youth interview	4052	76.2
Youth non-response	1269	23.8
Total	5321	100.0

Table 37: Nurse Health Assessment outcome, Wave 2 (conditional on individual Wave 2 response)

	Number	Percent
Interview and blood sample	9948	36.8
Interview, no blood sample	5685	21.1
Refusal	7622	28.2
Non-contact	1887	7.0
Other non-interview	1855	6.9
Total (eligible GPS)	26997	100.0

Table 38: Nurse Health Assessment outcome, Wave 3 (conditional on individual Wave 3 response)

	Number	Percent
Interview and blood sample	3381	37.9
Interview, no blood sample	1670	18.7
Refusal	1728	19.4
Non-contact	521	5.8
Other non-interview	1618	18.1
Total (eligible BHPS sample)	8918	100.0

Table 39: Household eligibility status by sample, Wave 4

	Number	Percent
GPS		
Eligible	21946	97.1
Whole household deceased	117	0.5
Other ineligible	528	2.3
Total	22591	100.0
EMBS		
Eligible	3229	96.8
Whole household deceased	4	0.1
Other ineligible	103	3.1
Total	3336	100.0
BHPS sample		
Eligible	6968	96.6
Whole household deceased	41	0.6
Other ineligible	207	2.9
Total	7216	100.0
Total		
Eligible	32143	97.0
Whole household deceased	162	0.5
Other ineligible	838	2.5
Total	33143	100.0

Table 40: Household outcome by sample if eligible, Wave 4

	Number	Percent
GPS		
Complete household	13494	61.5
Partial household	4191	19.1
Household grid	28	0.1
Refusal	2273	10.4
Non-contact	1771	8.1
Other non-response	189	0.9
Total	21946	100.0
EMBS		
Complete household	1511	46.8
Partial household	819	25.4
Household grid	15	0.5
Refusal	394	12.2
Non-contact	455	14.1
Other non-response	35	1.1
Total	3229	100.0
BHPS sample		
Complete household	4628	66.4
Partial household	1171	16.8
Household grid	18	0.3
Refusal	593	8.5
Non-contact	484	6.9
Other non-response	74	1.1
Total	6968	100.0
Total		
Complete household	19633	61.1
Partial household	6181	19.2
Household grid	61	0.2
Refusal	3260	10.1
Non-contact	2710	8.4
Other non-response	298	0.9
Total	32143	100.0

Table 41: GPS: Individual response, Wave 4 by type of sample member (persons over 15 in enumerated households)

	Number	Percent
OSM		
Full interview	27708	70.9
Proxy interview	2334	6.0
Refusal	5709	14.6
Other non-interview	3328	8.5
Total	39079	100.0
PSM		
Full interview	89	53.0
Proxy interview	30	17.9
Refusal	20	11.9
Other non-interview	29	17.3
Total	168	100.0
TSM		
Full interview	1174	49.1
Proxy interview	342	14.3
Refusal	548	22.9
Other non-interview	329	13.7
Total	2393	100.0
Total		
Full interview	28971	69.6
Proxy interview	2706	6.5
Refusal	6277	15.1
Other non-interview	3686	8.9
Total	41640	100.0

Table 42: EMBS: Individual response, Wave 4 by type of sample member (persons over 15 in enumerated households)

	Number	Percent
OSM		
Full interview	3761	57.7
Proxy interview	554	8.5
Refusal	1172	18.0
Other non-interview	1032	15.8
Total	6519	100.0
PSM		
Full interview	12	40.0
Proxy interview	6	20.0
Refusal	5	16.7
Other non-interview	7	23.3
Total	30	100.0
TSM		
Full interview	476	52.9
Proxy interview	110	12.2
Refusal	178	19.8
Other non-interview	136	15.1
Total	900	100.0
Total		
Full interview	4249	57.0
Proxy interview	670	9.0
Refusal	1355	18.2
Other non-interview	1175	15.8
Total	7449	100.0

Table 43: BHPS sample: Individual response, Wave 4 by type of sample member (persons over 15 in enumerated households)

	Number	Percent
OSM		
Full interview	8306	76.8
Proxy interview	348	3.2
Refusal	1338	12.4
Other non-interview	823	7.6
Total	10815	100.0
PSM		
Full interview	642	70.2
Proxy interview	76	8.3
Refusal	125	13.7
Other non-interview	71	7.8
Total	914	100.0
TSM		
Full interview	1049	60.4
Proxy interview	140	8.1
Refusal	346	19.9
Other non-interview	201	11.6
Total	1736	100.0
Total		
Full interview	9997	74.2
Proxy interview	564	4.2
Refusal	1809	13.4
Other non-interview	1095	8.1
Total	13465	100.0

Table 44: GPS: Individual response, Wave 4 by outcome in Wave 3 (OSM over 15 in enumerated households)

	Number	Percent
Adult interview Wave 3		
Full interview	25365	86.6
Proxy interview	596	2.0
Refusal	2094	7.1
Other non-interview	1250	4.3
Total	29305	100.0
Proxy interview Wave 3		
Full interview	517	23.0
Proxy interview	1129	50.3
Refusal	399	17.8
Other non-interview	201	8.9
Total	2246	100.0
Youth interview Wave 3		
Full interview	354	76.1
Proxy interview	34	7.3
Refusal	48	10.3
Other non-interview	29	6.2
Total	465	100.0
Interview non-response Wave 3		
Full interview	1453	20.7
Proxy interview	572	8.2
Refusal	3155	45.0
Other non-interview	1833	26.1
Total	7013	100.0
Total		
Full interview	27703	70.9
Proxy interview	2332	6.0
Refusal	5702	14.6
Other non-interview	3326	8.5
Total	39063	100.0

Table 45: EMBS: Individual response, Wave 4 by outcome in Wave 3 (OSM over 15 in enumerated households)

	Number	Percent
Adult interview Wave 3		
Full interview	3056	77.7
Proxy interview	122	3.1
Refusal	412	10.5
Other non-interview	341	8.7
Total	3931	100.0
Proxy interview Wave 3		
Full interview	147	26.5
Proxy interview	239	43.1
Refusal	86	15.5
Other non-interview	83	15.0
Total	555	100.0
Youth interview Wave 3		
Full interview	76	74.5
Proxy interview	5	4.9
Refusal	11	10.8
Other non-interview	10	9.8
Total	102	100.0
Interview non-response Wave 3		
Full interview	427	23.3
Proxy interview	184	10.0
Refusal	651	35.5
Other non-interview	570	31.1
Total	1832	100.0
Total		
Full interview	3706	57.7
Proxy interview	550	8.6
Refusal	1160	18.1
Other non-interview	1005	15.7
Total	6421	100.0

Table 46: BHPS sample: Individual response, Wave 4 by outcome in Wave 3 (OSM over 15 in enumerated households)

	Number	Percent
Adult interview Wave 3		
Full interview	7781	87.9
Proxy interview	114	1.3
Refusal	598	6.8
Other non-interview	355	4.0
Total	8848	100.0
Proxy interview Wave 3		
Full interview	78	22.2
Proxy interview	148	42.0
Refusal	75	21.3
Other non-interview	51	14.5
Total	352	100.0
Youth interview Wave 3		
Full interview	102	75.6
Proxy interview	10	7.4
Refusal	13	9.6
Other non-interview	10	7.4
Total	135	100.0
Interview non-response Wave 3		
Full interview	339	23.3
Proxy interview	74	5.1
Refusal	642	44.1
Other non-interview	402	27.6
Total	1457	100.0
Total		
Full interview	8300	76.9
Proxy interview	346	3.2
Refusal	1328	12.3
Other non-interview	818	7.6
Total	10792	100.0

Table 47: Adult self-completion outcome, Wave 4 by sample (conditional on Wave 4 individual response)

	Number	Percent
GPS		
Adult self-completion	26502	91.5
No self-completion	2469	8.5
Total	28971	100.0
EMBS		
Adult self-completion	3390	79.8
No self-completion	859	20.2
Total	4249	100.0
BHPS sample		
Adult self-completion	9058	90.6
No self-completion	939	9.4
Total	9997	100.0
Total		
Adult self-completion	38950	90.1
No self-completion	4267	9.9
Total	43217	100.0

Table 48: GPS: Self-completion outcome, Wave 4 by Wave 3 self-completion outcome (conditional on being an OSM)

	Number	Percent
Adult self-completion Wave 3		
Adult self-completion	22000	95.6
No self-completion	1022	4.4
Total	23022	100.0
No self-completion Wave 3		
Adult self-completion	2180	64.7
No self-completion	1190	35.3
Total	3370	100.0
Youth self-completion Wave 3		
Adult self-completion	337	95.2
No self-completion	17	4.8
Total	354	100.0
Total		
Adult self-completion	24517	91.7
No self-completion	2229	8.3
Total	26746	100.0

Table 49: EMBS: Self-completion outcome, Wave 4 by Wave 3 self-completion outcome (conditional on being an OSM)

	Number	Percent
Adult self-completion Wave 3		
Adult self-completion	2189	88.5
No self-completion	284	11.5
Total	2473	100.0
No self-completion Wave 3		
Adult self-completion	497	54.4
No self-completion	416	45.6
Total	913	100.0
Youth self-completion Wave 3		
Adult self-completion	67	88.2
No self-completion	9	11.8
Total	76	100.0
Total		
Adult self-completion	2753	79.5
No self-completion	709	20.5
Total	3462	100.0

Table 50: BHPS sample: Self-completion outcome, Wave 4 by Wave 3 self-completion outcome (conditional on being an OSM)

	Number	Percent
Adult self-completion Wave 3		
Adult self-completion	6692	96.2
No self-completion	261	3.8
Total	6953	100.0
No self-completion Wave 3		
Adult self-completion	499	48.7
No self-completion	526	51.3
Total	1025	100.0
Youth self-completion Wave 3		
Adult self-completion	95	93.1
No self-completion	7	6.9
Total	102	100.0
Total		
Adult self-completion	7286	90.2
No self-completion	794	9.8
Total	8080	100.0

Table 51: Youth interview outcome, Wave 4 by type of sample member (persons aged 10-15 in enumerated households Wave 4)

	Number	Percent
GPS		
Youth interview	2580	76.4
Youth non-response	797	23.6
Total	3377	100.0
EMBS		
Youth interview	559	63.8
Youth non-response	317	36.2
Total	876	100.0
BHPS sample		
Youth interview	909	77.8
Youth non-response	259	22.2
Total	1168	100.0
Total		
Youth interview	4048	74.7
Youth non-response	1373	25.3
Total	5421	100.0

Table 52: Youth interview outcome, Wave 4 by youth outcome in Wave 3 (OSM aged 10-15 in enumerated households Wave 4 and also enumerated Wave 3)

	Number	Percent
Youth interview Wave 3		
Youth interview	2706	85.8
Youth non-response	448	14.2
Total	3154	100.0
Youth non-response Wave 3		
Youth interview	445	49.0
Youth non-response	463	51.0
Total	908	100.0
Child under 10 Wave 3		
Youth interview	543	70.2
Youth non-response	231	29.8
Total	774	100.0
Total		
Youth interview	3694	76.4
Youth non-response	1142	23.6
Total	4836	100.0

Table 53: Percent of item non-response – individual questionnaires

	W1	W2	W3	W4	W5
Overall item non-response rate	2.8	3.9	2.6	2.0	2.1
UKHLS sample only	2.8	3.5	2.3	1.6	2.3
BHPS sample only	n.a.	4.1	2.7	2.2	2.0
Self-completion refused/proxy	5.4	5.0	6.1	7.5	5.6
Provided self-completion	2.1	3.7	1.9	0.9	1.2
Adult self-completion questions only*	3.3[1]	9.2[1]	0.6	0.6	0.4
"Extra 5 minutes" sample excluded	2.7	3.8	2.6	2.1	2.1
"Extra 5 minutes" sample only	3.2	4.8	2.8	1.3	2.2
"Extra 5 minutes" questions only*	1.4	11.7	2.0	1.0	1.3
Nurse Health Assessment questions only*	n.a.	1.4	1.4	n.a.	n.a.
Youth self-completion questions*[1]	6.2	4.9	15.5	7.0	16.8

Notes: All substantial variables included; interview checks, interviewer observations, derived variables and system variables excluded. Variables recording response to stable contact details excluded due to improbable missing values. The analysis is based on special licence data which include non-response to more sensitive questions such as foreign country of birth.
* Sample restricted to units who completed the relevant instrument.
[1] PAPI. Missing values have not been checked with respect to any routing within the instrument. I.e., there may be instances where the missing value is [-9] missing instead of [-8] inapplicable.

Table 54: Percent of item non-response – household questionnaire

	W1	W2	W3	W4	W5
Overall item non-response rate	1.2	0.8*	0.8*	1.3	1.7
UKHLS sample only	1.2	1.4	0.9	1.3	1.7
BHPS sample only	n.a.	0.6	0.8	1.4	1.7

Note: All substantial variables included; interview checks, interviewer observations, derived variables, imputed values, and system variables excluded. * There was an error in the script for collecting information on mortgages and house values for respondents who confirmed their homeownership status fed forward from the previous wave. The released data show are [-9] missing for affected households but have been recoded to [-8] inapplicable for this analysis because the questions were not asked.

Table 55: Percent item non-response for specific variables

Description	Variable name	wave 1	2	3	4	5
Marital status	mlstat	0.0	0.1	0.2	0.4	0.3
General health	sf1	0.2	0.1	0.1	0.1	0.0
SOC code for current job	jbsoc00	0.8	0.8	1.2	0.9	0.6
SIC code for current job	jbsic07	1.8	1.7	1.8	0.6	0.5
Gross pay last payment	paygl	15.1	11.7	11.8	11.0	11.7
Usual pay	payu	8.8	9.1	11.1	9.5	11.2
Day started current job	jbbgd	8.6	30.8	30.6	27.8	28.5
Month started current job	jbbgm	3.5	18.6	20.7	18.7	17.4
Year started current job	jbbgy	0.5	0.7	0.6	0.7	0.4
S/e net profit in last yearly account	jsprf	46.9	41.8	41.0	38.5	39.3
Supports a particular political party	vote1	0.5	0.3	0.5	0.3	0.3
Party would vote for tomorrow	vote3	9.4	10.6	11.5	14.0	14.7
Which political party closest to	vote4	1.4	1.2	1.1	1.1	1.0
Monthly amount saved	saved	n.a.	10.2	n.a.	9.6	n.a.
Value of property: home owners	hsval	8.9	7.5	7.6	6.3	6.0
Net amount of last rent payment	rent	6.0	4.7	4.8	4.3	4.2
Gross rent incl. Housing Benefit	rentg	8.4	8.2	9.2	8.9	7.1

Table 56: Frequencies of sample members with retrospectively corrected gender and age variables by sample group and wave.

	Corrected sex category		Corrected age		Total sample
	UKHLS	BHPS	UKHLS	BHPS	UKHLS+BHPS
Wave	N (%)	N (%)	N (%)	N (%)	N (%)
1	94 (0.12)	-	566 (0.73)	-	77309 (100)
2	87 (0.11)	22 (0.03)	148 (0.19)	16 (0.02)	77538 (100)
3	0 (0)	1 (<.01)	72 (0.10)	8 (0.01)	70751 (100)
4	0 (0)	2 (<.01)	31 (0.05)	8 (0.01)	65773 (100)
Total	181	25	817	32	

Table 57: Coding schema for recording within household relationships

Value	Value label	Value	Value label
1	husband/wife	16	step-brother/sister
2	partner/cohabitee	18	foster brother/sister
3	civil partner	19	brother/sister-in-law
4	natural son/daughter	20	grand-child
5	adopted son/daughter	21	grand-parent
6	foster child	22	cousin
7	stepson/stepdaughter	23	aunt/uncle
8	son-in-law/daughter-in-law	24	niece/nephew
9	natural parent	25	other relative
10	adoptive parent	26	employee
11	foster parent	27	employer
12	step-parent	28	lodger/boarder/tenant
13	parent-in-law	29	landlord/landlady
14	natural brother/sister	30	other non-relative
15	half-brother/sister		

Table 58: Listing of potential coding errors in the relationship grid

	Wave 1		Wave 2		Wave 3		Wave 4		Wave 5	
Relationship pair in household with …	N	%	N	%	N	%	N	%	N	%
respondent with too many natural parents	306	0.2	716	0.4	506	0.3	1,802	1.2	1,750	1.2
child that is older than their parent	230	0.1	292	0.2	178	0.1	706	0.5	930	0.7
grandchild that is older than grandparent	0	0.0	12	0.0	42	0.0	110	0.1	72	0.1
any pair of brothers-in-law who are too young to be married	446	0.2	32	0.0	52	0.0	12	0.0	42	0.0
child that is parent of somebody	352	0.2	436	0.2	204	0.1	1,968	1.3	1,856	1.3
person who is partner to >1 person	92	0.1	132	0.1	0	0.0	90	0.1	206	0.1
child that is partner of somebody	78	0.0	178	0.1	2	0.0	216	0.1	334	0.2
niece/nephew that is older than aunt/uncle	1,190	0.7	884	0.5	1,048	0.6	1,276	0.8	1,094	0.8
any pair of siblings >20 years apart in age	6,670	3.7	7,908	4.4	6,286	3.8	4,744	3.1	4,442	3.1
any of the above potential errors	9,564	5.2	8,078	4.5	7,266	4.4	7,066	4.6	6,586	4.6
Total number of relationship pairs	182,802	100	179,044	100	164,638	100	152,972	100	142,570	100

Source: *Understanding Society* (2015), Waves 1-5, 2009-2014.

Table 59: Change over time in the relationship status “natural child” from Wave 1 to Waves 2-5

Relationship ego to alter in later wave(s)	Original version		Edited version	
	Freq.	Percent	Freq.	Percent
Missing			2	0
husband/wife	19	0.01	12	0.01
partner/cohabitee	7	0.01	1	0
natural son/daughter	138,494	99.2	140,406	99.71
adopted son/daughter	85	0.06	88	0.06
foster child	4	0	5	0
stepson/stepdaughter	139	0.1	132	0.09
son-in-law/daughter-in-law	70	0.05	51	0.04
natural parent	10	0.01	1	0
parent-in-law	6	0	2	0
natural brother/sister	627	0.45	24	0.02
half-brother/sister	12	0.01	1	0
step-brother/sister	1	0		
brother-in-law/sister-in-law	5	0	2	0
grand-child	31	0.02	18	0.01
Cousin	4	0	1	0
aunt/uncle			7	0
niece/nephew	57	0.04	25	0.02
other relative	3	0	3	0
lodger/boarder/tenant	2	0	2	0
other non-relative	35	0.03	37	0.03
Total[1]	139,611	100	140,820	100

[1] Sample restricted to pairs of respondents whose relationship code equalled [4] ‘natural son/daughter’ on the respective relationship variable in Wave 1 and who were observed at least once more in Waves 2-5. All waves pooled.

Table 60: List of select data files

Filename	Description
Data from responding sample members	
*w*_hhresp	Substantive data from responding households plus derived variables
*w*_indresp	Substantive data for responding adults (16+) including proxies and telephone interviews from individual questionnaires including self-completion, plus derived variables, weights and imputed variables.
*w*_youth	Substantive data from responding youths (age 10-15), plus derived variables and weights.
Data from enumerated sample members	
*w*_indall	Household grid data for all persons in household, including children and non-respondents
*w*_child	Childcare, consents and school information of all children in the household
*w*_egoalt	Kin and other relationships between pairs of individuals in the household
Cross-wave files	
xwavedat	Stable characteristics of individuals
xivdata	Interviewer characteristics
xwaveid	Individual and household identifiers across all waves
Paradata	
*w*_hhsamp	Data from Address Record File for issued households
*w*_callrec	Information about interview outcome at each call
*w*_issue	Information about interview outcomes at each issue including interviewer number

Table 61: Description of the Primary Sampling Unit variable (*w*_psu)

Value	Sample	Notes
1 – 575	former BHPS sample in England, Scotland and Wales	Identical to the BHPS variable *w*psu
701 – 1999	former BHPS Northern Ireland sample	Corresponds to initial (BHPS Wave 11) sampled households, as these were selected in a one-stage design
2001 – 4640	UKHLS-GPS in England, Scotland and Wales	Corresponds to the postal sectors used as PSUs, see Berthoud et al. (2009)
46424 – 7035	UKHLS-GPS in Northern Ireland	Corresponds to Wave 1 sampled households, as these were selected in a one-stage design
7048 – 51789	UKHLS-EMBS	Corresponds to Wave 1 sampled households, as these were selected in a one-stage design within the high minority density domain, see Berthoud et al. (2009)

Note: There was an error in b_psu and c_psu for Northern Ireland BHPS households in the Wave 2 and Wave 3 data releases. This has been corrected from the Wave 4 release.

Table 62: Description of the stratification variable (w_strata)

Values	Sample	Notes
1 – 151	former-BHPS sample in England, Scotland and Wales	Identical to the BHPS variable *w*strata
701	former-BHPS Northern Ireland sample	Northern Ireland treated as a single stratum
2001 – 3320	UKHLS-GPS in England, Scotland and Wales	Corresponds to groups of two or more PSUs in selection order, as they were selected systematically from an implicitly ordered list, see Berthoud et al. (2009)
3321	UKHLS-GPS in Northern Ireland	Northern Ireland treated as a single stratum
3322 – 5117	UKHLS-EMB	Corresponds to the postal sectors in the high minority density domain as selections were made independently from each, see Berthoud et al. (2009)

Table 63: Naming conventions for weights

Wave	Target population (units of analysis)	Survey instrument (questionnaire)	Sample[1]	Weight Type
w_	Xxx	yy	zz	_aa
a_	hhd: household	en: enumeration	us: GPS & EMB	_xw: cross-sectional analysis weight
b_	psn: persons 0+	in: interview	bh: BHPS	_lw: longitudinal weight
c_	ind: persons 16+	px: interview or proxy	ub: GPS, EMBS and BHPS	_xd: x-sectional design weight
d_	yth: persons 10-15	5m: “Extra-5-minutes”	91: BHPS original sample	_li: longitudinal inclusion weight
…		sc: self-completion	01: BHPS original sample + boosts	
		ns: nurse visit		
		bd: blood		

Notes: [1] Not further described in this table are weights available for the GPS only. These are marked by “gp” and there is only one type of such weights - the design weights for the GPS. That weight should be used by advanced users only.

Table 64: Components of net income variables on *Understanding Society*

Personal monthly income	
*w*_fimngrs_dv	Gross monthly personal income gross (imputed)
*w*_netinc1	Net monthly personal income (imputed), no taxes deducted other than taxes on earnings
*w*_fimngrs_if	Imputation flag fimngrs_dv and netinc1
Components of (imputed) personal gross monthly income: fimngrs_dv	
*w*_fiyrinvinc_dv	Gross annual income from savings and investments
*w*_fibenothr_dv	Gross monthly income from benefits and other sources
*w*_fimnlabgrs_dv	Gross monthly labour income
Components of (imputed) personal net monthly income: netinc1	
*w*_inc1lab	Net monthly personal labour income (imputed)
*w*_inc2misc	Monthly personal miscellaneous income (imputed)
*w*_inc3prben	Monthly personal private benefit income (imputed)
*w*_inc5inv	Monthly personal investment income (imputed)
*w*_inc6pen	Monthly personal pension income (imputed)
*w*_inc7sben	Monthly personal social benefit income (imputed)
Components of (imputed) personal net monthly labour income: inc1lab	
*w*_inc1alabem	Net monthly personal earnings from main job (imputed); same as *w*_paynu_dv
*w*_inc1blabse	Net monthly personal self-employment income (imputed); *w*_seearnnet_dv
*w*_inc1clabj2	Net monthly personal earnings from second job (imputed); jb2pay_dv (gross monthly earnings from second job, imputed) MINUS taxes and NI contributions
Household monthly income	
w_hhnetinc1	Net monthly household income (imputed), no taxes deducted other than taxes on earnings
w_dep9ctax	Council tax estimated
w_hhnetinc3	Net monthly household income (imputed), minus taxes on earnings, NI contributions and council tax liability

Notes: Personal incomes stored in data file *w*_indresp. Household incomes and estimated council tax stored in data file *w*_hhresp. Variables in italics are available with the Special Licence version of the data, see UKDS SN6931.

Table 65: Sample sizes for selected characteristics of responding households

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Number of people in household[1]					
1	7,569	7,624	6,992	6,557	6,285
2	10,138	10,393	9,361	8,731	8,260
3	5,040	5,070	4,648	4,286	3,929
4	4,498	4,650	4,214	3,949	3,652
5 or more people	2,924	2,771	2,567	2,352	2,243
Total	30,169	30,508	27,782	25,875	24,369
Composition of household (w_hhtype_dv)					
1 male, aged 65+, no children	990	1,079	1,048	1,000	979
1 female, age 60+, no children	2,470	2,764	2,595	2,475	2,432
1 adult under pensionable age, no children	4,109	3,784	3,351	3,083	2,875
1 adult, 1 child	1,117	1,000	839	737	660
1 adult, 2 or more children	1,256	1,047	900	802	695
Couple both under pensionable age, no children	3,885	3,766	3,313	3,001	2,749
Couple 1 or more over pensionable age, no children	3,808	4,283	4,005	3,808	3,721
Couple with 1 child	2,169	2,165	1,937	1,763	1,572
Couple with 2 children	2,745	2,841	2,595	2,420	2,226
Couple with 3 or more children	1,453	1,402	1,275	1,161	1,061
2 adults, not a couple, both under pensionable age, no children	846	807	697	690	654
2 adults, not a couple, one or more over pensionable age, no children	482	537	507	494	476
2 adults, not a couple, 1 or more children	455	465	440	395	356
3 or more adults, no children, incl. at least one couple	2,063	2,334	2,210	2,153	2,082
3 or more adults, 1-2 children, incl. at least one couple	1,347	1,415	1,325	1,185	1,145
3 or more adults, >2 children, incl. at least one couple	21	11	13	10	8
3 or more adults, no children, excl. any couples	570	456	414	381	368
3 or more adults, 1 or more children, excl. any couples	383	352	318	317	310
Total	30,169	30,508	27,782	25,875	24,369
Children aged 0-2 in household[2]					
No	26,975	27,697	25,287	23,665	22,401
Yes	3,166	2,807	2,490	2,210	1,968
Total	30,141	30,504	27,777	25,875	24,369
Children aged 3-4 in household[2]					
No	27,941	28,380	25,842	24,137	22,839
Yes	2,200	2,124	1,935	1,738	1,530
Total	30,141	30,504	27,777	25,875	24,369

Children aged 5-11 in household[2]					
No	24,709	25,213	22,936	21,323	20,187
Yes	5,432	5,291	4,841	4,552	4,182
Total	30,141	30,504	27,777	25,875	24,369
Children aged 12-15 in household[2]					
No	26,523	26,850	24,428	22,804	21,534
Yes	3,618	3,654	3,349	3,071	2,835
Total	30,141	30,504	27,777	25,875	24,369
Housing tenure (w_tenure_dv)					
Owned outright	8,603	9,607	8,915	8,452	8,194
Owned with mortgage	10,367	10,797	9,852	9,064	8,361
Local authority rent	3,761	3,540	3,124	2,794	2,536
Housing Association rented	2,445	2,351	2,131	1,980	1,859
Rented from employer	266	253	224	200	173
Rented private unfurnished	2,544	2,364	2,219	2,032	1,914
Rented private furnished	1,939	1,458	1,172	1,042	968
Other	123	88	53	56	50
Total	30,048	30,458	27,690	25,620	24,055
Crowding[3]					
not crowded	25,250	25,911	23,680	22,116	20,967
crowded	4,919	4,597	4,102	3,759	3,402
Total	30,169	30,508	27,782	25,875	24,369
Households in relative poverty[4]					
not poor	22,868	25,394	23,830	22,520	21,543
poor	7,301	5,114	3,952	3,355	2,826
Total	30,169	30,508	27,782	25,875	24,369
Behind with mortgage/rent payments					
yes	2,880	2,423	2,196	1,979	1,736
no	18,046	17,996	16,196	14,861	13,852
Total	20,926	20,419	18,392	16,840	15,588
Behind with Council Tax payments					
yes	2,585	2,098	1,858	1,639	1,430
no	25,708	25,937	23,623	22,233	21,137
Total	28,293	28,035	25,481	23,872	22,567
Up to date with all bills					
up to date with all bills	27,497	28,545	26,003	24,364	22,949
behind with some bills	2,089	1,666	1,460	1,258	1,109
behind with all bills	425	225	207	161	161
Total	30,011	30,436	27,670	25,783	24,219

[1] Recode of w_hhsize.[2] Recode of w_nch02_dv, w_nch34_dv, w_nch0511_dv, and w_nch1215_dv, as appropriate.[3] Crowding is defined here as households with more than one person per room. Derived from w_hsbeds, w_hsrooms and w_hhsize.[4] Defined here as having access to less than 60% of the median income. The income measure is the household net income (w_hhnetinc1) and this has been equivalised using the old OECD equivalent factor (w_ ieqmoecd_dv).

Table 66: Sample sizes for selected characteristics of responding adults

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Sex (w_sex_cr)					
Male	21,038	22,374	20,207	19,145	18,246
Female	26,694	28,341	25,696	24,072	22,795
Total	47,732	50,715	45,903	43,217	41,041
Age group: 13 categories (w_agegr13_dv)					
16-17 years old	1,626	1,813	1,501	1,413	1,344
18-19 years old	1,359	1,503	1,419	1,372	1,273
20-24 years old	3,493	3,299	2,934	2,795	2,673
25-29 years old	3,983	3,621	3,039	2,726	2,467
30-34 years old	4,190	4,151	3,648	3,328	3,054
35-39 years old	4,563	4,505	3,900	3,527	3,210
40-44 years old	4,727	4,970	4,525	4,146	3,788
45-49 years old	4,368	4,734	4,279	4,095	3,932
50-54 years old	3,835	4,219	4,010	3,824	3,692
55-59 years old	3,401	3,813	3,432	3,395	3,269
60-64 years old	3,578	3,967	3,613	3,244	3,031
65 years or older	8,609	10,110	9,603	9,352	9,308
Total	47,732	50,705	45,903	43,217	41,041
Ethnic group (racel_dv[1])					
British/English/Scottish/Welsh/ Northern Irish	35,920	40,088	36,645	34,537	32,960
Irish	718	1,175	1,046	960	891
Gypsy or Irish traveller	17	12	10	10	12
Any other white background	1,470	1,370	1,161	1,077	995
White and Black Caribbean	345	283	269	266	281
White and Black African	146	116	97	100	97
White and Asian	198	176	162	160	148
Any other mixed background	185	151	140	139	140
Indian	1,911	1,509	1,332	1,262	1,186
Pakistani	1,441	1,177	1,087	1,025	964
Bangladeshi	1,130	827	717	696	643
Chinese	322	229	172	175	159
Any other Asian background	628	509	442	417	372
Caribbean	1,149	859	765	713	636
African	1,459	1,038	883	789	735
Any other Black background	87	74	61	58	59
Arab	282	188	150	138	125
Any other ethnic group	271	208	184	168	157
Total	47,679	49,989	45,323	42,690	40,560

Born in the UK (ukborn[1])					
Yes, England	31,319	31,495	28,885	27,589	26,552
Yes, Scotland	3,412	4,442	3,988	3,678	3,496
Yes, Wales	2,044	3,143	2,914	2,720	2,529
Yes, Northern Ireland	1,929	3,126	2,823	2,576	2,340
Not born in UK	9,016	7,070	6,018	5,542	5,129
Total	47,720	49,276	44,628	42,105	40,046
Migrant generation (generation[1])					
1st generation	9,016	7,070	6,018	5,542	5,129
2nd generation: born UK, at least one parent not born UK	5,109	4,580	4,170	3,906	3,673
3rd generation: born UK, both parents born UK, at least one grandparent not born UK	3,331	3,100	2,747	2,533	2,367
4+ generation: born UK, both parents and all grandparents born UK	27,718	28,447	25,164	22,964	21,405
Other: born UK, no data for any parents or grandparents	56	1,173	1,545	1,980	2,320
Other: born UK, both parents born UK, no data for any grandparents	257	2,398	2,735	3,140	3,232
Total	45,487	46,768	42,379	40,065	38,126
Region (w_gor_dv)					
North East	1,916	1,836	1,651	1,549	1,513
North West	5,091	5,049	4,564	4,241	4,002
Yorkshire and the Humber	3,956	3,843	3,557	3,417	3,295
East Midlands	3,550	3,644	3,385	3,303	3,153
West Midlands	4,344	3,968	3,561	3,378	3,252
East of England	4,174	4,187	3,816	3,637	3,461
London	7,652	6,095	5,297	4,966	4,702
South East	5,774	5,843	5,303	5,030	4,855
South West	3,613	3,853	3,575	3,411	3,279
Wales	2,262	3,895	3,597	3,338	3,094
Scotland	3,403	4,660	4,203	3,854	3,660
Northern Ireland	1,997	3,729	3,377	3,066	2,748
Total	47,732	50,602	45,886	43,190	41,014

De facto marital status (w_mastat_dv)					
Single and never married/in civil partnership	11,000	11,127	9,969	9,466	9,117
Married	24,089	26,374	23,955	22,218	21,033
In a registered same-sex civil partnership	87	102	112	114	126
Separated but legally married	1,132	940	842	800	727
Divorced	3,135	3,272	3,013	2,878	2,761
Widowed	2,893	3,155	2,885	2,689	2,565
Separated from civil partner	11	14	7	9	10
A former civil partner	1	2	3	1	2
A surviving civil partner	4	4	4	5	5
Living as couple	5,363	5,710	5,110	4,929	4,618
Total	47,715	50,700	45,900	43,109	40,964
Current economic activity (w_jbstat)					
Self employed	3,376	3,572	3,286	3,163	3,042
In paid employment	21,743	23,465	21,329	20,113	19,271
Unemployed	3,201	2,889	2,459	2,216	1,863
Retired	9,816	11,433	10,765	10,264	9,971
On maternity leave	365	339	267	240	229
Looking after family or home	3,546	3,265	2,807	2,522	2,228
Full-time student	3,418	3,440	2,983	2,858	2,726
Long term sick or disabled	1,835	1,967	1,679	1,532	1,422
On a government training scheme	77	55	56	38	33
Unpaid worker in family business	40	37	26	25	26
Working in an apprenticeship	0	0	32	42	35
Doing something else	306	250	212	199	192
Total	47,723	50,712	45,901	43,212	41,038
Current job: Eight Class NS-SEC (w_jbnssec8_dv)					
Large employers & higher management	1,207	1,271	1,131	1,093	1,041
Higher professional	2,175	2,239	2,045	2,019	1,982
Lower management & professional	7,203	7,836	7,194	6,872	6,676
Intermediate	3,659	3,949	3,508	3,349	3,127
Small employers & own account	2,513	2,669	2,437	2,384	2,306
Lower supervisory & technical	1,939	2,124	1,914	1,742	1,627
Semi-routine	4,820	5,177	4,667	4,390	4,294
Routine	2,607	2,898	2,597	2,352	2,249
Total	26,123	28,163	25,493	24,201	23,302

Current job: Employment Status 2000 (w_jbes2000)					
Self-employed: large establishment (25+ employees)	55	50	54	45	43
Self-employed: small establishment (1-24 employees)	647	646	586	500	466
Self-employed: no employees	2,721	2,908	2,780	2,744	2,692
Manager: large establishment (25+ employees)	3,582	3,804	3,392	3,209	3,108
Manager: small establishment (1-24 employees)	1,701	1,846	1,647	1,513	1,424
Foreman or supervisor	3,116	3,394	3,082	2,904	2,787
Employee (not elsewhere classified)	14,459	15,686	14,210	13,447	12,873
Total	26,281	28,334	25,751	24,362	23,393
Current job SIC 2007[2]					
Agriculture, forestry and fishing	235	297	254	246	219
Mining and quarrying	69	81	73	77	88
Manufacturing	2,329	2,636	2,378	2,247	2,151
Electricity, gas, steam and air conditioning supply	164	184	156	157	122
Water supply; sewerage, waste management and remediation activities	132	165	148	153	139
Construction	1,524	1,613	1,436	1,344	1,255
Wholesale and retail trade; repair of motor vehicles and motorcycles	3,581	3,857	3,443	3,317	3,146
Transportation and storage	1,354	1,322	1,195	1,144	1,097
Accommodation and food service activities	1,396	1,465	1,272	1,152	1,129
Information and communication	952	953	876	868	842
Financial and insurance activities	973	985	900	855	797
Real estate activities	274	285	280	252	238
Professional, scientific and technical activities	1,446	1,498	1,436	1,416	1,360
Administrative and support service activities	1,244	1,290	1,129	1,121	1,157
Public administration and defence; compulsory social security	1,809	2,088	1,825	1,701	1,661
Education	2,890	3,215	2,963	2,885	2,776
Human health and social work activities	4,155	4,540	4,218	4,050	3,952
Arts, entertainment and recreation	609	667	637	624	573
Other service activities	648	694	629	599	562
Activities of households as employers	54	52	57	53	48
Activities of extraterritorial organisations and bodies	24	18	11	16	15
Total	25,862	27,905	25,316	24,277	23,327

Highest qualification (w_hiqual_dv)					
Degree	10,336	10,609	10,082	9,947	9,784
Other higher degree	5,340	5,606	5,195	4,957	4,755
A-level etc.	8,915	10,142	9,433	9,108	8,766
GCSE etc.	9,826	10,610	9,687	9,012	8,435
Other qualification	4,934	5,110	4,514	4,096	3,829
No qualification	8,348	8,053	6,766	5,947	5,353
Total	47,699	50,130	45,677	43,067	40,922
Fathered any children (w_father)					
Yes		699	584	547	478
No		14,517	13,648	12,978	12,382
Awaiting birth of child		280	269	227	204
Total		15,496	14,501	13,752	13,064
Has had pregnancy (w_preg)					
Pregnant at last interview		619	548	509	427
Yes, has had pregnancy		936	803	675	574
No pregnancies		12,877	11,941	11,109	10,459
Total		14,432	13,292	12,293	11,460
Long-standing illness or disability (w_health)					
Yes	16,882	17,699	16,105	15,123	14,009
No	30,742	32,980	29,756	28,062	27,010
Total	47,624	50,679	45,861	43,185	41,019
Satisfaction with life overall (w_sclfsato)					
Completely dissatisfied	1,033	1,184	1,093	1,083	978
Mostly dissatisfied	1,697	2,111	2,667	2,570	2,432
Somewhat dissatisfied	2,830	3,421	3,240	3,521	3,249
Neither satisfied or dissatisfied	3,908	3,948	3,540	3,858	3,862
Somewhat satisfied	7,049	7,292	7,232	6,678	6,715
Mostly satisfied	17,453	19,687	18,503	17,190	16,188
Completely satisfied	5,588	5,781	4,381	4,008	3,791
Total	39,558	43,424	40,656	38,908	37,215

Notes: Based on the wave-specific indresp data files. Only refers to respondents with a full adult interview. Unbalanced sample.
[1] Information merged in from data file xwavedat.
[2] Recode of w_jbsic07_cc to SIC 2007 categories A-U.

Table 67: Select indicators of individual change

Indicator of change[1]	Wave 2	Wave 3	Wave 4	Wave 5	Pooled[2]	Balanced[3]
Moved to different region (w_gor_dv)	250	288	357	361	1,256	719
Became employed (w_jbstat)	1,881	2,002	1,859	1,864	7,606	3,925
Became self-employed (w_jbstat)	623	671	679	661	2,634	1,597
Became unemployed (w_jbstat)	1,792	1,952	1,843	1,646	7,233	4,254
Started marriage/cohabitation (w_mastat)	588	606	536	531	2,261	1,244
Stopped marriage/cohabitation (w_mastat)	270	260	305	238	1,073	632
Became widowed (w_mastat)	132	180	156	123	591	381
Started living in owner occupied accommodation (w_tenure_dv)	322	388	412	457	1,579	883
Exited poverty (see notes to Table 65)	3,959	3,093	2,402	2,198	11,652	6,406
Entered poverty (see notes to Table 65)	1,960	2,176	2,010	1,729	7,875	4,339
NSSEC 8-category: lower (w_jbnssec8_dv)	924	1,029	1,051	1,040	4,044	2,173
NSSEC 8-category: higher (w_jbnssec8_dv)	786	794	760	731	3,071	1,789
SIC 2007: lower (w_jbsic07_cc)	469	565	534	525	2,093	1,134
SIC 2007: higher (w_jbsic07_cc)	504	544	581	599	2,228	1,199
New highest qualification (w_hiqual_dv)	950	931	899	834	3,614	1,359
Became disabled (w_health)	3,621	3,515	3,279	3,199	13,614	8,331
No longer disabled (w_health)	3,158	3,725	3,401	3,047	13,331	8,225
Less satisfied with life (w_sclfsato)	7,697	9,527	9,634	9,060	35,918	20,876
More satisfied with life (w_sclfsato)	6,748	8,365	8,185	8,454	31,752	18,606
Number of observations[4]	50,715	45,903	43,217	41,041	228,608	111,400

Notes: [1] Change defined as wave-on-wave change. Information from the BHPS Waves 1-18 not fed forward, i.e., actual sample sizes may be higher. Original variable used to measure change in brackets. [2] Sum of all occurrences of wave-on-wave change.[3] Sum of all occurrences of wave-on-wave change for adult respondents with a full interview in all waves, incl. Wave 1. [4] Based on the wave-specific indresp data files. Respondents with a full adult interview only. Balanced sample and pooled sample refer to person-year observations.

Table 68: List of *Understanding Society* data products distributed through the UKDS

Study type	Study no.[1]	Study title[2]
core	6614	*Understanding Society*: Waves 1-5, 2009-2014: Standard End-User Licence
core	6931	*Understanding Society*: Waves 1-5, 2009-2014: Special Licence Access
core	6676	*Understanding Society*: Waves 1-5, 2009-2014: Secure Access
core	7251	*Understanding Society*: Waves 2-3 Nurse Health Assessment, 2010-2012
core	7587	*Understanding Society*: Waves 2-3 Nurse Health Assessment, 2010-2012: Special Licence Access
link	7615	*Understanding Society*: Interviewer Survey, 2014
link	7533	*Understanding Society*: Waves 1-3, 2009-2012: Special Licence Access, Geographical Accessibility
link	7642	*Understanding Society*: Wave 1, 2009-2011: Linked National Pupil Database: Secure Access
link-id	7182	*Understanding Society*: Wave 1, 2009-2010: Special Licence Access, School Codes
link-id	6666	Special Licence Access, Local Authority District
link-id	6668	Special Licence Access, Westminster Parliamentary Constituencies
link-id	6671	Special Licence Access, Local Education Authorities
link-id	6675	Special Licence Access, Travel to Work Areas
link-id	6672	Special Licence Access, Strategic Health Authorities
link-id	6673	Special Licence Access, Primary Care Organisations
link-id	7453	Special Licence Access, Acorn Type 2013
link-id	6669	Special Licence Access, Census Area Statistics Wards
link-id	7454	Special Licence Access, Census 2001 Rural-Urban Indicators
link-id	7630	Special Licence Access, Census 2011 Rural-Urban Indicators
link-id	7245	Special Licence Access, Census 2001 Middle Layer Super Output Areas
link-id	7249	Special Licence Access, Census 2011 Middle Layer Super Output Areas
link-id	6670	Special Licence Access, Census 2001 Lower Layer Super Output Areas
link-id	7248	Special Licence Access, Census 2011 Lower Layer Super Output Areas
link-id	6674	Special Licence Access, Census 2001 Output Area Classification
link-id	7629	Special Licence Access, Census 2011 Output Area Classification

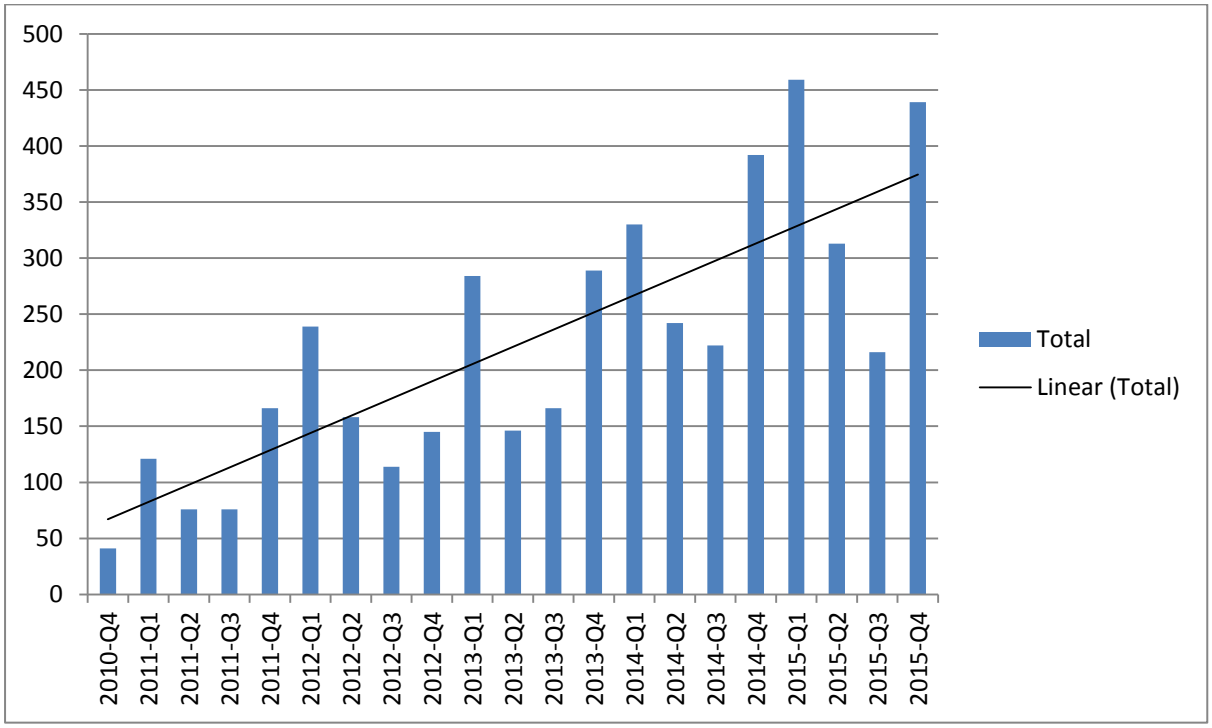
[1] All products can be accessed directly by replacing ## by the Study number in the following URL: https://discover.ukdataservice.ac.uk/catalogue/?sn=##
[2] Unless stated otherwise, the complete study title begins with "*Understanding Society*: Wave 1-5, 2009-2014:"

Table 69: Unique downloads of data from the *Understanding Society* data family 2010-2015

Year	UKHLS Standard Licence (SN 6614)	UKHLS Special Licence (SN6931)	UKHLS Secure Access (SN6676)	LSOA 2001 (SN6670)
2010[1]	41	0	0	0
2011	439	0	0	2
2012	656	23	7	15
2013	885	33	4	36
2014	1186	24	12	18
2015	1427	36	8	20

Notes: Figures as of 25th January 2016.
[1] No special licence or secure access data available in 2010.

Figure 1: Quarterly UKDS downloads for Study 6614



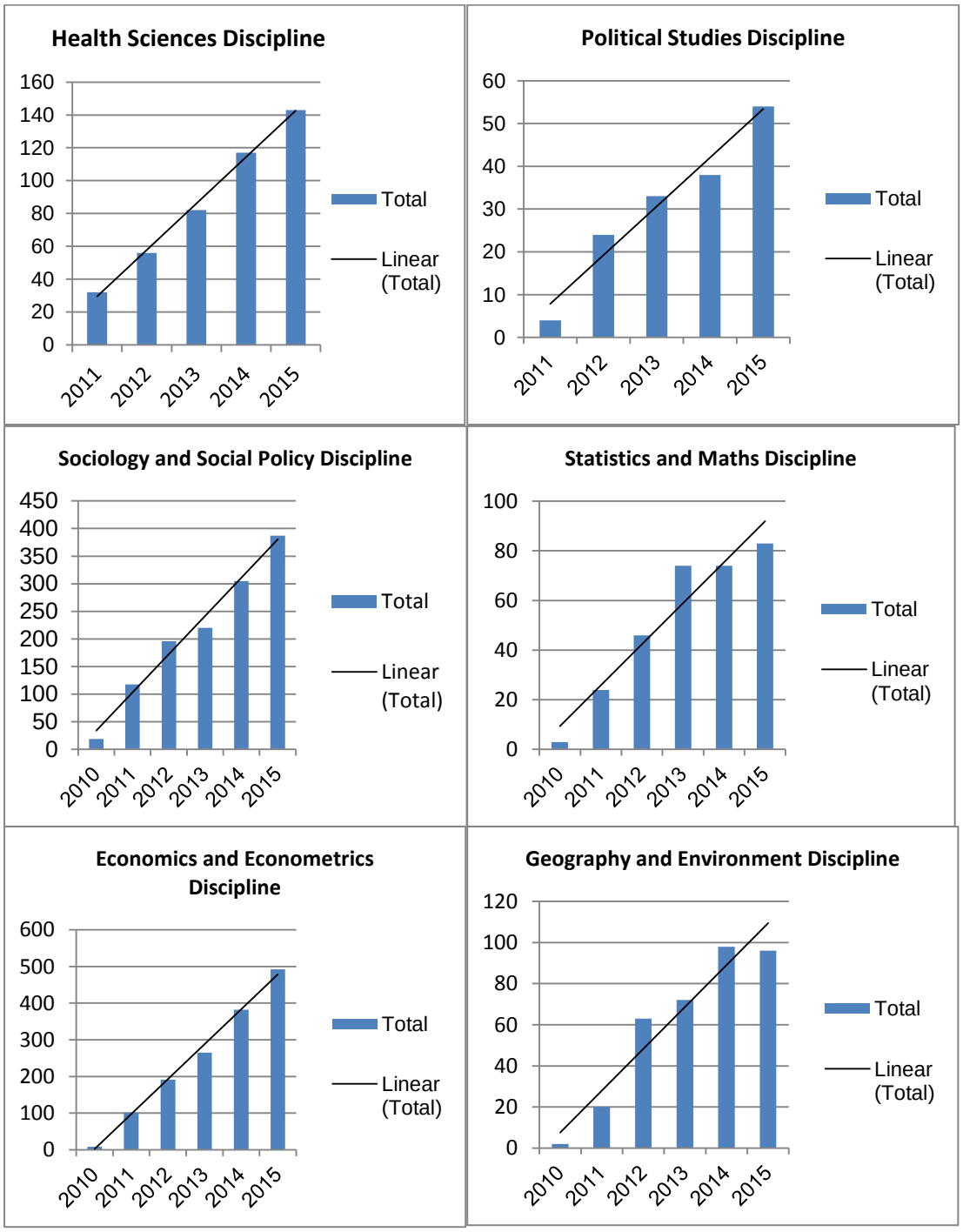
Notes: Figures as of 25th January 2016.Includes datasets 6614, 6849, 7251, 7615, GN 33428, GN 33429 and teaching dataset 7549.

Table 70: Unique users by sector

Sector	Unique Users
UK Higher Education	3867
Higher Education (UK)	3730
Research Centre	137
UK Government	251
Central Government	223
Local Government	28
UK Third Sector	96
Charity	38
Membership Organisation	17
Think Tank	40
Trade Union	1
UK Commercial	75
Commercial	75
Total	4289

Notes: Figures as of 25th January 2016.
Includes datasets 6614, 6849, 7251, 7615, GN 33428, GN 33429 and teaching dataset 7549.

Figure 2: Unique annual users of *Understanding Society*, by discipline.



Notes: Figures as of 25th January 2016.
Includes datasets 6614, 6849, 7251, 7615, GN 33428, GN 33429 and teaching dataset 7549.

Table 71: Publications using Understanding Society, by type

Record Type	Publications
Book/Chapter Publication	46
Book	9
Book Chapter	37
Academic Papers	178
Journal Article	84
Working/Research Papers	94
Parliamentary Papers	3
Report	74
Total	301

Notes: Figures as of 25th January 2016.

Table 72: Number of questions posted on the Online User Support Forum, by topic

Main topic[1]	Open	Closed	Total[2]
Data analysis	4	75	79
Data documentation	1	92	93
Data inconsistency		17	17
Data releases	1	16	17
Ethnicity		11	11
External data linkage		15	15
Questionnaire design		12	12
Special license		14	14
Derived variables		20	20
Health		11	11
Income		27	27
Weights	1	43	44
Other[3]	1	41	42
Total	8	394	402

[1] Users are asked to categorise their post into an existing category but can also generate new categories. When users did not specify the category, the decision is made by the Redmine Administrator or member of staff to whom the question is assigned.
[2] Accessed on 10th November 2015.
[3] Summed statistics for topics with less than 10 posts.

Table 73: Number of questions posted in the Online User Support Forum, by main study and wave

Main study[1]	Open	Closed	Total[2]
BHPS	2	76	78
Wave 1	-	42	42
Wave 2	-	47	47
Wave 3	1	21	22
Wave 4	1	5	6
Wave 5	1	1	2
Wave 6	-	-	-
Wave 7	-	-	-
All waves	3	186	189
Nurse Health Assessment	-	11	11
Total	8	389	397

[1]Users are asked to categorise their post into an existing category. When users did not specify the category, the decision is made by the Redmine Administrator or member of staff to whom the question is assigned.
[2] Accessed on 10th November 2015.

Table 74: *Understanding Society* training courses and number of participants, 2009-2015

Training event[1]	Number of participants
Biomarkers in *Understanding Society*[2]	116
Introduction to Innovation Panel	14
Using BHPS with Stata	125
Using UKHLS with Stata	230
Using UKHLS with Stata for Transport Analysts	15
Comparative research using UKHLS and German SOEP	32
Managing and Analysing UKHLS Data	13
Using Weights in UKHLS	22
Total	579

Notes: Figures as of 25th January 2016.
[1] Only courses organised by ISER included in the count. Unless otherwise stated these are events which include a hands-on session using the data. There have been a number of events in which *Understanding Society* team members were involved and which focused on introducing the Study which are not included in the count.
[2] These were not hands-on training sessions. Comparable figures for such events on the Main Study are not available.

Table 75: Unique attendances at training courses, by discipline

Discipline	Unique Attendances
Not specified	41.8%
Sociology and Social Policy	20.2%
Economics and Econometrics	14.7%
Health Sciences	10.0%
Statistics and Maths	4.0%
Political Studies	2.8%
Business and Finance	2.6%
Geography and Environment	2.4%
Others	1.6%
Grand Total (N=579)	100%

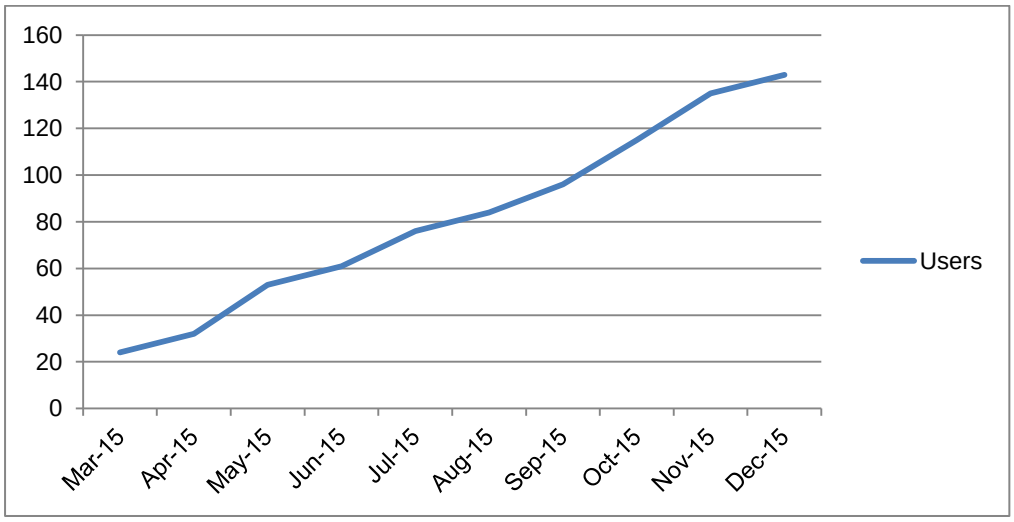
Notes: Figures as of 25th January 2016.

Table 76: Unique registrations for online training courses, by training course

Training course	Live since	Registered users[1]
Introduction to UKHLS using Stata	March 2015	143
Introduction to UKHLS using SPSS	Sept 2015	33
Introduction to BHPS using Stata	April 2015	14
UKHLS for Transport Analysis	July 2015	11
Total		201

Notes: Date accessed 09.12.2015.
[1] Count based on email addresses. No attempt has been made to check for duplicate registrations. Users may sign up for different courses.

Figure 3: New registrations for 'Understanding Society using Stata' Moodle course, by month.



Notes: Figures as of 10.November 2015.

Table 77: Information on ethical reviews of the Study and its components

Main Survey
Ethics Committee of the University of Essex: 6 July 2007 (Waves 1-2) 17 December 2010 (Waves 3-5) 20 August 2013 (Waves 6-8), with further amendments approved 31 July 2014 and 1 July 2015
Linkage to health records
National Research Ethics Service Oxfordshire REC A (08/H0604/124): 21 October 2008 National Research Ethics Service Royal Free Hospital & Medical School (08/H0720/60): 18 June 2008 National Research Ethics Service Southampton REC A (11/SC/0274): 28 September 2011/24 November 2011
Nurse Health Assessment and IBIO pilot
National Research Ethics Service Oxfordshire REC A (10/H0604/2): 9 April 2010. National Research Ethics Service Oxfordshire REC A (10/H0604/62): 19 August 2010. National Research Ethics Service Oxfordshire REC A (10/H0604/70): 20 January 2011