

**Understanding Society**

**Working Paper Series**

**No. 2019 – 15**

**November 2019**

**Improving population and sub-group coverage: who is missing and what can be done about it?**

**Magda Borkowska (University of Essex)**



## Non-Technical Summary

Sufficient coverage of all key subgroups of the population that the survey aims to represent is one of the important aspects of the representativeness of the sample. There are two key aspects:

- The extent to which the sample mirrors the UK population in terms of some key characteristics.
- The extent to which the sample of particular subgroups is large enough to enable meaningful analysis.

This report examines both of these issues and compares the *Understanding Society* estimates with key benchmark administrative and survey data where possible.

- At Wave 8 (2016/2017), compared to ONS' 2017 Annual Population Survey, *Understanding Society* continues to be cross-sectionally representative in terms of sex, age groups and geographical regions. There is a small under-representation of younger ages of people, those living in London, some ethnic minority groups and a small over-representation of older age groups, people living in Yorkshire and the Humber and those who do not classify themselves into any of the standard categories of ethnicity ('other' ethnic group).
- Using ONS definition of hard to reach groups, the Wave 1 sample of *Understanding Society* had similar prevalence of such groups as the 2011 Census, although they were more likely to have dropped out of the sample by Wave 8. A revised weighting strategy may improve their estimation in subsequent waves in the future.

One useful aspect of panel studies - such as *Understanding Society* - is that individuals flow into the sample over time, which means that members of groups that are not included in the initial sampling frame (such as residents of institutions) might join the study at some point. If such transient groups flow in and out of the study with some frequency, (e.g. being in care) then even if the group is small at any point in time, a panel study will observe an increasing number of people who experience being in this group as the study matures.

- We were able to find participants that represented 'missing populations' - international migrants, children in care, people in residential homes, prisoners, and people who are homeless. In general, the estimates of the annual flow of people into such groups are lower than might be expected based on administrative records although the Study does accumulate significant numbers of participants from such groups. For example, across the first 8 waves, *Understanding Society* includes 16,413 unique observations of non-UK born individuals, 312 people who were evicted from rented accommodation and 219 children who were living in foster care.
- We have identified a number of improvements that are being implemented to the leavers and joiners questions, and we are adding a follow up surveys with people moving into institutions and with those moving abroad.

## Improving population and sub-group coverage: who is missing and what can be done about it?

Dr Magda Borkowska (University of Essex)

**Abstract** Sufficient coverage of all key subgroups of the population that the survey aims to represent is one of the important aspects of the representativeness of the sample. There are, however, different understandings of coverage (Kruskal and Mosteller 1979). In this report we look at the *Understanding Society* sample coverage in relation to two aspects: first, the extent to which the sample mirrors the UK population in terms of some key characteristics; and second, the extent to which the sample of particular subgroups is large enough to enable meaningful analysis. In particular, we discuss the coverage of hard-to-reach groups, known to be commonly under-represented in surveys.

**Keywords:** population studies, representativeness, hard to reach groups, missing populations

**Acknowledgement:** I would like to thank: Michaela Benzeval, Chris Bollinger, Jon Burton, Tom Crossley, Paul Fisher, Laura Fumagalli, Karon Gush, Nicole James, Peter Lynn, and Alina Pelikh who provided insight and expertise that greatly assisted this paper. I am grateful for research assistance provided by Yuliya Kazakova who calculated many of the administrative data estimates used here.

Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service. This methodological briefing has been funded by the Economic and Social Research Council UK Population Lab Innovation Development Grant (ES/S016651/1).

University of Essex. Institute for Social and Economic Research, NatCen Social Research, Kantar Public. (2018). *Understanding Society: Waves 1-8, 2009-2017 and Harmonised BHPS: Waves 1-18, 1991-2009*. [data collection]. 11th Edition. UK Data Service. SN: 6614, <http://doi.org/10.5255/UKDA-SN-6614-12>

**Corresponding author:** Magda Borkowska, University of Essex, [m.borkowska@essex.ac.uk](mailto:m.borkowska@essex.ac.uk)

## Introduction and aims

Sufficient coverage of all key subgroups of the population that the survey aims to represent is one of the important aspects of the representativeness of the sample. There are, however, different understandings of coverage (Kruskal and Mosteller 1979). In this report we look at the *Understanding Society* sample coverage in relation to two aspects: first, the extent to which the sample mirrors the UK population in terms of some key characteristics; and second, the extent to which the sample of particular subgroups is large enough to enable meaningful analysis. In particular, we discuss the coverage of hard-to-reach groups, known to be commonly under-represented in surveys.

The paper is divided into two sections. Section 1 focuses on hard-to-reach groups, which are included in the initial sampling frame of the Study and are relatively large in size. The first part of Section 1 investigates how closely *Understanding Society* sample mirrors population estimates (calculated based on administrative data or where appropriate, based on benchmark survey data) in the initial and most recent wave (Wave 1 and Wave 8) in terms of key sociodemographic characteristics such as: sex, age group, region of residence, ethnicity, economic activity, country of birth, dwelling type, household tenure and household composition. The presented estimates focus on the subgroups traditionally considered as hard-to-reach, such as: the youngest and oldest age groups, ethnic minority groups, immigrants, people living in multiple-occupied dwellings, those in private-rented accommodation, those with pre-school children and single-person households. These benchmark comparisons aim to assess how well *Understanding Society* sample covers some of the main societal groups across waves and whether the sample remains cross-sectionally representative over time.

The second part of Section 1 assesses the effectiveness of field operations across different areas through comparing household response rate by key area-level characteristics, known to be associated with propensity to respond to surveys. We conclude this section by providing recommendations on what type of improvements could potentially be made to increase coverage of the discussed hard-to-reach groups.

Section 2 investigates the extent to which *Understanding Society* sample covers small and particularly vulnerable subgroups, sometimes considered as ‘missing populations’ due to being largely outside of the scope of household surveys. These groups include: international migrants, children in care, people living in care homes, people who move into institutional settings, prisoners and homeless.

There is a growing recognition in the data science community that the absence of those small groups can substantively affect key policy-relevant estimates (such as poverty rates, health and educational outcomes) since those who are missing from the data are likely to be the most vulnerable (ODI 2015). In relation to the small, hard-to-reach groups, we focus on two issues: first; the extent to which such groups ‘flow’ in and out of the Study (for example, we look at the number and proportion of different demographic subgroups who move into institutional settings) and second; what direct and indirect information is already collected by the Study that would allow us to identify those who are likely to be in (or fall into) one of the categories of vulnerable subgroups. We conclude Section 2 by providing suggestions on what additional information could be collected to better capture small, mobile and/or particularly transient groups.

## Methods and data

In this section we describe the analytical strategies and types of data used in different sections of the paper. Some of the further details and caveats are also discussed in the main body of the paper, under relevant sections.

The analysis presented in the first part of the paper provide a comparison between *Understanding Society* sample in terms of key variables of interest and relevant administrative or, where appropriate, other gold standard survey data used to produce official government statistics. For *Understanding Society* estimates, we additionally report 95% confidence intervals and the number of observations for each of the subgroups. Where possible, we also report confidence intervals for the benchmark estimates that we compare *Understanding Society* with. In most cases, we use the whole *Understanding Society* UK sample, however, where appropriate, we restrict the sample to smaller geographies (e.g. England and Wales, only England or only Scotland depending on what administrative data exists). For making comparisons with external data sources, we utilise all *Understanding Society* sample components, that is: General Population Sample (GPS), BHPS sample, Ethnic Minority Boost Sample (EMB) and Immigrant and Ethnic Minority Boost (IEMB); and include all members of the sample, that is: Original Sample Members (OSM), Temporary Sample Members (TSM) and Permanent Sample Members (PSM). All estimated proportions (except household response rates) are weighted and the details of the weights used are provided in the notes under the tables.

In the first part of Section 1, we assess whether *Understanding Society* continues to be cross-sectionally representative over time by looking at the comparison between Wave 8 (2016/2017) of the total *Understanding Society* sample and: i) 2017 ONS mid-year population estimates<sup>1</sup> for gender, age group, and region variables (Table 1); ii) 2017 Annual Population Survey<sup>2</sup> for ethnicity variable (Table 2) and iii) 2017 Labour Force Survey<sup>3</sup> for economic activity variable (Table 3).

---

<sup>1</sup> The mid-year population estimates are the official set of population estimates for the UK and its constituent countries, the regions of England and Wales and local authorities. The estimates account for long-term international migrants (people who change their country of usual residence for a period of 12 months or more), but do not account for short-term migrants (people who come to or leave the country for a period of less than 12 months). A combination of registration, survey and administrative data are used to estimate the different components of population change. The mid-year population estimates are used directly as a base for other secondary population statistics, such as population projections, population estimates for the very old and population estimates for small geographical areas. They are also used in the weighting of survey estimates. Description of the methodology of mid-year population estimates can be found here: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/methodologyguideformid2015ukpopulationestimatesenglandandwalesjune2016>

<sup>2</sup> The Annual Population Survey (APS) is a continuous, cross-sectional household survey, covering the UK. The APS is not a stand-alone survey; it uses data combined from 2 waves of the main Labour Force Survey (LFS), collected on a local sample boost. These boosts are sponsored by the Department for Work and Pensions (DWP), Department for Business, Innovation and Skills (BIS), the Welsh Government and the Scottish Government. The APS is a widely used source for many government statistics. For the note on methodology see: <https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/methodologies/annualpopulationsurveyapsqmi>

Additionally, in Table 4, we show estimates of the groups identified by ONS as hard-to-reach, who are known to have relatively high non-response in the 2001 census and who were targeted in the design of the 2011 census (Abbot and Compton 2014). These groups have been originally classified by ONS into three priority groups (high, medium, low) based on their size and evidence of non-response. We calculate *Understanding Society* estimates of these hard-to-reach subgroups a based on *Understanding Society* Waves 1 and 8. Estimates from Wave 1 are compared with the 2011 England and Wales census figures. For Wave 8, we only provide estimates based on *Understanding Society* as there are no administrative data or official ONS statistics for 2016/2017 that we could compare *Understanding Society* estimates against to assess their accuracy.

In the second part of Section 1, we compare household level response rate of *Understanding Society* GPS sample in Wave 1 and Wave 8 in different types of areas (Table 5). Area characteristics are measured on LSOA-level and the estimates are based on the 2011 Census data for England and Wales except for: IMD quintiles (data obtained from Ministry of Housing and Local Government, restricted to England only) and share of Jobseeker's Allowance Claimants in LSOA (data based on records of JSA claimants from Jobcentre Plus local offices<sup>4</sup>). The presented estimates of household response rate are restricted to the GPS sample because Ethnic Minority Boost (Wave 1) and Immigrant and Ethnic Minority Boost (Wave 6) sample components are oversampled from areas of high concentration of ethnic minorities. Those areas tend to be disproportionately more likely to be located in economically worse-off areas and in the areas with higher immigration level, therefore the relationship between the variables considered here and household non-response cannot be examined in a comparable way to the GPS sample. The base for the household response rate analysis consists of all sampled households/addresses located in England and Wales known or assumed to be eligible for the survey (General Population Sample only). The household response rate estimates are expressed in unweighted percentages.

In Section 2, we assess *Understanding Society* coverage of small subgroups, traditionally considered as outside the scope of household surveys by looking at the estimated shares and/or frequencies of: (i) international migrants (Table 6), (ii) foster children (Table 7), (iii) people living/moving into institutions (Table 8); and (iv) people subject to involuntary house moves (Table 9) based on *Understanding Society* total sample (including GPS, BHPS, EMBS and IEMB components) in each wave.

We discuss the coverage of (i)-(iv) in the context of administrative estimates of stocks and flows of small hard-to-reach groups (migrants, children in care, people living in care homes, prisoners, homeless living in temporary accommodations) estimated based on available administrative data (details of the administrative estimates are available in Tables A1-A4 in the Appendix) for years 2010 to 2017, which broadly correspond to *Understanding Society* fieldwork period for Waves 1 to 8. It is important to keep in mind, however, that

---

<sup>3</sup> Labour Force Survey (LFS) is a large, cross-sectional survey based on a random sample of UK households. LFS collects the data quarterly and is recommended by ONS as one the main sources for UK labour market statistics. For the note on methodology see: [https://webarchive.nationalarchives.gov.uk/20110827023812/http://www.statistics.gov.uk/downloads/theme\\_labour/What\\_exactly\\_is\\_LFS1.pdf](https://webarchive.nationalarchives.gov.uk/20110827023812/http://www.statistics.gov.uk/downloads/theme_labour/What_exactly_is_LFS1.pdf)

<sup>4</sup> Data obtained from nomis- government official labour market statistics website

*Understanding Society* single wave fieldwork spans over more than 2 calendar years, therefore, the comparisons with annual administrative estimates are only approximate. Given that the main goal of the presented analysis is to assess wave to wave changes, we present the results by wave rather than by calendar years.

In relation to:

- (i) International migrants (defined as non-UK born individuals), *Understanding Society* figures are based on the total UK sample of persons with completed individual or proxy interview at each wave. Administrative estimates are based on combined official data for constituent UK countries obtained from: ONS Migration Statistics Quarterly Reports (for England), StatsWales (for Wales), Northern Ireland Statistics and Research Agency (for Northern Ireland) and National Records of Scotland (for Scotland). For each year, the share of migrants is calculated in relation to mid-year population estimates. We assume here that the proportion of the general population and migrant population living in households each year is approximately the same as the proportion of the general UK population living in households reported in the 2011 censuses.
- (ii) *Understanding Society* estimates of foster children are based on the total sample of all enumerated persons aged 0-17 living in England at the time of the interview. Administrative estimates are based on combined official data for constituent UK countries obtained from: Department of Education (England), Department of Health (Northern Ireland), Children's Social Work Statistics (Scotland) and Welsh Government (Wales). For each year, the share of children in care is calculated in relation to mid-year population estimates of children aged 0-17. We assume here that the proportion of children aged 0-17 living in households each year is approximately the same as the proportion of children aged 0-17 reported in the 2011 censuses.
- (iii)-(iv): People living/moving into institutions and people experiencing involuntary house moves, we present unweighted frequencies of the number of observations available in each wave of the total *Understanding Society* UK sample as well as cumulative number of unique observations across waves. We do not attempt to calculate population estimates here as the number of observations is small, and, there are no reliable, population-level external estimates that we could compare *Understanding Society* figures with.

In the second part of Section 2, we investigate the frequencies of responses to questions asked about those who leave the Study (leavers) and asked to those who join the study (joiners) (Tables 10-11).

## Representativeness of *Understanding Society* sample by subgroups traditionally considered as hard to survey

Previous assessment of *Understanding Society* representativeness in relation to sample coverage (Lynn and Borkowska 2018) compared the *Understanding Society* GPS sample estimates of some of the key sociodemographic characteristics (namely gender, age group, ethnicity, region, economic activity, having limiting long-term illness and having car or van in the household) from Wave 1 (2009/2010) with the 2011 Census figures. This comparison has shown that *Understanding Society* General Population Sample (GPS) aligns well with the census estimates with the exception of small under-representation of males (45.4% *Understanding Society* Wave 1 compared to 48.6% 2011 census), people living in Greater London (9.9% *Understanding Society* Wave 1 compared to 12.8% 2011 census), and people with a severely limiting long-term illness (8.6% *Understanding Society* Wave 1 compared to 10.3% 2011 census).

In this section, we look at Wave 8 (2016/2017) *Understanding Society* estimates in terms of gender, age group, region, ethnicity and economic activity. We do not include estimates of the proportion of people with limiting long-term illness or those with/without car or van in the household based on Wave 8 because, to the best of our knowledge, there are no official annual estimates that we could compare them against (e.g. ONS does not produce annual population estimates for these variables).

The results presented in Table 1 show that, in general, *Understanding Society* continues to be cross-sectionally representative in terms of sex, age groups and geographical regions. There is a small underrepresentation (up to 2 percentage points) of younger age groups (under 40) and, small overrepresentation of older age groups (up to 1.5 percentage points) compared to 2017 annual population estimates.<sup>5</sup> The Wave 8 sample size of each 10 year cohort is about 6,000 people. In terms of representativeness of people living in different geographical regions, there is a small underrepresentation of people living in London and small overrepresentation of people living in Yorkshire and the Humber (about 1 percentage point).

Compared to the 2017 Annual Population Survey, Wave 8 of *Understanding Society* seems to have some underrepresentation of ethnic minority groups (less than 1 percentage point in case of Indian, Pakistani/Bangladeshi and Black/Black British groups) and some overrepresentation (about 1.5 percentage point) of the 'other category'. These differences, however, especially in the case of 'other' category might be partially explained by the differences in coding of the ethnicity variable (e.g. *Understanding Society* categories had to be collapsed in order to be comparable to the smaller number of ethnicity categories available in APS). Another reason for under-representation of ethnic minorities is the cumulated effect of differential non-response over 8 waves for one part of the ethnic minority sample (EMB component) and over 3 waves for the other part of the sample (IEMB component). The new, improved weighting (currently under development) aims to correct for some of these issues.

---

<sup>5</sup> Mid-year population estimates are considered as a source of 'gold standard' benchmark estimates, however, it is important to keep in mind that they are also subject to regular revisions.

Compared to the Labour Force Survey estimates from December 2016-February 2017, *Understanding Society* Wave 8 estimates differ up to 3.5 percentage points in terms of categories of economic activity. Differences in estimates of the share of 'employed' and 'economically inactive' are likely to be at least partially explained by the differences in the definitions used. The breakdowns from *Understanding Society* are based on self-declared main economic activity. This means that some students, for example, might be economically active but still report 'full-time student' as their main economic activity. It is also important to keep in mind that both 'ethnicity' and 'economic activity' benchmark estimates used here are based on survey data and therefore are also subject to all potential sources of survey errors.

Table 1: *Understanding Society* Wave 8: Sex, Age and Region

	<i>Understanding Society</i> Wave 8 (2016/2017)		2017 Mid-year Population Estimates %
	Obs	Wtd % , 95% CI	
<b>Sex</b>			
male (16+)	17,923	47.9 [47.4, 48.3]	48.9
female (16+)	21,366	52.1 [51.7, 52.6]	51.1
Total	39,289		
<b>Age Group</b>			
16-19 years old	2,329	6.1 [5.8, 6.4]	5.6
20-29 years old	5,118	13.6 [13.1, 14.2]	16.3
30-39 years old	5,848	14.4 [13.9, 15.0]	16.3
40-49 years old	6,982	16.2 [15.7, 16.7]	16.1
50-59 years old	6,929	17.7 [17.2, 18.2]	16.5
60-69 years old	5,886	14.9 [14.4, 15.4]	13.2
70 years or older	6,197	17.1 [16.5, 17.7]	15.9
<b>Region (GOR)</b>			
North East	1,355	4.4 [4.0, 4.8]	4.1
North West	3,963	11.1 [10.4, 11.8]	11
Yorkshire and the Humber	3,461	9.1 [8.5, 9.8]	8.2
East Midlands	2,717	7.8 [7.2, 8.3]	7.3
West Midlands	3,355	8.8 [8.2, 9.3]	8.8
East of England	3,192	9.7 [9.0, 10.3]	9.3
London	5,411	10.9 [10.2, 11.6]	13.1
South East	4,535	13.8 [13.1, 14.6]	13.7
South West	3,015	8.9 [8.3, 9.4]	8.6
Wales	2,522	4.7 [4.4, 5.1]	4.8
Scotland	3,198	8.1 [7.6, 8.7]	8.4
Northern Ireland	2,550	2.8 [2.6, 3.0]	2.8

Note: *Understanding Society* figures based on persons with completed individual or proxy interview (therefore limited to people aged 16+ at time of Wave 8) from the total Wave 8 sample (including GPS, BHPS, EMBS and IEMB components). Weighted proportions use

h\_indpxui\_xw weight (adult main and proxy interview cross-sectional weight). Numbers of observations are unweighted. Mid-year population estimates use the figures for the UK collated by ONS. Mid-year population estimates refer to the total resident population, whereas Understanding Society only aimed to represent the household population, which accounted for approximately 98% of the total population aged 16 or over in 2011. ONS does not produce confidence intervals for mid-year population estimates, therefore only point estimates are reported here.

Table 2: *Understanding Society* Wave 8: Ethnicity

Ethnicity	<i>Understanding Society</i> Wave 8 (2016/2017)		2017 Annual Population Survey %, 95%CI
	Obs	Wtd% , 95%CI	
White (British)	29,600	88.9 [88.3, 89.4]	88.1 [88.0, 88.2]
Indian	1,603	1.8 [1.6, 2.0]	2.4 [2.3, 2.5]
Pakistani/Bangladeshi	2,290	1.6 [1.4, 1.8]	2.4 [2.3, 2.5]
Black or Black British	1,742	1.8 [1.6, 2.0]	2.7 [2.6, 2.8]
Mixed	721	1.1 [1.0, 1.3]	1.1 [1.0, 1.2]
Other	2,247	4.8 [4.4, 5.3]	3.3 [3.2, 3.4]
Total	38,203		

Note: *Understanding Society* estimates calculated as in Table 1. APS estimates based on household population aged 16+. For details of APS survey see: Methods and data section.

Table 3 *Understanding Society* Wave 8: Economic Activity

Economic activity	<i>Understanding Society</i> Wave 8 (2016/2017)		Labour Force Survey (Dec-Feb 2017) %
	Obs	Wtd %, 95%CI	
Employed	21,978	56.1 [55.3, 56.8]	60.5
Unemployed	1,566	3.8 [3.5, 4.2]	4.6
Inactive	15,630	40.1 [39.4, 40.8]	36.5
Total	39,174		

Note: *Understanding Society* estimates calculated as in Table 1. Figures from LFS obtained from the ONS Labour Market statistics summary data tables published on 16<sup>th</sup> April 2019 (<https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentanddemoployeetypes/bulletins/uklabourmarket/april2019/relateddata> ) The ONS summary statistics do not include confidence intervals, therefore only point estimates are reported here.

The estimates of the hard-to-reach groups (Table 4) show, quite reassuringly, that the estimated subgroup shares (in the population of England and Wales) based on Wave 1 of *Understanding Society* sample broadly align with the 2011 Census estimates. In terms of the estimated shares of younger and older populations, there is small underrepresentation of students and small underrepresentation of the 80+ category. The estimates of the share of particular ethnic minority groups are aligned well with the 2011 census estimates, and, each of the ethnic groups identified as hard-to-reach (except Chinese ethnic group) is represented in the Wave 1 sample by over 1,000 individuals allowing meaningful subgroup comparisons

(the number of observations is slightly lower in Wave 8). The share of long-term migrants (defined as non-UK born individuals who plan to stay in the UK for the next 6 months) is by about 2 percentage points under-estimated in Wave 1 compared to 2011 Census. The cross-sectional sample sizes of the non-UK born individuals are relatively large across waves— 9,335 residents of England and Wales in Wave 1 sample and 6096 in Wave 8 sample. There is also less than three percentage point underestimation of people living in private-rented accommodation and less than 2 percentage point overestimation of households with pre-school children compared to the 2011 census figures. In relation to Wave 8 sample, the estimated shares of some ethnic minority groups and non-UK born individuals are lower than what we would expect. As mentioned in the previous paragraph, the new revised weights aim to correct some of these issues.

Table 4 *Understanding Society* Wave 1 and Wave 8 by key societal subgroups considered as hard-to-reach

Priority (as reported by Abbot & Compton 2014)	Group	2011 Census (England and Wales) %	Wave 1 Obs	Wave 1 GPS + EMB sample Wtd%, 95% CI	Wave 8 Obs	Wave 8 GPS + BHPS + EMB + IEMB Wtd%, 95% CI
<b>High</b>	Young adults aged 20-29	16.8	8,154	17.0 [16.5, 17.5]	5,118	13.6 [13.1, 14.2]
	Students*	8.2	3,497	6.9 [6.5, 7.3]	2,167	6.1 [5.8, 6.5]
	Short-term migrants**	N=181,412				
	Bangladeshi eth group	0.6	1,181	0.5 [0.48, 0.61]	1,471	0.5 [0.4, 0.6]
	Black African eth group	1.5	1,478	1.4 [1.3, 1.5]	920	1.2 [1.0, 1.4]
<b>Medium</b>	Black Caribbean eth group	1.1	1,182	1.0 [0.89, 1.04]	750	0.7 [0.6, 0.8]
	Indian eth gr	2.5	1,951	2.6 [2.4, 2.8]	1,592	1.9 [1.7, 2.2]
	Pakistani eth group	1.7	1,513	1.4 [1.3, 1.5]	1,471	1.2 [1.0, 1.4]
	Chinese eth gr	0.8	315	0.4 [0.3, 0.5]	150	0.3 [0.2, 0.4]
	Long-term migrants (non-UK born)	15.2	9,335	12.9 [12.4, 13.3]	6,096	9.8 [9.3, 10.4]
<b>Low</b>	Multiple- occupied dwellings (hh-level)	Shared dwelling 0.3	2	0.009 [0.002, 0.04]		
	Pre-school (under 5 –hh level)	11.8	4,101	13.5 [13.1, 13.9]	2,078	9.4 [8.8, 10.0]
	Private-rented accommodation	17.3	7,082	14.7 [14.1, 15.3]	4,036	13.2 [12.4, 13.9]
	One-person household	15.9	6,545	16.4 [15.9, 16.9]	4,727	16.1 [15.5, 16.7]
	Low-income household	No data***	***		***	
Adults aged 80+	5.7	1,691	5.2 [4.9, 5.5]	1,656	6.1 [5.7, 6.5]	

Note: *Understanding Society* figures based on persons with completed individual or proxy interview (therefore limited to people aged 16+ at the time of the interview) living in England and Wales. Weighted proportions use adult main and proxy interview cross-sectional weights (except for two household-level estimates: households in multiple-occupied dwellings and households with pre-school children, which are weighted by household grid weights). Numbers of observations are unweighted. The 2011 census

estimates refer to England and Wales (comparable estimates exist for Northern Ireland and Scotland but combining them is not straight forward therefore the analysis here has been limited to England and Wales only). All census estimates except for proportion of households in 'multiple-occupied dwellings' and the 'proportion of households with pre-school children' are based on all usual residents aged 16+ living in England and Wales. Census estimates refer to the total resident population, whereas *Understanding Society* only aimed to represent the household population, which accounted for approximately 98% of the total population aged 16 or over in 2011. \*The category of 'students' is defined slightly differently in the 2011 census and in *Understanding Society*. The census category refers to all full time students (whether economically active or inactive), whereas *Understanding Society* category refers to those full-time students who self-declared being a student as their main economic activity. \*\*Short-term migrants are defined as those living in England and Wales who were born outside the UK, and who intended to stay in the UK for a period of between 3 and 12 months. \*\*\*There is no data/estimates based on the 2011 census on low-income households. *Understanding Society* Income Annex (forthcoming) discusses in detail how the Study estimates align with other major UK surveys.

## Household response by type of areas

Another important aspect of good survey coverage is successful field operations that maximise response and minimise differential non-response. Although the existing *Understanding Society* weights account for differential non-response in order to ensure that the population-level estimates are not biased, we recognise the importance of maintaining high response rate among all societal subgroups and all types of geographical areas. In this section we look at the household-level response rate in different types of areas by using auxiliary information obtained from administrative data. The previous assessment of Wave 1 household-level response in different types of areas (Lynn et al. 2012) showed that there is only a modest association between household-level response rate and a few of the examined area-level characteristics. All the assessed characteristics used Medium-level Super Output Area (MSOA) as a geographical unit, and, most of the data used were based on the 2001 census (2011 census was not available at the time). The results showed that, in Wave 1, there was a slightly lower response rate in areas with relatively high share of single-person households, relatively high full-time employment (up to 3 percentage points difference between the first and the last quartile), relatively high proportions of people in higher managerial and professional occupations (up to 4 percentage point difference), relatively high burglary rate (up to 6 percentage point difference), and where lower share of people drive to work (up to 5 percentage point difference).

In this paper, we look at household-level response rate in Wave 1 and Wave 8 of the GPS sample by a number of additional small-area level indicators. The presented indicators cover some of the most important area-level characteristics included in the ONS 2009 Hard-To-Count Index (HtC)<sup>6</sup>. The 2009 HtC index was developed to predict non-response in the 2011 Census and used as the main tool for allocation of resources for the census field operations. HtC index has been also shown to be associated with household non-response in General Household Survey, Labour Force Survey, Family Resource Survey and Expenditure and Food

---

<sup>6</sup> For details of the methodology of the 2009 Hard-to-Count index see: <https://www.ons.gov.uk/file?uri=/census/2011census/consultationsusersandlocalpartners/censusadvisorygroups/censusgeneralcag/ag0917tcm77190524.pdf>

Survey (ONS 2009). Similar to the 2009 HtC index methodology, we use small-level geographical units (Lower level Super Output Areas-LSOAs). Table 5 shows the unweighted household-level response rate by quintiles of LSOA-level measures of: (a) 2010 Index of Multiple Deprivation, (b) share of unemployed in the area, (c) share of inactive in the area, (d) share of Job Seeker's Allowance Claimants (JSA), (e) Share of non-UK born, (f) share of non-White British, (g) dwelling density quintiles, (h) share of residents aged 16-29.. The above measures can be generally grouped into two types: those related to the economic conditions of the area (a-d), and those related to mobility/population turnaround in the area (e- h).

In relation to the latter group of measures, the results show that in the initial wave, there was a moderate negative association between the share of non-UK born and non-White British in LSOA and household response rate (up to 10 percentage points difference between the first and the last quintile; see: Table 5). Most of this difference has been due to the higher rates of non-contact rather than refusal, which is not surprising given that such areas have higher turnaround of the population, which is likely to make the initial contact more challenging. Interestingly, at the latest wave (Wave 8), the negative association between the share of non-white British or foreign born in the area and response rate almost completely disappeared (except the quintile with the very highest share of either of these two groups, in which the response rate was slightly lower than in other areas). These results suggest that the field operations have been equally successful in re-contacting people living in highly mobile areas as in less mobile areas once the initial contact has been made. Another area characteristic modestly associated with lower household response rate at Wave 1 was relatively high dwelling density (up to 7 percentage point difference between the first and the last quintile). This is generally in line with the literature, which shows that people living in more busy places (e.g. big cities) are less likely to respond to surveys<sup>7</sup>.

Variables related to the economic condition of the area such as quintiles of economic inactivity, unemployment, share of JSA claimants, and Index of Multiple Deprivation were not related to the overall household response rate at Wave 1 but they seemed to be moderately associated with the nature of non-response. Economically worse-off areas tended to have a relatively higher rate of non-contact whereas better-off areas tended to have a relatively higher rate of refusals (see: Table 5). This relationship between the nature of non-response and economic condition of the area was also noticeable at the latest wave (Wave 8) and resulted in a slightly lower overall household response-rate in less well-off areas. This suggests that people living in economically deprived areas might be more difficult to contact, but, once the successful contact is established, they are at least equally likely (or even slightly more likely) to agree to participate in a survey compared to those living in better-off areas. The geographical spread of different types of areas in terms of share of non-UK born (variable most strongly associated with household response rate among the examined mobility-related characteristics) and economic inactivity rate (variable most strongly associated with household response rate among the examined economic characteristics) is illustrated in Figure 1.

---

<sup>7</sup> For an overview of the existing evidence see: ONS (2009) "Predicting Patterns of Household Non-Response"



Table 5 Household response rate by small area characteristics

Wave 1 (GPS sample)					Wave 8 (GPS sample)				
	Response	Non-contact	Refusal	Other		Response	Non-contact	Refusal	Other
Total (Eng)	57.4	7.4	34.3	1.0	87.7	3.9	5.0	3.4	
Total (Eng & Wales)	57.7	7.5	33.9	0.9	87.3	4.1	5.1	3.6	
<b>2010 IMD quintiles (LSOA-level, England only)</b>									
Q1 - most deprived	57.8	11.2	29.1	2.0	84.4	6.4	5.6	3.7	
Q2	56.1	8.5	34.1	1.3	87.1	4.8	4.7	3.4	
Q3	57.4	6.7	35.1	0.8	87.2	3.6	5.6	3.6	
Q4	58.0	5.8	35.9	0.4	89.1	2.9	4.6	3.4	
Q5 - least deprived	57.6	4.8	37.3	0.4	89.8	2.5	4.5	3.2	
<b>Unemployment quintiles (LSOA-level, England and Wales)</b>									
Q1- lowest unemployment rate	58.8	5.3	35.7	0.3	88.7	2.8	5.0	3.6	
Q2	57.8	5.8	36.0	0.4	88.4	3.3	4.8	3.5	
Q3	57.1	7.2	35.0	0.7	88.3	3.3	4.7	3.7	
Q4	56.9	8.4	33.5	1.2	85.8	5.1	5.5	3.6	
Q5- highest unemployment rate	57.9	10.8	29.3	2.0	84.6	6.6	5.5	3.3	
<b>Inactivity quintiles (LSOA-level, England and Wales)</b>									
Q1-lowest inactivity rate	58.8	5.5	35.2	0.5	89.4	2.9	4.2	3.5	
Q2	57.1	5.9	36.6	0.4	88.7	2.9	4.7	3.7	
Q3	57.5	6.9	34.9	0.7	87.5	3.9	5.5	3.2	
Q4	56.6	9.0	33.2	1.1	86.8	4.6	4.8	3.8	
Q5- highest inactivity rate	58.4	10.2	29.4	2.0	83.1	6.7	6.4	3.7	

**Share of JSA claimants (2011) quintiles (LSOA-level, England and Wales)**

	Response	Non-contact	Refusal	Other	Response	Non-contact	Refusal	Other
Q1- lowest share of JSA claimants	58.2	5.1	36.4	0.2	88.5	2.7	5.2	3.7
Q2	57.8	5.6	35.9	0.6	89.0	2.6	4.8	3.6
Q3	57.4	6.8	35.2	0.6	87.7	3.9	4.8	3.6
Q4	57.0	8.6	33.3	1.2	86.6	5.2	5.1	3.1
Q5- highest share of JSA claimants	57.9	11.4	28.7	2.0	84.1	6.6	5.7	3.7

**Share of non-UK born quintiles (LSOA-level, England and Wales)**

Q1- lowest share of non-UK born	61.0	5.9	32.9	0.3	88.2	3.8	4.8	3.2
Q2	60.3	5.7	33.6	0.4	87.9	3.5	4.9	3.7
Q3	58.4	6.6	34.3	0.7	87.2	3.8	5.3	3.7
Q4	57.2	7.5	34.6	0.7	87.4	4.0	5.0	3.6
Q5- highest share of non-UK born	51.1	12.1	34.1	2.7	85.1	5.8	5.4	3.7

**Share of non-White British quintiles (LSOA-level, England and Wales)**

Q1- lowest share of non-White British	61.5	5.4	32.9	0.2	88.2	3.4	5.0	3.4
Q2	60.0	5.7	33.8	0.5	87.7	3.8	5.0	3.5
Q3	58.4	6.7	34.4	0.6	87.4	4.0	4.9	3.7
Q4	56.7	8.0	34.6	0.7	87.0	4.2	5.3	3.5
Q5- highest share of non-White British	51.3	12.0	33.9	2.8	85.4	5.6	5.3	3.8

**Dwelling density<sup>8</sup> quintiles (LSOA-level, England and Wales)**

	<b>Response</b>	<b>Non- contact</b>	<b>Refusal</b>	<b>Other</b>	<b>Response</b>	<b>Non- contact</b>	<b>Refusal</b>	<b>Other</b>
Q1- lowest dwelling density	60.9	5.9	32.9	0.3	88.2	2.9	5.1	3.8
Q2	59.6	5.9	33.9	0.6	88.0	4.0	4.8	3.3
Q3	57.4	6.7	35.2	0.7	88.0	3.4	5.2	3.4
Q4	57.1	7.4	34.5	0.9	86.1	5.0	5.5	3.4
Q5 - highest dwelling density	53.3	11.7	32.9	2.2	85.8	5.6	4.7	4.0

**Share of residents aged 16-29 quintiles (LSOA-level, England and Wales)**

Q1- lowest share of 16-29	58.2	4.8	36.8	0.3	88.9	2.5	5.0	3.6
Q2	58.9	5.5	35.2	0.4	89.1	3.0	4.7	3.2
Q3	58.2	6.7	34.6	0.6	87.0	3.9	5.6	3.4
Q4	57.5	8.4	33.1	1.0	85.7	5.7	5.2	3.4
Q5-highest share of 16-29	55.7	12.2	29.8	2.3	85.0	5.9	4.8	4.3

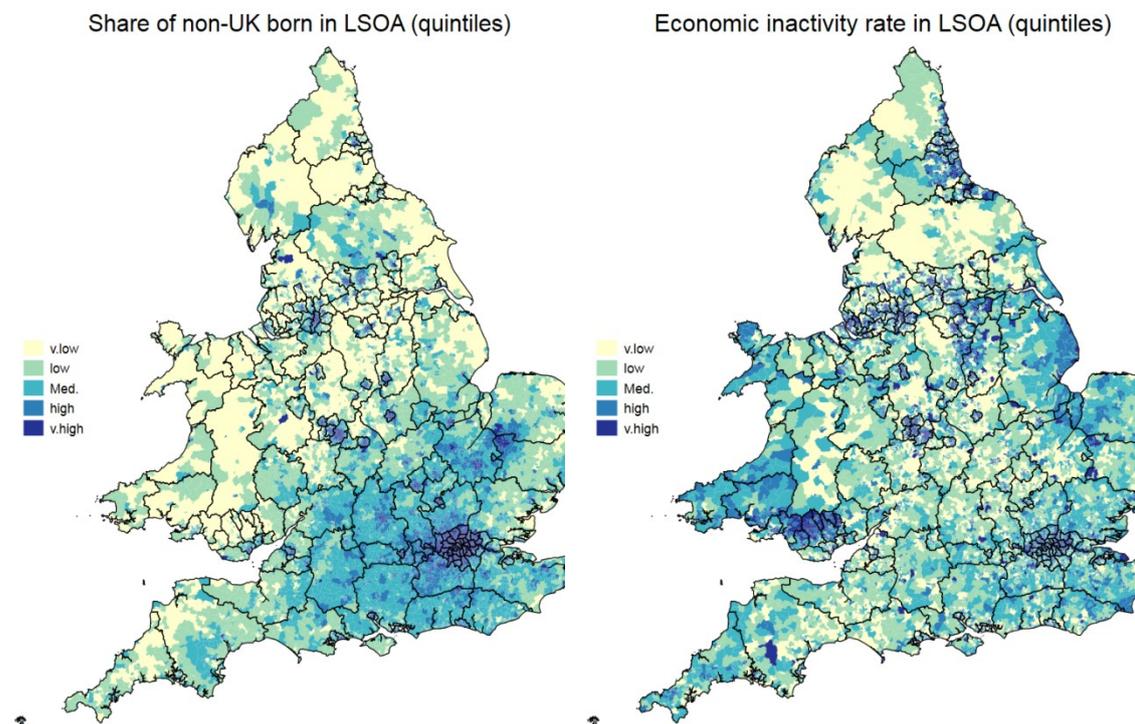
Note: Unweighted percentages.

Figure 1 Distribution of LSOA-level share of non-UK born (left) and economic inactivity rate (right) in England and Wales based on 2011 Census

---

<sup>8</sup> Dwelling density calculated by dividing number of households by the land-based area in hectares of each LSOA (based on 2011 census data)

## Potential directions for improving coverage of large hard-to-reach subgroups



First, as noted in the previous section, there is some evidence suggesting that the rates of non-contact with the sampled households are slightly higher in areas with high turnaround of the population (areas with high shares of non-UK born and non-White British residents) and in the areas that are relatively less well-off and more detached from the labour market. This suggests that it might be worth putting more resources into field operations in these two types of areas, especially given that the presented descriptive evidence suggests that once the initial contact is established, people living in those areas are equally likely (as those in other areas) to agree to take part in the Study. Given that the lower household response rate in highly mobile places was more pronounced in the initial wave, paying special attention to these type of areas would be especially important in case of the new boost samples.

One of possible changes might be increasing the minimum number of contact attempts and ensuring that fieldwork agency always makes the required number of contact attempts before the household can be coded as non-contact. Currently, the mean number of call attempts is 9, and the median is 6. For households which end up with a non-contact outcome, the average number of calls is 12. However, there are still 6.6% of non-contact households with 5 or fewer call attempts, which is one of the issues we are currently trying to minimise.

## Coverage of subgroups traditionally considered as being outside the scope/sampling frame of household surveys

### International migrants

In this section we return to the *Understanding Society* coverage of migrants in more detail. Tracking highly mobile individuals in surveys is inevitably challenging and international migrants are one of the most prominent examples of highly mobile subgroups, e.g. they tend to move home more frequently than the natives, which makes tracking them continuously challenging. *Understanding Society* estimates and sample sizes of long-term migrants (defined as non-UK born individuals, see: Table 6) cover non-UK born population reasonably well. In Wave 1(2009/2010), the estimated share of non-UK born was around 12 percent of the total UK population, which is about the same as the share calculated based on the official 2010 administrative migration statistics<sup>9</sup>. The estimated proportion of long-terms migrants in subsequent waves somewhat diverges from the administrative migration statistics (the share of migrants based on *Understanding Society* data is underestimated). It is worth keeping in mind, however, that the definition of the *Understanding Society* population is slightly different than the definition of the population applied to estimate the share of migrants based on administrative data. Cross-sectional *Understanding Society* estimates from Waves 1 to 5 refer to the UK population that has been continuously living in the UK since 2009/2010 and estimates from Waves 6 to 8 refer to the UK population that has been continuously living in the UK since 2014/2015 whereas estimates based on administrative data refer to the UK population in a single year. Therefore, administrative estimates are designed to capture all new immigrants who arrive to the UK each year, which is not possible to do in a longitudinal survey. In terms of sample sizes,

*Understanding Society* contains between 6392 (Wave 8) and 9744 (Wave 1) non-UK born individuals in each wave, and in total 16,413 unique observations of non-UK born individuals that were present in the Study in at least one of the waves. This large sample size provides unique opportunities for exploring research questions related to migration and for analysing foreign-born population living in the UK.

Another aspect of migration-related coverage is the estimated share of those who emigrate from the country. At the moment, *Understanding Society* does not conduct interviews with sample members who moved abroad. Currently we only have 109 reported cases of the *Understanding Society* household leavers who have been reported as moved abroad during enumeration as this question was (so far) only asked in Wave 8 (see Table 10). However, our most recent funding award included resources to undertake surveys of those people who emigrate, and improve our identification of them. In time, therefore, these data should improve, as well as provide insights on international family networks.

---

<sup>9</sup> For the details of the administrative data used see: Methods and Data section

Table 6 International long-term migrants in the UK: administrative estimates of stock and flows between 2010 and 2017 (left) and *Understanding Society* cross-sectional estimates for Waves 1-8 (right)

Administrative estimates for the UK				<i>Understanding Society</i> estimates for the UK			
	Stock	Share flow in	Share flow out	Wave	Obs	Wtd % , 95%CI	Cumulative N of unique obs
2010	11.53	0.9	0.6	1	9,744	12.2 [11.8, 12.6]	9,744
2011	12.11	1.0	0.5	2	7,828	10.2 [9.8, 10.6]	11,153
2012	12.28	0.8	0.6	3	6,779	10.6 [10.2, 11.1]	11,643
2013	12.36	0.8	0.5	4	6,298	10.5 [10.0, 10.9]	12,018
2014	12.81	0.9	0.5	5	5,885	10.2 [9.8, 10.6]	12,262
2015	13.16	1.0	0.5	6	8,391	10.9 [10.3, 11.5]	15,688
2016	13.94	1.0	0.5	7	7,170	10.3 [9.7, 10.9]	16,159
2017	14.21	0.9	0.5	8	6,392	9.6 [9.1, 10.2]	16,413

Note: Weighted proportions use adult main and proxy interview cross-sectional weights. Numbers of observations are unweighted.

### Children looked after

Children in care<sup>10</sup> are another important subgroup to consider as their developmental and educational needs and outcomes are likely to be very different from children who live with their parents. United Nations/Lumos estimates suggest that there are about 8million children living in institutional care worldwide and that over 80% of them are not orphans (have at least one living parent)<sup>11</sup>. In the UK, based on the combined administrative data from Department of Education (England), Department of Health (Northern Ireland), Children's Social Work Statistics (Scotland) and Welsh Government (Wales), we estimated that in 2017 there were about 96,415 children looked after by local authorities (see Table A1 in the Appendix). Based on the administrative data for England, we estimate that almost three quarters of children in care live in foster care<sup>12</sup>. This means that the vast majority of

<sup>10</sup> The definition of looked-after children (children in care) is found in the Children Act 1989. A child is looked after by a local authority if a court has granted a care order to place a child in care, or a council's children's services department has cared for the child for more than 24 hours. On reaching the age of 18, children cease to be considered looked-after by a council.

<sup>11</sup> Statistics presented at the 2019 UN World Data Forum (<https://unstats.un.org/unsd/undataforum/index.html>)

<sup>12</sup> This estimate is based on the figures for England obtained from the Department of Education

children in care should be included in the sampling frame of *Understanding Society* (or any other household survey). Between 2010 and 2017, children in care constituted about 0.7 % of the total UK population of children aged 0-17(living in households)<sup>13</sup>; about 0.3% (of all children aged 0-17) were admitted to care each year (based on care orders), and about 0.3% were leaving care each year. Table 7 compares estimates of the share of foster children in England based on the administrative data with *Understanding Society* estimates based on the English subsample. Unfortunately, the precise estimates of foster children are not consistently provided for Scotland, Wales and Northern Ireland and thus the comparison had to be restricted to England only. The estimated proportion of foster children in relation to all children aged 0-17 based on the *Understanding Society* sample varies between 0.33% (Wave 1) and 0.15% (Wave 4). The population statistics for 2010-2017 suggest that foster children constituted about 0.44% of all children. This means that there is a small underestimation of this subgroup in *Understanding Society* sample; however, similar to the case of international migrants, it is important to keep in mind that there is a large turnaround of foster children population and therefore, some of them have stopped being looked after at some point during the 8 wave period. *Understanding Society* UK sample contains between 39 (Wave 8) and 67 (Wave 2) observations of foster children aged 0-17 in each wave, and in total, it contains 219 unique observations of foster children that were present in at least one of the waves. As shown in Table 7, the sample of foster children grows considerably with time, which provides promising opportunities for researchers interested in looking at the outcomes of foster children.

---

<sup>13</sup> See Table A1 in the Appendix

Table 7 Estimates of foster children in England based on: DfE data (left) and *Understanding Society* data (right)

Department of Education estimates (England only)				<i>Understanding Society</i> estimates (England only)			<i>Understanding Society</i> (UK)	
Year	Foster placements - % of the population of children (0-17yo)	Number of children in foster placements	Total looked after children % of the population of children (0-17yo)	Wave	Obs	Wtd %, 95%CI (base: all children 0-17yo)	Obs	Cumulative N of unique obs
2010	0.42	46,890	0.58	1	57	0.33 [0.22, 0.49]	66	66
2011	0.43	48,150	0.58	2	47	0.29 [0.2, 0.43]	67	104
2012	0.44	50,030	0.59	3	33	0.17 [0.1, 0.28]	59	133
2013	0.44	50,600	0.60	4	31	0.15 [0.09, 0.26]	59	158
2014	0.44	50,880	0.60	5	30	0.17 [0.1, 0.31]	48	174
2015	0.44	51,570	0.60	6	31	0.18 [0.11, 0.31]	42	193
2016	0.44	51,430	0.60	7	30	0.22 [0.14, 0.35]	46	210
2017	0.45	53,010	0.62	8	27	0.21 [0.12, 0.35]	39	219
Total N of unique observations					145		219	

Note: Estimated shares of foster children based on *Understanding Society* data are weighted by enumerated person cross-sectional weights. Numbers of observations are unweighted.

## People living in care homes

Another subgroup commonly considered as missing from household surveys is the population of people living in care homes<sup>14</sup>. Institutions are explicitly excluded from the sampling frame for the Study (and other household studies), however, people will move from *Understanding Society* households into care homes (and vice versa).

We estimate, based on combined data from NHS digital (for England), Social Care Wales, Scottish Government and Department of Health (Northern Ireland) that between 2008 and 2016, between 2.4% and 3.1% of the UK population aged 65 and over was living in care homes (the estimated percentage varies by year). Based on existing administrative data, it is not possible to calculate the share of older people going into or out of care homes for the whole of the UK but it is possible to calculate such estimates for Scotland. On average, in Scotland, about 1.4% of those aged 65 plus start living in care homes each year and about 0.3% leave care homes each year (see Table A2 in the Appendix). If the share of people aged 65 and over going into care homes each year was the same in Scotland as in the rest of the UK, and, if all *Understanding Society* sample members aged 65 and over had the same chances of going into care homes as the Scottish population, we should observe between 128 and 140 individuals aged 65 plus who move into care homes in each wave. Of course, the initial *Understanding Society* sample consisted only of those living in private households, therefore we do not have information on those living in care homes or other communal establishments in Wave 1 (except 19 cases, where members of the sampled households were reported as being away in care/nursing home, see: Table 10). In subsequent waves, however, efforts were made to conduct interviews with everyone who moved into all types of institutions (except for prison). Currently, the Study does not consistently collect detailed information about the type of institution that someone has moved into, therefore, it is not possible to provide precise number of sample members who moved into care homes in each wave (Note: We know that there were 24 cases of people moving into care homes in wave 2 and 29 such cases in wave 3 – see Table 10, variable *lvwhy*<sup>15</sup>). However, given that we have information about the age of the respondent, it is likely that many of those aged 65 and over who moved into institutions, moved into care homes. Table 8 shows that there were between 2 and 11 older sample members living in an institution in each wave for whom individual or proxy interview was obtained, and that, in total, we observed 36 older sample members who moved into institutions at one of the waves (and who participated in the survey). In Table 8, we also report the total number of people living in institutions (cumulative number of unique observations across waves and the number of observations in each wave) for whom individual or proxy interview was obtained as well as the number of households who were reported as having their address based in an institution in each wave. There are a couple of observations that one can make based on those numbers. First, about half of all interviews obtained in institutional settings are with people aged 65 plus, which suggests that care homes are likely to be the most common institutional setting that we have in our sample. Secondly, the number of households, which addresses are located in institutions is considerably larger than the number of interviews completed by people living in institutions. This is not surprising given that institutional settings might have many

---

<sup>14</sup> People living in care homes/accessing long-term support are defined as those whose intention at time of admission to a care home was to stay as a permanent resident, regardless of how long they actually stayed.

<sup>15</sup> The coding of the *lvwhy* variable is discussed in the later section

restrictions with respect to who can access their premises, and, if majority of them are indeed care homes, it might be that older people suffer from health conditions which limit their ability to complete an interview. The third possible reason of why we observe much fewer interviews with people living in institutions than we would expect (based on the number of households which addresses are located in institutions) is that interviewers might be less inclined and/or able to conduct interviews in such settings. To date, exploring institutional settings and circumstances of people living in them has not been the focus of the Study, however, as part of the most recent funding, we included plans for an ‘exit’ interview for all those (or a relative) who move into institutions. This will enable us to collect more information about the type of institutions that people live in and hence the feasibility of obtaining interviews from people living in institutions in the future if we putting more resources to this.

**Table 8 *Understanding Society* sample members who live in institutions by wave**

Wave	Number of people aged 65+ living in institutions who completed an individual or proxy interview (N of full individual interviews in brackets)	Cumulative N of unique observations of people aged 65+ living in institutions who completed an individual or proxy interview	Total number of people living in institutions for who completed an individual or proxy interview (N of full individual interviews in brackets)	Cumulative N of unique observations of the total number of people living in institutions who completed an individual or proxy interview	Total number of sampled households for whom identified address was an institution
w2	2 (1)	2	6 (5)	6	18
w3	8 (7)	9	17 (16)	21	48
w4	7 (5)	12	10 (8)	24	17
w5	9 (7)	20	18 (16)	39	42
w6	11 (11)	27	24 (24)	56	97
w7	5 (5)	31	13 (13)	66	27
w8	9 (9)	36	16 (16)	78	34

Note: Unweighted frequencies based on the total UK *Understanding Society* sample (including GPS, BHPS, EMBS and IEMB components)

### **Prisoners**

Currently, there are 32 cases of *Understanding Society* household leavers (from waves 4-8) who have been reported as now living in a prison and 30 out of these 32 cases are unique based on pidp. At Wave 1, there were additionally 6 members of the originally sampled households, who were reported as being away in prison or young offenders’ institution. Since Wave 2, at each wave, there have been between 1 and 17 reported cases of household members being away in prison. If *Understanding Society* sample members aged 18 and over had identical chances of going into prison as the general UK population (we estimated that around 0.23% of the total adult population aged 18 plus is admitted to prison every year

based on the administrative data for England and Wales<sup>16</sup>), we should observe between 88 and 113 individuals at each wave leaving the Study and being admitted to prison. Similarly, if new joiners had identical chances of being released from prison as the general population (around 0.18% of the total adult population aged 18 plus is being released from prison every year), we should observe (based on the absolute number of new joiners and excluding re-joiners), about one to two ex-prisoners joining the Study at each wave. Since *Understanding Society* does not conduct interviews in prisons, tracking sample members who become prisoners has not been the main focus of the Study to date. Given the likely reluctance of survey participants to report cases of their household members being admitted to prison because of the associated social stigma, it is not surprising that the numbers of reported prison admissions are lower than what we would expect based on the general population statistics.

## Homeless

Homeless are particularly vulnerable and difficult to count population. There are a few types of homelessness according to the official government classification. The first group - rough sleepers- are defined as “people sleeping, about to bed down or bedded down on the street, in doorways, parks, tents, bus shelters, cars, barns, shed and other places not designed for habitation”<sup>17</sup>. Although authorities in all four UK countries do make an effort to count and estimate the number of rough sleepers, according to the UK Statistics Authority, this numbers are not reliable and not comparable between UK countries due to the differences in methodology used. *Understanding Society* also does not conduct interviews with people without address of residence, therefore this category of homeless is currently not possible to investigate further. The second group - statutory homeless - defined as “households over the course of a year which the local authority has agreed it has a duty to house under the 1996 Housing Act”<sup>18</sup> is easier to capture in administrative statistics. However, not all statutory homeless are housed by local authorities<sup>19</sup>. Those who are, can be placed in ‘temporary accommodation’ for any duration of time (from one night to indefinite). Type of ‘temporary accommodation’ might be one of the following: night/winter shelters, hostels, B&Bs, woman’s refuges and private and social housing. Based on the MHCLG statistics for England (see Table A3 in the Appendix), we estimated that between 2004 and 2016 about 0.3 % of all households (in England) were living in temporary accommodation. *Understanding Society* does not currently collect information on whether someone was placed in temporary accommodation under the 1996 Housing Act, however, those placed in social or private housing could potentially be part of the initial *Understanding Society* sample, and those who

---

<sup>16</sup> Figures for England and Wales obtained from the Ministry of Justice Offender Management Statistics

<sup>17</sup> Ministry of Housing, Communities & Local Government “Rough Sleeping Statistics 2018”, retrieved from:  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/781567/Rough\\_Sleeping\\_Statistics\\_2018\\_release.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/781567/Rough_Sleeping_Statistics_2018_release.pdf)

<sup>18</sup> Local Authority Homelessness Statistics (England)  
<https://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-7586#fullreport>

<sup>19</sup> Until 2018 it was mainly households in ‘priority’ groups, since 1<sup>st</sup> April 2018, when the Homelessness Reduction Act 2017 came into force in England, the right to being housed by local authorities has been extended.

moved from private household addresses into hostels and B&Bs could have been interviewed in later waves (although as shown in the previous section, we currently have a very small number of interviews conducted in institutional/communal establishment settings). We do have, however, information that could potentially help to identify people in vulnerable housing situations. This is especially important given that, according to MHCLG, most homelessness is hidden. As noted by CRISIS (charity devoted to the issue of homelessness), “this means staying with family and friends, sofa surfing, living in unsuitable housing such as squats or in ‘beds in shed’ situations (The Homelessness Monitor: England 2018). All these situations leave the person extremely vulnerable. The majority of the hidden homeless will have slept rough at some time (The hidden truth about homelessness, 2011)”<sup>20</sup>. One type of information collected in *Understanding Society* that can be used to identify people in vulnerable housing situations is the question about ‘the reason why someone moved houses’ which include types of involuntary moves. As shown in Table 9, the declared reasons for involuntary moves between Wave 2 and Wave 8 included: being evicted from rented accommodation (between 31 and 61 observations per wave); moving because of health reasons (between 12 and 42 observations per wave); moving to accommodation with no stairs (between 2 and 18 observations per wave); and moving to sheltered accommodation/institution (between 1 and 11 observations per wave). In total we observed 312 unique individuals who were evicted from rented accommodation across waves, 232 individuals who moved because of health reasons, 48 who moved to accommodation with no stairs, and 22 who moved to sheltered accommodation. Similar to the case of foster children, the sample of people who, at some point, were evicted from rented accommodation increases quite considerably with time, which might provide unique opportunities for housing researchers interested in looking at people in particularly vulnerable housing situations.

Furthermore, *Understanding Society* also contains information on whether someone is living in a concealed household<sup>21</sup> (on average, across waves, about 30% of sample members live in concealed households; estimates not shown here), which could also be a sign of potentially undesirable and unstable housing situation (although not always); and we collect information on events (such as relationship breakdowns), which might increase a person’s chances of being in a vulnerable housing situation. According to the House of Commons briefing paper (Statutory Homelessness in England 2019<sup>22</sup>), the three most common reasons cited by people who applied for temporary accommodation are: family and friends no longer able or willing to accommodate the household (24%), the end of an Assured Shorthold Tenancy (AST) in the private rented sector (20%), and relationship breakdown (19%).

---

<sup>20</sup> <https://www.crisis.org.uk/ending-homelessness/homelessness-knowledge-hub/types-of-homelessness/>

<sup>21</sup> Concealed household is defined as household with more than one family/benefit unit. A benefit unit is defined (following DWP definition) to be a single adult or a married or cohabiting couple and any dependent children; since January 2006 same-sex partners (civil partners and cohabittees) have been included in the same benefit unit.

<sup>22</sup> <http://researchbriefings.files.parliament.uk/documents/SN01164/SN01164.pdf>

Table 9 *Understanding Society* involuntary moves

Note: Unweighted frequencies based on the total UK *Understanding Society* sample (including GPS, BHPS, EMBS and IEMB components)

Reason for move	Number of cross-sectional observations by wave (Cumulative N of unique observations in brackets)						
	w2	w3	w4	w5	w6	w7	w8
Evicted form rented accommodation	31 (31)	43 (74)	53 (124)	47 (171)	40 (211)	43 (253)	61 (312)
Health reasons	38 (38)	53 (91)	34 (122)	12 (134)	42 (176)	37 (213)	20 (232)
Moved to bungalow/accommodation with no stairs	10 (10)	2 (12)	2 (14)	4 (18)	6 (24)	18 (42)	6 (48)
Moved to sheltered accommodation/institution	4 (4)	0 (4)	0 (4)	6 (10)	0 (10)	1 (11)	11 (22)

## **Review of the *Understanding Society* questions related to tracking hard to reach individuals**

### **Leavers**

*Understanding Society* household grid module contains a number of questions related to changes in household composition that can be useful for identifying some of the hard to reach and mobile populations. Frequencies of the main variables related to the sample members who left the household or who are temporarily away from the household address are summarised in Table 10 (Note: Majority of these details are currently not publically available. The *Understanding Society* data team is currently reviewing the possibility of releasing them together with the main data). Not all of the variables presented in Table 10 are available for all waves due to the changes in the questionnaire design between waves. In particular, the changes in the questions about household leavers between Waves 1 and 2 were expected as we had learned a lot about household initial conditions at Wave 1; and; the changes in the questionnaire design in Wave 4 were largely due to the preparation for the move to web. Nevertheless, we should aim to improve consistency of the leavers' questions coding across waves and make sure that we provide careful documentation about all the changes made to the questions about household leavers (including those temporarily away from the household). The numbers presented in Table 10 show that the main reasons why people are leaving the household are related to changes in family circumstances such as: setting up own home, separation/divorce or new cohabitation/marriage/moving with a partner. However, there is a considerable share of household leavers (on average about one fifth of leavers) for whom we have no concrete information about either the reason for leaving or the type of place where they are currently living (see frequencies of lvwhy 'other' category and clstat 'somewhere else' category). These responses are separate from 'refusals' and 'don't knows', which means that they could potentially be coded into more detailed response categories in future waves. Similarly, some of the earlier variables had more detailed coding of institutions where the household leaver went to (e.g. information about a child being in children's care home was coded separately in Wave 1 'absreason' question, and information about household member leaving to old person home was coded separately in 'lvwhy' question in Waves 2 and 3). In future waves, the more detailed coding of the institutions that sample members are currently living in/moving into could be potentially re-introduced. It is important to keep in mind, however, that even if we do not know the reason why someone left or the type of accommodation they moved into, we still attempt to collect information about their new physical address (i.e. the contact attempt will be made at the subsequent wave if the leaver has OSM or PSM status).

### **Joiners**

In each wave, there is considerable number of people joining the Study. The recorded reasons why a new person is reported in the household are summarised in Table 11. The vast majority of new joiners are babies born between waves (there are between 562 and 1,118 newborns reported at each wave). The other reported reasons include: partners moving in with the sample members (between 438 and 721 cases at each wave), people moving into their parents' or relative house (between 323 and 595 cases at each wave), people moving into shared accommodation (between 152 and 314 cases at each wave), and people moving from college or university (between 65 and 137 cases at each wave). Similar to the questions asked about household leavers, there is a considerable share of responses coded as 'other reasons' (between 152 and 314 at each wave), which could potentially be

coded into more detailed categories. Currently, the retrospective information about new joiners is not collected as they are Temporary Sample Members, who have been considered (to date) primarily in terms of the contextual information that they provide in relation to Original Sample Members. However, collecting additional information about their previous place of residence could potentially enable us to identify some of the hard-to-reach groups discussed earlier in this paper (e.g. people moving from prisons, young offenders/children homes, homeless, etc.).

Table 10

Reason why someone left the household (lvwhy)	Number of observations							
	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
deceased		332	244	0	0	0	0	0
separated/divorced		369	300	284	241	207	199	183
to college/university		226	193	103	99	86	60	66
to old persons home		24	29	0	0	0	0	0
sent to prison		4	1	0	0	0	0	0
left for job		258	185	146	108	108	145	104
cohab/moved with partner		324	293	264	240	231	207	211
marriage/civil partnership		75	74	59	40	57	70	56
to set up own home		823	627	554	527	432	452	414
child/dep moved w adult/carer		271	198	196	185	136	137	115
returned home from college or university		0	0	53	36	22	31	31
other		571	441	316	266	242	280	272
Current place where household leaver is living now (clstat)								
deceased				200	172	170	239	154
their own home/flat				1429	1293	1182	1227	1112
working away from home				39	39	51	37	49
halls of residence				78	75	69	62	60
boarding school				2	1	3	0	1
prison				17	4	6	3	2
other institution (such as a nursing home or hospital)				30	31	22	28	29
somewhere else				597	520	364	402	361
Reason member is absent (absreason)								
boarding school		6						
college/university		169						
hospital		4						
care/nursing home		19						
retirement home		1						
local authority children's care home		3						
prison/young offenders institution		6						
other		25						

Note: Unweighted frequencies based on the total UK *Understanding Society* sample (including GPS, BHPS, EMBS and IEMB components)

Table 11

New joiner reason (ynew)	Number of observations						
	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
new baby	1,118	995	862	733	553	683	562
marriage/cohab/partner	721	699	654	566	525	505	438
from college/university	137	96	78	86	65	65	66
from institution	7	5	2	6	3	1	3
resident last wave/never left	282	201	112	91	104	225	146
moved in with parent/relative	595	539	495	442	371	378	323
shared accommodation	425	325	322	267	235	263	155
other	314	274	243	215	198	204	152

Note: Unweighted frequencies based on the total UK *Understanding Society* sample (including GPS, BHPS, EMBS and IEMB components)

### Potential directions for improving coverage of small/transitory hard-to-reach subgroups

One potential direction of improving coverage of small and particularly vulnerable populations would be increasing efforts to interview people who moved into institutions. As indicated in earlier section, the number of interviews conducted with people living in institutions is lower than what we would expect based on the number of household addresses located in institutions. As part of the recently funded Wave 12 grant, we are planning to begin a specific 'leavers survey' for people moving into institutions either fielded to the person themselves or to a relative if this is not possible. This would be a first step to considering whether more interviews with sample members living within institutions are feasible.

In relation to the questionnaire content, the 'leavers' and 'joiners' questions could potentially be more tailored to track small but particularly interesting and potentially underrepresented groups. One improvement could simply be recording more details about the type of accommodation that the mover has moved into and the circumstances of the move (e.g. what type of institution, who they moved with, the reasons of move (if known)). The current plans (which part of Wave 12 grant) include conducting web-interviews with people who emigrate outside of the UK. Similar strategy could be considered for obtaining interviews with those who are currently in prisons. The questionnaire could also include more retrospective questions asked of new joiners, in particular, questions about the type of accommodation that someone was living in before joining the household, the number of previous home moves, and family circumstances. Such questions would allow us to obtain valuable information about potentially the most mobile part of the population that we have in the *Understanding Society* sample. As part of current bid for waves 13-15, if funded, we also plan to improve tracking important life events and introduce special event-triggered modules. The new strategy will involve sending out text messages every month asking respondents whether they have recently been: 1. Diagnosed with a new health condition or entered hospital as an in-/outpatient; 2. Became pregnant / partner became pregnant; 3. Changed jobs, started or stopped working; 4. Moved house; 5. Stopped or started living with a partner; and, if they answer 'yes' to any of the above, they will be asked to complete a

special module related to the event that happened to them. This strategy should both improve our ability to track people who experience important life changes as well as allow us to better understand the impact of these changes on people's lives.

## References

Abbott, O. and G. Compton, 2014. "Counting and estimating hard-to-survey populations in the 2011 Census" in *Hard-to-Survey Populations* by Tourangeau, R. B. Edwards, T. P. Johnson, K. M. Wolter, Cambridge University Press, retrieved from: <https://www.cambridge.org/core/books/hardtosurvey-populations/counting-and-estimating-hardtosurvey-populations-in-the-2011-census/7C414EE6710D70F9AAD135EFB0B5A354>

Kruskal, W. and Mosteller, F., 1979. Representative sampling, I: Non-scientific literature. *International Statistical Review/Revue Internationale de Statistique*, pp.13-24.

Lynn, P., and M. Borkowska, 2018. "Some Indicators of Sample Representativeness and Attrition Bias for BHPS and *Understanding Society*", *Understanding Society Working Paper Series*, No. 2018-01

Lynn P., J. Burton, O. Kaminska, G. Knies and A. Nandi, 2012. "An Initial Look at Non-Response and Attrition in *Understanding Society*" *Understanding Society Working Paper Series*, No. 2012 – 02

ODI 2015. "Exclusion in household surveys: Causes, impacts and ways forward", by C. Chávez Villegas and E. Samman

ONS 2009. "Predicting Patterns of Household Non-Response"  
<https://www.ons.gov.uk/file?uri=/census/2011census/consultationsusersandlocalpartners/censusadvisorygroups/censusgeneralcag/ag0917tcm77190524.pdf>

## Appendix

Table A1: Children in care in the UK: Stock and flows based on combined administrative estimates

Year	Stock % of the population of children (0-17yo) living in hh	Estimated total number of children looked after	Flow in % of the population of children (0-17yo) living in hh	Flow out % of the population of children (0-17yo) living in hh
2010	0.68	88,128	0.28	0.24
2011	0.69	89,652	0.27	0.25
2012	0.70	91,682	0.28	0.25
2013	0.71	92,664	0.28	0.28
2014	0.71	93,013	0.29	0.29
2015	0.70	93,360	0.29	0.29
2016	0.71	94,272	0.29	0.29
2017	0.72	96,415	0.30	0.28

Note: Administrative estimates based on combined official data for constituent UK countries obtained from: Department of Education (England), Department of Health (Northern Ireland), Children's Social Work Statistics (Scotland) and Welsh Government (Wales). For each year, the share of children in care is calculated in relation to mid-year population estimates of children aged 0-17. We assume here that the proportion of children aged 0-17 living in households each year is approximately the same as the proportion of children aged 0-17 reported in the 2011 censuses.

Table A2: People living in care homes: stock (available for the UK) and flow (available for Scotland) of those aged 65 and over

Year	UK (stock)		Scotland (flows)	
	% of people aged 65+	Estimated number of people accessing long-term support	Flows in % of people aged 65+	Flows out % of people aged 65+
2008-2009	3.1	314,299	1.5	0.4
2009-2010	3.0	303,361	1.6	0.4
2010-2011	2.9	300,666	1.5	0.4
2011-2012	2.8	299,700	1.4	0.3
2012-2013	2.7	296,510	1.4	0.3
2013-2014	2.6	291,706	1.3	0.3
2014-2015	2.5	285,334	1.3	0.3
2015-2016	2.4	281,186	1.4	0.2

Note: Stock estimates based on combined official data for constituent UK countries obtained from: NHS (England), Social Care Wales (Wales), Scottish Government (Scotland) and Department of Health (Northern Ireland). Estimates of flows based on the Scottish Government statistics. The base for each year is the population of people aged 65+ obtained from the mid-year population estimates.

Table A3: Prisoners stock (UK) and flows (England and Wales)

UK Stock			England and Wales Flows	
Year	Number	% of UK population aged 18+	First prison reception - % of the E&W population aged 18+	Releases- % of the E&W population aged 18+
2010	94,044	0.19	0.27	0.21
2011	95,812	0.19	0.27	0.19
2012	96,465	0.19	0.25	0.19
2013	93,970	0.19	0.24	0.18
2014	94,868	0.19	0.22	0.16
2015	94,893	0.18	0.21	0.16
2016	94,372	0.18	0.19	0.16
2017	n.a.*	n.a.	0.18	0.15

Note: Stock estimates based on combined official data for constituent UK countries obtained from: Ministry of Justice Offender Management Statistics Quarter (England and Wales), Scottish Government 'Prison Statistics and population projections', and Department of Justice (Northern Ireland). Flows estimates calculated based on the Ministry of Justice data for England and Wales. The base for each year is the population of people aged 18+ obtained from the mid-year population estimates. \*n.a.-data not available

Table A4: People living in temporary accommodation (stock)

Year	England		Total number of children in TA	Wales		Scotland	
	Total number of households in TA	Total number of HH in TA % of total number of HH*		Total number of families accommodated	Total number of HH in TA % of total number of HH	Total number of households in TA	Total number of HH in TA % of total number of HH
2004	101,030	0.48	124,630	n.a.**	n.a.	n.a.	n.a.
2005	98,730	0.47	127,620	n.a.	n.a.	n.a.	n.a.
2006	89,510	0.42	122,080	n.a.	n.a.	n.a.	n.a.
2007	79,500	0.37	112,260	n.a.	n.a.	n.a.	n.a.
2008	67,480	0.31	98,880	n.a.	n.a.	n.a.	n.a.
2009	53,370	0.25	77,990	n.a.	n.a.	n.a.	n.a.
2010	48,010	0.22	69,050	n.a.	n.a.	n.a.	n.a.
2011	48,920	0.22	69,460	n.a.	n.a.	n.a.	n.a.
2012	53,140	0.24	76,740	n.a.	n.a.	n.a.	n.a.
2013	56,940	0.26	80,970	n.a.	n.a.	n.a.	n.a.
2014	61,930	0.28	93,980	n.a.	n.a.	n.a.	n.a.
2015	69,140	0.31	106,240	1,779	0.13	n.a.	n.a.
2016	75,740	0.34	118,930	1,953	0.14	n.a.	n.a.
2017	79,720	n.a.	122,400	1,971	n.a.	43,531	1.80

Note: Estimates calculated based on official statistics obtained from: Ministry of Housing and Local Government (for England), Scottish Government (for Scotland), and Welsh Government (for Wales). \*Base: Total number of households in England by year obtained from the mid-year population estimates. \*\*n.a.-data not available