

Understanding Society: design overview

Nick Buck

Stephanie McFall

Institute for Social and Economic Research

University of Essex

nhb@essex.ac.uk

(Received May 2011 Revised October 2011)

Abstract

'Understanding Society', the UK Household Longitudinal Study, builds on the success of the British Household Panel Survey (BHPS). This paper describes some of the key elements of the design and conduct of the study and suggests how Understanding Society is distinctive as a longitudinal survey. Its large sample size offers new opportunities to study sub-groups that may be too small for separate analysis using other studies. The new content included in Understanding Society, not least the bio-measures, provides exciting prospects for interdisciplinary research across the social and medical sciences. The Innovation Panel is already proving to be an invaluable resource for research in longitudinal survey methodology. Finally, the inclusion of the BHPS sample within Understanding Society enables that long-running panel to continue into the future, opening up inter-generational research and the opportunity to look at very long-term trajectories of change. This paper also describes the four sample components: the general population sample, ethnic minority boost sample, the Innovation Panel, and participants from the BHPS. Each component has a multi-stage sample design, mostly with stratification and clustering. A complex weighting strategy is being developed to support varied analyses. This overview also describes the instruments, methods of data collection, and the timetable for data collection. A summary of the survey content is provided. With the data becoming available, the user community is beginning to benefit from this investment in longitudinal studies

Keywords *Understanding Society*; household panel study; design

Introduction

Understanding Society, the UK Household Longitudinal Study, is a major investment in longitudinal studies, the sort of large scale investment in science which has up to now has been more characteristic of the physical and biological sciences. It substantially expands the research resources available to social scientists and researchers in other disciplines, including biomedical and environmental scientists. The study brings potentially huge long-term contributions for the understanding of the UK in the early twenty-first century.

Understanding Society builds on the success of the British Household Panel Survey (BHPS). The BHPS has been heavily used by government departments and by researchers within and outside

the UK. It has been accessed by more than 2000 users and generates more than 150 publications per year. However the ambitions of *Understanding Society* support a much wider range of research than the BHPS. It is an important addition to the UK's rich portfolio of longitudinal studies.

Understanding Society forms part of an international network of studies including the Panel Study of Income Dynamics (Hill 2001), the German Socio-Economic Panel Study (Wagner et al 1993), the Household, Income, and Labour Dynamics in Australia Survey (Wooden et al 2002), the Swiss Household Panel (Budowski et al 2001), the Survey of Labour and Income Dynamics in Canada (Webber 1994) and other active household panels in South Africa, Israel, Korea, China. The household panel design was established in the Panel Study of Income

Dynamics (PSID) in the USA in the late 1960s. This design has proved extremely powerful in understanding the dynamics of populations and the determinants of behaviour and outcomes at household and individual level.

This paper provides an overview of *Understanding Society*, focusing firstly on its design as a household panel study which makes it distinct from other longitudinal studies. This will include description of design features which will contribute to the scope of research questions that can be examined with the study. Secondly, it describes the sample design and thirdly, the plan for data collection. Finally it provides some overview of the data collected, via the survey questionnaire and by other means.

Understanding Society as a household panel study

Longitudinal studies have advanced social science methods, enhanced the understanding of major social changes, and supported better assessment of policy interventions. The UK has taken a prominent role in the development of longitudinal studies, especially with its unique portfolio of birth cohort studies, the Office for National Statistics Longitudinal Study linking census and vital records for England and the comparable Scottish Longitudinal Study and Northern Ireland Longitudinal Study, the English Longitudinal Study of Ageing, the Longitudinal Study of Young People in England, the British Household Panel Survey, and many others. These studies provide an understanding of social change, the trajectories of individual life histories and particularly the processes of individual development across the life course, and the dynamic processes which underlie social and economic life.

Understanding Society is a major addition to this portfolio. As a longitudinal study it follows individuals over time, regularly collecting data about each participant, together with similar data about all other members of his or her household. The design allows it to provide information on the persistence of such states as unemployment, child poverty or disability; on factors which influence key life transitions, such as marriage and divorce, labour force entries and exits and retirement; and, as the study matures, information on the effects of earlier life circumstances on later outcomes. It will support research relevant to the formation and evaluation

of policy and will also foster the development of improved and more reliable analytical techniques. Cross-sectional data, based on only a single observation of each individual, cannot achieve these aims.

Longitudinal studies have a major role in understanding the social and economic changes facing all types of society, as the range of studies cited above suggests, since they collect data about different time points within an individual's life, or indeed look across generations, by collecting and linking data from different points in the lives of parents and children. Longitudinal analysis can provide very different understandings than cross-sectional ones. Specific examples of anticipated uses for the longitudinal data are included in the other papers in this Special Section. Some important issues for which we anticipate contributions from the longitudinal data of *Understanding Society* include:

- The analysis of the incidence of states such as poverty or unemployment over time. The distribution of such states has important temporal dimensions (Hill and Jenkins 2006; Jenkins and Micklewright 2007). The experience of long term unemployment or persistent poverty has different implications for other outcomes such as health, than short term or transitory occurrences (McLeod and Shanahan 1996). Longitudinal studies are uniquely placed to collect the information necessary to analyse these effects.
- The measurement of the rates of transition between states, and the factors associated with them. These might include spells of illness or transitions in partnerships (Aassve, Burgess, Propper and Dickson 2006; Ermisch and Francesconi 2000). The analysis of associations between the life courses of different household members, and of their mutual interactions, is enabled by *Understanding Society* interviews with all household members from age 10. The combined data can be used to examine the dynamics of household formation and dissolution and associated outcomes.

- The analysis of the association between change in the different domains (e.g. health and the labour market), to understand causal ordering.
- The use of repeated survey measures from the same individual allows researchers to separate out person-specific time invariant effects, including those which may be unmeasured, and hence reduce biases caused by associations between these person-specific factors and the change in the phenomenon of interest (e.g. see Hedeker and Gibbons 2006).
- The accumulation of life history data over the waves of the panel, which can be used to analyse the long-term accumulation of personal and financial resources and their implications for other behaviours and outcomes.

The household panel design of *Understanding Society* contrasts with that of the cohort design of many other longitudinal studies. In these cohort studies, a sample of individuals from a particular age group is selected and followed. There are birth cohorts, where the sample is selected around birth and followed thereafter, cohorts of young people, often selected from a particular school year, or ageing cohorts, where people over a threshold age are followed. In the household panel design, a sample of the whole population is selected in their household context. It is important to stress here that the longitudinal elements, just as in the cohort studies, are the individual people. It is not a longitudinal study of households, since arguably households have no coherent existence over time, and focusing analysis only on households whose composition does not change between waves leads to severe biases (see Duncan and Hill 1985). Rather, it is a study of individuals in their changing household contexts and this context is very important for analysis of many life domains (Giles 2001).

There are three key distinctive features of the household panel design compared with the cohort design. Firstly, while a birth cohort study is representative of the population in that particular cohort, the household panel is a representative survey of the whole population of all ages.

Research from studies with a household panel design supports direct inferences about the whole population. Since the study has a full range of age cohorts, and because births to sample members join the sample, there is a representative sequence of new cohorts constantly replenishing the study. Consequently, studies with a household panel design are an important complement to cohort studies by supporting generalisations beyond specific cohorts.

Secondly, it is a survey based on households. Multiple social environments shape behaviours and life circumstances. For example Skew and Wolke, in this Special Section, examine bullying in relation to school and home. However, households can be closely observed in *Understanding Society*. Economic welfare, income and material conditions are normally assessed at the household level, because of the degree of sharing of resources. Households also provide a context for understanding the social and cultural resources available to individuals, both children and adults. The collection of comparable data from each individual in the household at each wave provides a natural way of collecting rich household level data. It also provides a very important resource for the study of how households are changing and the demographic processes which lead to household composition change.

Thirdly and related, the household focus also provides a way of understanding the inter-relationships between individuals within households and families. Many of the key decisions which individuals make are influenced by other household members. A focus on households also provides an opportunity to investigate the inter-relationship with families. Families clearly extend beyond household boundaries and not all household members will consider themselves family members. The study supports research on the interactions over time with family members outside the household, and the evolution of relationships within the household. For example, household panel surveys have contributed to the study of resident and non-resident parents and their contributions to the developmental outcomes for children (Ermisch 2008). They also permit examination of changing patterns in partnerships, such as dissolution and cohabitation, timings of marriages and births, and re-partnering in relation to childbearing and employment outcomes.

Observing multiple generations and all siblings allows examination of long-term transmission processes and isolates the effects of commonly shared family background characteristics. Articles in this issue make use of household relationships in their analyses. Meadows and Arber find a shift with age in the couple level influences on sleep maintenance. Booker and Sacker report that the well-being of younger children is more strongly influenced by a parent with chronic illness, than are older children.

The sustained collection of short-term data makes it possible to accumulate long-term sequences of high quality biographical information across multiple domains. The central purpose of *Understanding Society* is to understand the individual dynamics of change experienced by the population of the UK. With a national sample covering the whole population, *Understanding Society* will also provide representative cross-sectional population estimates for each wave. This Special Section focuses on data from Wave 1 collected in 2009. Nevertheless, its real strength will be the provision of nationally representative longitudinal data at the individual and household level across a range of substantive domains.

Design features which broaden research questions

Five key areas extend the research potential of the household panel design beyond that of the BHPS. The five extensions are:

Size of the survey

The survey's large sample size is a key attribute. The target of 40,000 households across the study's samples will permit exploration of questions for which other longitudinal surveys are too small to support effective research. Many relatively rare events or sub-populations can be studied with the survey. It permits analysis of small sub-groups, people who moved to the UK as children or disabled people, or regional and sub-regional levels. It allows examination of the effects of geographical variation in policy, for example differences between the countries of the UK. The large sample size also allows high-resolution analysis of events in time, for example focussing on single-year age cohorts. As an example, with a total of approximately 1,000 births to sampled women per year, it will be possible to study births to teenage mothers.

An emphasis on ethnicity research

Understanding Society is also noteworthy for its ability to contribute to the understanding of ethnic minorities, which are relatively poorly covered by other longitudinal studies. The UK population can be characterized as having a relatively large number of minority groups, each with small population shares. Study of ethnic variation in general population surveys requires over-representation, using boost samples of minorities in order to achieve sufficient sample size. An ethnic minority boost was also incorporated in the Millennium Cohort Study (Dex and Joshi 2005).

Understanding Society will provide important new information about ethnic minorities through over-sampling of ethnic minorities and the collection of additional measurements relevant to their life experiences. These additional measures are asked of members of the boost sample (over-sample), of members of ethnic minority groups not sampled as part of the boost sample, and by a comparison group of around 1,000 adults from the general population sample (Berthoud et al 2009). Examples of the additional ethnicity relevant measures include ethnic identity, discrimination and harassment, questions about ethnicity and social networks, and questions about remittances or financial payments to families in the country of origin. The ethnicity strand of *Understanding Society* also informs questions for the whole sample. The objective is that the whole study should contribute to the understanding of the UK as a multi-ethnic society. In this Special Section, Nandi and Platt describe the development of ethnic identity measures. It will be interesting to see how the relative importance of identity dimensions change over time. More frequent assessment of ethnic identity in young adults highlights the potential for the shifting importance of its dimensions.

A multi-topic survey

The goal of *Understanding Society* is to support research across the range of social sciences as well as biomedical sciences and other physical sciences, such as environmental sciences. While the research agenda of household panel studies has in the past focused particularly on issues of family dynamics and household organisation, household income and welfare, and labour market participation, *Understanding Society* will also support inter-

disciplinary research on such issues as health and well-being, social participation and a range of other behaviours. *Understanding Society* is extending this agenda with questions tapping psychological attributes and attitudes related to environmental behaviours. Consequently, a long-term content plan has been developed to prioritize measures within the questionnaires and to provide breadth and balance of coverage in topic.

Collection of biomeasures

The fourth feature is a special but extremely important case of expansion in content. It involves the collection of a relatively wide range of biomeasures and other health indicators (Weinstein, Vaupel and Wachter 2008). We use biomarker or biological sample to refer to an assay generated from a biological specimen. We use the broader term, biomeasure, to refer to a range of biological, anthropometric, functional, and sensory measures (Jaszczak, Lundeen and Smith 2009).

The study has been designed to become a biosocial survey, to support biomedical and social science research. Health scientists will find the study a rich source of information about social and economic factors that may influence health status at a single time point, or the trajectory of health outcomes. The particular markers included are relevant to major health outcomes of cardiovascular and metabolic conditions and have been related to social and economic resources and attainments. For example, there is utility to having health measures based on direct physical measurements and samples, in contrast to self-reported health measures, where the reporting may be influenced by the individual's life circumstances. There is also substantial interest in gene-environment linkages as potential explanations of behaviour and disease etiology.

Innovation in measures and methods of data collection

The study involves innovation in both data collection methods and in the type of data collected. The plans include the augmentation of conventional interview data with novel data collection methods and measures, including linkage to spatial and administrative data and the collection of qualitative, visual and audio data. The study also aims to advance methods of survey data collection. Key to this is an Innovation Panel of approximately

1,500 households, which is a testbed for research related to longitudinal survey methods. It is intended both to guide decisions about the study and contribute to the development of longitudinal survey research methods more generally. Methodological questions are addressed in a realistic survey context similar to the general *Understanding Society* design. That is, the same methods of data collection are used in conjunction with a similar set of questionnaires. Uhrig (2011) summarises the experiments included in the first two waves of the Innovation Panel, for example, methodological experiments on incentives, alternative question wording and the use of dependent interviewing on measures of change (Bottazzi, Crossley and O'Dea 2008; Pudney 2010).

Sample design

Understanding Society has four sample components: a) the general population sample, b) the ethnic minority boost sample, c) the Innovation Panel, and d) the sample of participants from the BHPS. The sample designs are similar in having multi-stage sample designs mostly with stratification and clustering. However, the sample design of each sample component has some unique features, (Lynn 2009) which are discussed below.

General population sample component

The general population sample is a stratified, clustered, equal probability sample of residential addresses drawn to a uniform design throughout the whole of the UK (including north of the Caledonian Canal). The Northern Ireland sample is not clustered. Within Great Britain, the Primary Sampling Units (PSUs) are postal sectors stratified by nine regions of England plus Scotland and Wales), population density and minority ethnic density. 2,640 postal sectors were selected systematically, with probability proportional to size (number of addresses). Within each sampled sector, 18 addresses were selected systematically, resulting in an equal-probability sample of a total of 47,520 addresses in Great Britain. In Northern Ireland, 2,400 addresses were selected systematically from the Land and Property Services Agency list of domestic properties, thus making a total of 49,920 selected addresses in the UK. Since constraints of survey capacity meant that fieldwork needed to be spread over a two year period, the overall sample was divided into 24 monthly sub-samples, each independently representative of the UK population.

This means that differences over time within a wave can be compared using nationally representative samples, and annual or quarterly subsets can be independently analysed.

Ethnic minority boost sample component.

The goal for the ethnic minority boost sample was to provide samples of at least 1,000 adults in each of the five largest ethnic minority groups: Indian, Pakistani, Bangladeshi, Caribbean and African. Such a sample would support group-specific analyses of these ethnic groups (Berthoud et al 2009). While the sampling targets are defined in terms of numbers of adults, the sample is of households.

The sampling approach first identifies geographic areas with at least 5% density of ethnic minority groups. Because the 2001 Census was becoming outdated, the density estimates were adjusted using more recent survey estimates. These high density sectors were 36 per cent of the total sectors and accounted for 85% of all members of minorities. Further sub-sampling of the high density areas was done to increase the efficiency of the yield. Thus, a higher sampling fraction was used for areas anticipated to yield three or more households while successively smaller fractions were used for areas expected to yield two, one or zero ethnic minority households.

At selected addresses, households were screened for the presence of a member of a minority ethnic group. The screening question was, "Do you come from or have parents or grandparents who come from any of the following ethnic groups?" The response categories are Indian, mixed Indian, Pakistan, Bangladeshi, Sri Lankan, Caribbean/West Indian, mixed Caribbean/West Indian, North African, Black African, African Asian, Chinese, Far Eastern, Turkish, or Middle Eastern/Iranian, or other. At the screening stage, all households with the smaller ethnic groups were selected and there is some de-selection of larger ethnic minority groups, e.g. Indians.

Following the first six months of data collection the procedures were reviewed and modified. One change was to increase the number of addresses issued in areas estimated to be high in Bangladeshi, the smallest of the five main ethnic groups.

The screening question also identified persons in the following categories in addition to the five target groups: Chinese, other Far Eastern, Sri

Lankan, and Middle Eastern. While it is useful to be able to identify members of these ethnic groups, the number of cases is well below 1,000. White minorities were not selected in the screening but can be identified by survey questions in the general sample.

BHPS sample component.

Understanding Society incorporates the BHPS sample members into the overall sample design beginning in Wave 2. The extensive longitudinal data of the BHPS has great scientific value, including the opportunity for early longitudinal analyses of *Understanding Society*. The BHPS was a random sample of Great Britain, excluding the Scottish Highlands and Islands. In its first wave in 1991, it achieved a sample of 5,500 households. Boost samples of Scotland and Wales were added in 1999 and of Northern Ireland in 2001. These modifications were motivated by interest in analyses in these countries, related to political changes associated with devolution in the UK.

In planning the timing of fieldwork for the BHPS sample, it was necessary to balance fully integrating the sample into *Understanding Society* as against creating a discontinuity in the BHPS series. After consultation, it was decided that it was most important to ensure the integration of BHPS into the new study (Laurie 2010). So instead of having its fieldwork concentrated between September and December, as was the practice up to 2008, fieldwork is distributed evenly over the 12 months of the first year of data collection beginning in January 2010, as part of wave 2 of *Understanding Society*. This does introduce a one-off longer gap between interviews for the BHPS sample. From wave 2 onwards the BHPS sample has the same questionnaire as the *Understanding Society* general population sample. Jenkins and Taylor in this issue present rates of employment from 1991 to 2009 from BHPS and *Understanding Society* data. This allowed examination of the effects of two major recessions on employment patterns of young and older persons. Their discussion of how to maximise the comparability of questions, response categories, and sample composition illustrate important decisions for analyses integrating the two surveys.

The Innovation Panel.

The final sample component is the Innovation Panel (IP). As noted, the purpose of the IP is to test methods of data collection and data collection instruments relevant to the conduct of the main survey. As far as practical, it has design and procedures identical to the other samples. Its sample design began with 2,760 households drawn from 120 areas of Britain. Northern Ireland, and Scotland north of the Caledonian Canal are not included.

Use of the whole sample

The overall achieved initial sample was targeted at 40,000 households: approximately 26,000 from the general population sample, 4,000 from the ethnic minority boost, 1,500 from the Innovation Panel, and 8,400 from the BHPS participants at wave 18 of that study. The total achieved numbers across these four components were 39,802 households containing 101,086 individuals, including children.

Understanding Society has a complex sample design and will be used in various ways by data analysts. Consequently the weighting strategy is also complex. *Understanding Society* must provide weights for the household and individual levels, for units that respond or do not respond to different instruments, e.g. the self-completion instrument, for responding to different combinations of study waves, and for the diverse sample components. Sampling information, including primary sampling unit and strata identifiers, will be available on the data set.

In general, weights are the product of a design weight to convey the probability of selection, adjustment for non-response, and sometimes post-stratification, to make the distribution a closer match to the population distribution.

Units in the major sample components have different probabilities of selection. For example, the members of different ethnic minority groups in the boost sample have different probabilities of selection. In addition, the countries in the former BHPS sample have different sampling fractions, including boost samples for Scotland, Wales and Northern Ireland. Different weights may also be used for analyses which combine the sampling components. For example, when combining the general population component with the former BHPS, the weights adjust for the fact that the BHPS

sample does not contain immigrants since its initial period of recruitment.

The development of weights also takes the time pattern of response into consideration. For example, weights for complete longitudinal responses will be produced. These would take into account differential probabilities of attrition after wave 1. They would include those for Waves 1 and 2 or Waves 1, 2, and 3. Weights for other combinations of waves will be developed to support important analyses based on data from those waves. Cross-sectional weights and weights for single year samples waves will also be produced. This brief summary of the weighting strategy can be supplemented by Lynn and Kaminska (2010).

Following respondents over time and eligibility for interview

The composition of the household, the first stage of sampling, determines the rules for following individual respondents over time. The individuals found at selected households in the first wave are designated as Original Sample Members (OSM). We attempt to retain OSM respondents as part of the sample as long as they live in the UK. Individuals joining the household of an OSM after the sample selection/first interview are temporary sample members (TSM). However, births to an OSM are also classified as OSMs. We attempt to interview TSM participants in successive waves as long as they live in the household of an OSM. In sum, TSMs are not followed for interviews when they leave the household, but OSMs are.

The following rules mimic the demographic processes by which the population is reproduced, including births and deaths, partnership formations and dissolutions, and emigration. They provide a natural sampling method over time, which represents the evolving pattern of households and families in the UK. The one exception is that there is no direct way in which the following rules capture immigrants into the UK. Apart from immigration, the sample remains representative of the UK population as it changes over time, subject to weighting for attrition. Whether and how to sample new immigrants remains an issue to be decided in the future development of the study.

Data Collection

This section describes some important features of the study in relation to data collection, including

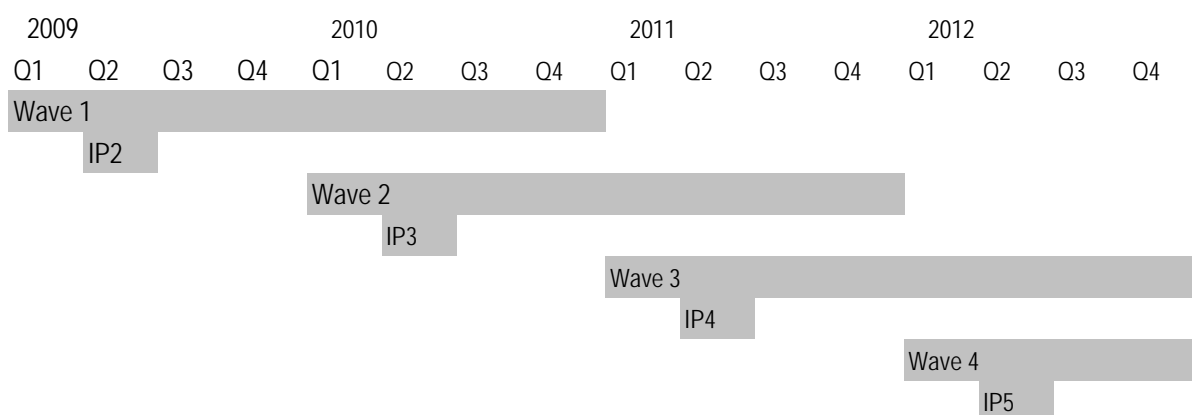
the timing of data collection, modes of data collection, and description of the survey instruments and topics.

Understanding Society is a household panel survey with annual measurements. A sample of households is selected, data are collected from all adult household members, and sample members are followed in subsequent years. Each wave is collected over two years or 24 months, because of the large sample size. Thus, the first wave of data was collected between January 2009 and December 2011. The second wave of data collection began in January 2010 with those interviewed in the first month of the first wave and concluded in December 2012 with those interviewed in the last month of

the first wave. The Innovation Panel is collected in the spring of the year before the main survey wave.

The 12 month interval between interviews with participants, as with the BHPS, is important from an analysis perspective. It captures information about life changes over relatively short periods, before events are forgotten, and provides a high frequency of repeated measures of states that cannot reliably be reported retrospectively (e.g. health and well-being). The timetable for data collection of the first four waves over the period 2009 to 2012, including the innovation panels within this period, is shown in Figure 1. It illustrates the pattern of overlapping waves which is part of the design.

Figure 1: Timetable for data collection waves 1 to 4 by quarter (Q) 2009-2012



Note. IP2, IP3 etc: Innovation panel wave 2, 3 etc.

Most of the data collection uses computer assisted personal interviewing (CAPI). There are several instruments for members in selected households. The structure is similar to the BHPS. One household member completes the household enumeration grid and the household interview, which takes about 15 minutes. Each person aged 16 or older has the individual adult interview (32 minutes) and self-completed questionnaire (8 minutes). Youth aged 10 to 15 are asked to respond to a self-completion questionnaire, which is a paper and pencil instrument. Information about younger children is provided by the responsible parent in the household and adult interviews. There is a brief

proxy interview about adults unable to be interviewed.

The initial four waves of data collection are face to face, a mode of administration that is typically more costly but more likely to reduce attrition when we are establishing the study. An experiment in the second wave of the Innovation Panel compared groups issued to face-to-face interviewing, vs those initially issued to telephone administration with varied procedures for interviewing outstanding household members face-to-face (Lynn, Uhrig and Burton 2010). This experiment has provided information about the reduction in response rates and costs of different

interviewing options. Implementation of multi-mode data collection is planned for Wave 5.

The questionnaire instruments and survey materials have been translated into nine languages: Bengali, Punjabi in Urdu and Gurmukhi scripts, Welsh, Arabic, Somali, Cantonese, Urdu, and Gujarati. Software development of the Language Management Utility (LMU) (Harkness 2003) supports the work of multiple translators and is important for the CAPI scripting of non-Roman scripts and languages not read left to right. In the translated interviews, interviewers can toggle between English and the alternative questionnaire.

Response Rates

The overall response rate at wave 1 for the general population sample at the household level response rate was 57.2%, which is somewhat below the target rate of 60%. The response rate for the ethnic minority boost sample after screening for eligibility was 56.9%, somewhat above the 55% target rate. These rates are typical for multi-purpose surveys of this sort in the UK. Surveys with a more specific focus of particular relevance for the sampled individuals, e.g. interviewing mothers about their children, tend to get rather higher response rates. The target for household response rate in Wave 2 is greater than 80% and for Wave 3 is greater than 90%. Burton, Laurie and Lynn (2011) provide more information about wave 1 response rates. They also present some results from a comparison of the characteristics of the *Understanding Society* wave 1 year 1 sample with the 2009 Labour Force Survey. They find that the two surveys have very similar sample distributions on the characteristics compared. There is no difference between the surveys on housing tenure status but the *Understanding Society* sample appears to have a higher proportion of female participants than the LFS, a higher proportion of children younger than 16 years and a lower proportion of those aged 65 or older. The *Understanding Society* sample has a higher proportion of divorced individuals and a lower

proportion of those who are widowed. The *Understanding Society* sample contains a lower proportion of participants who are in paid employment and higher proportion of those who are not in paid employment or looking for work. Although these differences are statistically significant, the sample sizes are large and so the actual percentage differences between the two samples on these key characteristics are quite small.

Data collected

Questionnaire data

Following extensive consultation with users it was clear there was demand for a wide range of content to be included in the study. To meet this demand within the available questionnaire time, *Understanding Society* has adopted a model in which questions are organized in topical modules which appear annually or are rotated less frequently. Rotating modules vary in frequency, depending on the subject matter and expected rates of change. The annual core is approximately 50% of the interview length and includes, after wave 1, an annual event module concerning key events over the previous year.

Because some of the annual measures are BHPS questions, analysts can examine a longer time series for that sample component. Incorporation of many measures from the BHPS will allow longitudinal analysis in key areas to continue for the BHPS sample. In addition, the inclusion of some BHPS measures provides continuity with the new *Understanding Society* sample.

A key design task in the initial stages was to agree which measures were critical for the annual panel design, and which questions should be carried annually. Table 1 sets out the content of the annual repeated measures and the rotating modules being carried over the first four waves. The table also includes information about modules which are included as part of the additional coverage for the ethnic minority boost sample.

Table 1. Outline of questionnaire content

Annual repeated measures	Rotating modules (general population sample)	Ethnic minority boost: additional rotating modules
Basic demographic characteristics for all household members	Family background, place of birth, education, relationship and fertility information (at first interview only)	Language and functional English literacy
Housing characteristics	Leisure and cultural participation	Migration history
Housing expenditure	Transport and communication access, travel behaviour	Remittances
Household facilities, car ownership	Well-being, sleep quality	Employment discrimination
Consumption expenditure	Psychological/personality traits (e.g. "Big 5", trust, risk perceptions)	Harassment
Health status (e.g. SF12), disability	Health-related behaviour, diet, exercise	Dimensions of identity
Mental health (GHQ12)	Family and social networks	Additional items on political engagement
Education qualifications, aspirations and expectations	Family relationships	Additional items on family and social networks
Labour market activity and employment status, job search	Local neighbourhood	Financial literacy and financial inclusion
Current job characteristics, basic employment conditions, hours of paid work, second jobs	Social support	Religious practice
Childcare, other caring within and outside household	Environmental attitudes and behaviour	Civic capital/use of services
Income and earnings	Political engagement	
Life satisfaction	Employment conditions and job quality	
Political affiliation	Uses of time (e.g. domestic work, volunteering)	
Changes between waves - employment, fertility, partnering, geographic mobility, education and training, diagnosis of health conditions	Wealth, assets and debts	

The broad outline of content in the self-completion questionnaire asked of 10-15 year olds is set out in table 2. Once again some of this content is repeated each wave and some is

repeated at intervals, usually of two years. Some of this content is also carried forward to young adults aged 16-19 who answer the main individual questionnaire.

Table 2: Outline of content of youth self-completion

Relationships with family and friends
 Social networks
 Illicit/risky behaviour
 Experience of education and aspirations
 Bullying at school and between siblings
 Use of leisure time
 Health, diet and obesity, exercise
 Self-esteem and satisfaction with life
 Strength and Difficulties Questionnaire (SDQ)
 Future aspirations for job, family, independence
 Social and political attitudes and values
 Financial behaviour and paid work
 Caring responsibilities
 Ethnic and religious identity

Links to Other Types of Data

The ability to link *Understanding Society* survey data with other data sources is a central goal for the study. The added data will greatly enhance its scientific research capacity.

Administrative data.

Administrative data can be used to supplement the interview data and reduce respondent burden, by adding data that would otherwise be collected in the interview and be a potential source of validation for the survey data (Lane 2010). Respondents were asked in Wave 1 for consents to link health and educational administrative records for themselves and for their children. The health records are held by the National Health Service (NHS), the NHS Central Registers, the health departments for England, Wales, Scotland and Northern Ireland, the General Registration Office and the Office for National Statistics. Consent to link to education records was requested of parents of children aged 4-15 and by young adults aged 16-24 who were currently attending school or had attended school in the UK in the past. At wave 1 approximately 70% of respondents gave consent to linkage. The majority of linked data will be accessed by researchers in a secure environment since it does increase disclosure risk.

We will illustrate the linkage process in terms of health data linkages in England. For those who consented, personal information from their survey

data will be sent to the NHS to establish a flag in the Central Register. The Central Register records the registration of the individual with a general practitioner and is updated following births, moves, name changes, and major events like marriages and deaths. With the flag established, the study will be notified when the Central Register is updated for a study participant. In addition, we will link to medical records like the Hospital Episode Statistics (HES). The type of data in HES includes dates of the episode, information about the facility, procedures and treatments, diagnoses, and waiting times. The data systems in Scotland, Wales and Northern Ireland vary somewhat but similar procedures will be followed for the linkages to the extent possible.

Spatial data.

This study will have the ability to link survey data to geographic designations of relevance for a variety of research interests, including parliamentary constituencies, local educational authorities, travel to work areas for local labour market analysis, local authority districts, and primary care trusts. The rural-urban classification categorizes localities by population size and the population density of surrounding areas. Spatial data will be released with additional safeguards to protect participants from disclosure of their identities. For a useful description of these geographical classifications see: *A Beginners' Guide to UK Geographies* (Office for National Statistics 2010).

Biomeasures.

Biomeasures have been included in several major longitudinal surveys including the 1946 and 1958 British Cohort studies (Ferri, Bynner and Wadsworth 2003), and the English Longitudinal Study of Ageing (ELSA) (Marmot et al 2003). The addition of biomeasures to *Understanding Society* is useful for the examination of objective biological, anthropometric and functional measures within a large sample that spans many ages, and which can be studied within a household context. The measurement of biomarkers in BHPS participants will permit researchers to immediately examine questions which rely on longitudinal psycho-social data.

Collection of the biomeasures began with a sample of adults from the general population sample of Wave 2. Data collection was conducted in a separate visit by trained nurses. The measures include anthropometric information (height, weight, waist circumference and body fat from bioelectrical impedance), blood pressure, lung function (spirometry), grip strength, and the collection of whole blood through venipuncture. The blood can be used for a range of analytes including total cholesterol, high density lipids, c-reactive protein, cystatin-c, and glycated haemoglobin (HbA1C). Respondents have been asked to provide written consent to store blood for future research and for genetic studies. Access to these samples will be regulated by a Data Access Committee established by the Economic and Social Research Council (ESRC).

Qualitative studies.

In the longer term, *Understanding Society* will be significantly enriched by the collection of a wider

range of data, which will help to address research issues that questionnaire data alone cannot address. Examples include using the survey data as a sampling frame for longitudinal qualitative research, the use of diaries to collect more accurate measurements such as time use information or specific health behaviours, and experiments to test specific hypotheses. An example of such a structured experiment is an experiment on trust (Ermisch et al 2009).

Conclusion

Understanding Society is designed to provide high quality longitudinal data to answer research and policy needs over the coming decades. Every effort is being made to conduct the study to the highest standards of best practice in the methodology of conducting longitudinal surveys. This paper has described some of the key elements of the design and conduct of the study and suggested how *Understanding Society* is distinctive as a longitudinal survey. With the data becoming available, the user community is beginning to benefit from this investment. The large sample size offers new opportunities to study sub-groups that may be too small for separate analysis in other studies. The new content included in *Understanding Society*, not least the biomeasures, provides exciting prospects for inter-disciplinary research across the social and medical sciences. The Innovation Panel is already proving to be an invaluable resource for research in longitudinal survey methodology. Finally, the inclusion of the BHPS sample within *Understanding Society* enables this long running panel to continue into the future, opening up inter-generational research and the opportunity to look at very long-term trajectories of change.

References

- Aassve A, Burgess S, Propper C, and Dickson M. (2006) [Employment, family union and childbearing decisions in Great Britain](#). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169, 781-804.
- Bartley M, Sacker A and Clarke P. (2004) [Employment status, employment conditions, and limiting illness: Prospective evidence from the British Household Panel Study 1991-2001](#). *Journal of Epidemiology and Community Health*, 58, 501-506.
- Berthoud R, Fumagalli L, Lynn P, and Platt L. (2009) *Design of the Understanding Society ethnic minority boost sample*. [Understanding Society Working Paper No. 2009-02](#), ISER, University of Essex, Colchester. <http://research.understandingsociety.org.uk/publications/workingpaper/2009-02>
- Bottazzi R, Crossley TF and O'Dea C. (May 2008) *UKHLS Innovation Panel consumption questions: Preliminary analysis* Institute for Fiscal Studies, . London.
- Budowski, Monica, Robin Tillmann, Erwin Zimmermann, Boris Wernli, Annette Scherpenzeel & Alexis Gabadinho. (2001) The Swiss Household Panel 1999-2003: Data for research on micro-social change, *ZUMA Nachrichten* 49, Jg. 25, November 2001.

- Burton J, Laurie H and Lynn P. (2011) Appendix: *Understanding Society* Design Overview. In S. L. McFall and C. Garrington. eds. [Understanding Society: Early findings from the first wave of the UK's household.](#)
- Dex S and Joshi H. (2005) Children of the 21st century: From birth to nine months. Policy Press, Bristol.
- Duncan G and Hill M. (1985) Conceptions of Longitudinal Households: Fertile or Futile?, *Journal of Economic and Social Measurement*, 13, 361-375.
- Ermisch J. (2008) [Child support and non-resident fathers' contact with their children.](#) *Journal of Population Economics*, 21, 827-853.
- Ermisch J and Francesconi M. (2000) [The increasing complexity of family relationships: Lifetime experience of lone motherhood and stepfamilies in Great Britain.](#) *European Journal of Population-Revue Europeenne De Demographie*, 16(3), 235-249.
- Ermisch J, Gambetta D, Laurie H, Siedler T and Uhrig, SCN. (2009) [Measuring people's trust.](#) *Journal of the Royal Statistical Society, series A*, 172, 749-769.
- Ferri E, Bynner J and Wadsworth M. eds. (2003) *Changing Britain, Changing Lives. Three Generations at the Turn of the Century.* Institute of Education, London.
- Giles P. (2001) An overview of the Survey of Labour and Income Dynamics (SLID). *Canadian Studies in Population*, 28, 363-375.
- Harkness J. (2003) Questionnaire translation. In JA Harkness, FJR van de Vijver and PP Mohler. eds. *Cross-Cultural Survey Methods* (pp. 35-56). John Wiley & Sons Inc, Hoboken, New Jersey.
- Hedeker D and Gibbons RD. (2006) Longitudinal data analysis. John Wiley & Sons, Hoboken, N.J.
- Hill MS. (1992) *The Panel Study of Income Dynamics: a User's Guide.* Sage Publications, Newbury Park CA.
- Hill M and Jenkins S. (2006) Poverty amongst British children: chronic or transitory. In B Bradbury, S Jenkins and J Micklewright. eds. *The dynamics of child poverty in industrialized countries* (pp. 92-132). Cambridge University Press, Cambridge, UK.
- Jaszczak A, Lundeen K and Smith S. (2009) Using nonmedically trained interviewers to collect biomeasures in a national in-home survey. *Field Methods*, 21, 26-48.
- Jenkins S and Micklewright J. eds. (2007) *Inequality and poverty re-examined.* Oxford University Press, Oxford.
- Lane J. (2010) Linking administrative and survey data. In PV Marsden and JD Wright. eds. *Handbook of Survey Research* 2nd ed. Pp. 659-680. Emerald Group Publishing, Bingley.
- Laurie H. (2010) *Continuity and innovation in the design of Understanding Society: the UK Household Longitudinal Study.* [Understanding Society working paper, No. 2010-02](#), Colchester. <http://research.understandingsociety.org.uk/publications/working-paper/2010-02>
- Lynn P. (2009) *Sample design for Understanding Society.* [Understanding Society Working Paper 2009-01, No. 2009-01](#), Colchester: ISER, University of Essex. <http://research.understandingsociety.org.uk/publications/working-paper/2009-01>
- Lynn P and Kaminska O. (2010) [Weighting strategy for Understanding Society](#) (No. 2010-05). ISER, University of Essex, Colchester. <http://research.understandingsociety.org.uk/publications/working-paper/2010-05>
- Lynn P, Uhrig SCN and Burton J. (2010) [Lessons from a randomised experiment with mixed-mode designs for a household panel survey](#) (No. 2010-03), Understanding Society working papers, Colchester: University of Essex
- Marmot M, Banks J, Blundell R, Lessof C and Nazroo J. eds. (2003) *Health, Wealth, and Lifestyles of the Older Population in England: the 2002 English Longitudinal Study of Ageing.* Institute for Fiscal Studies, London.
- McLeod J D and Shanahan MJ. (1996) [Trajectories of poverty and children's mental health.](#) *Journal of Health and Social Behavior*, 37, 207-220.
- Office for National Statistics. (2010) A Beginners' Guide to UK Geographies. Retrieved June 17, 2010, from http://www.statistics.gov.uk/geography/beginners_guide.asp
- Pudney S. (2010) [An experimental analysis of the impact of survey design on measures and models of subjective wellbeing](#) (No. 2010-01) Understanding Society working paper, University of Essex, Colchester.
- Uhrig SCN. (2011) Using experiments to guide decision making in *Understanding Society*: Introducing the Innovation Panel. In SL McFall and C Garrington. eds. *Understanding Society: Early findings from the first wave of the UK's household longitudinal study* (pp. 117-124). Institute for Social and Economic Research, University of Essex, Colchester.
- Wagner G, Burkhauser R, and Bheringer F. (1993) The English language public use file of the German Socio-Economic Panel. *Journal of Human Resources* 28, 429-433.
- Webber M. (1994) Survey of labour and income dynamics: an overview. In *Report1994*, Cat. 75-201E. Statistics Canada, Ottawa.
- Weinstein M, Vaupel JW and Wachter KW. eds. (2008) *Biosocial Surveys.* National Academies Press, Washington, DC.
- Wooden M, Freidin S and Watson N. (2002) The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1. *Australian Economic Review*, [Volume 35, Issue 3](#), 339-348.