An assessment of the national representativeness of new mothers and women of childbearing age in Understanding Society

Paul Fisher (University of Essex)

Non-technical summary

Understanding Society is an excellent data source for addressing questions around pregnancy and child development. It collects annual data on the mother-to-be and possibly the father and other siblings pre-birth, possibly over a significant time period. As the child grows up, ongoing data collection is conducted covering multiple topics, enabling novel research on how different aspects of the lives of the child and family interact with each other over time. The ethnic minority boost samples and coverage of the whole UK means the Study can inform science and policy across the development of the child, while examining differences in different parts of the UK or for different ethnic and social groups. This report demonstrated that:

- The *Understanding Society* sample sizes of mothers with newborns and women of childbearing age are large, allowing for common types of analysis on small subgroups of interest such as the most disadvantaged mothers. For example, from Waves 1 to 8, the pooled sample of single mothers is over 15,000, it is nearly 6,000 for women with newborns and over 5,000 are poor single mothers. Even when focusing on the number of unique cases the sample sizes are still substantial.
- Second, the *Understanding Society* samples of mothers with newborns and women of childbearing age line up closely to their *Family Resource Survey* (FRS) counterparts in terms of their demographic profile and measures of deprivation (poverty rates, share single mothers, living in low-cost housing) even as the panel ages.
- Third, *Understanding Society* opens up longitudinal research possibilities that are simply not possible with cross-sectional data such as the FRS. For example, women with newborns experience higher persistent poverty rates than the overall sample of women of childbearing age. Second, a substantial share of both mothers with newborns and women of childbearing age live in poverty for multiple years and this trend has been stable since the financial crisis.

In sum, the longitudinal dimension of *Understanding Society* makes it uniquely placed to enrich our understanding of the experience of parents and their children. We identified a number of actions for the Study to improve identification of pregnancies and new babies.

An assessment of the national representativeness of new mothers and women of childbearing age in Understanding Society

Paul Fisher (University of Essex)

Abstract: The aims of this report are threefold. The first is to document the representativeness of Understanding Society subgroups relevant for pregnancy and child development; and show how the representativeness has evolved over time. It does this by comparing estimates from Understanding Society to those from a high-quality cross-sectional survey – the Family Resources Survey (FRS). Our comparisons are cross-sectional in nature. This is relevant as many uses of Understanding Society are cross-sectional. Moreover, longitudinal measures such as transitions, are calculated as the difference between two cross-sections, and so it is important to get the latter right. Second, we document the available Understanding Society sample sizes of relevant subgroups. These are women of childbearing age and women with newborns, which we further subset according to measures of disadvantage. Third, we demonstrate the unique value of Understanding Society for research in this area. We report on the persistent poverty rates of our subgroups of women, something which has not been previously done for the UK, and is not possible without longitudinal data.

Keywords: representativeness, childbearing, mothers, pregnancy, child development

Corresponding author: Paul Fisher, Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, pfishe@essex.ac.uk

1. Introduction

There are substantial research agendas around pregnancy and child development. Understanding Society is an excellent data source for addressing questions in these fields. Central to its appropriateness is that it is a large probability sample and representative of the UK population and its subgroups, and that the available subgroup sample sizes are sufficiently large for analysis purposes. Additionally, unlike birth cohorts that rarely have information pre-conception, or on other family members, Understanding Society will have been collecting annual data on the mother and possibly the father and other siblings for some time. This means that information on family circumstances for parents pre-birth and the child around birth is collected at the time, rather than retrospectively. As the child grows up, ongoing data collection is conducted within the main study covering multiple topics, enabling novel intra generational analyses as well as research on siblings on how different aspects of the lives of the child and family interact with each other over time. The ethnic minority boost samples in Understanding Society and coverage of the whole UK means the Study can inform science and policy across the development of the child, while examining differences in different parts of the UK or for different ethnic and social groups.

Given the wealth of research opportunities better data collection around pregnancy create for science and policy, it is important to confirm that the survey remains representative of these key groups in practise, and not just in principle. The aim of this report is to therefore update our quality checks on the representativeness and adequacy of Understanding Society for the purposes of research around pregnancy and children.

Users of Understanding Society data on parents and their children are addressing questions on a vast number of different topics such as: the dynamics of living standards, fertility timing, child development, intra-family relations, pregnancy, the dynamics of parent and child health, labour market behaviour of families, and partnering behaviour to name just a few. One thing these agendas have in common is that they often make inferences about the whole population. For example, a reoccurring theme is a focus on inequalities, e.g. how are children of poor mothers developing compared to those of rich ones. To empirically study inequalities, representative data is needed that covers all groups of the population, including the poorest. Related, unrepresentativeness may make analysis impossible if it leads to too small samples of groups of interest, such as the most disadvantaged groups. As data producers, it is important for us to periodically document the representativeness of Understanding Society mothers and mothers to be, so that researchers can remain confident in the ability of the data to answer the key research questions in this area.

A key feature of the Understanding Society design is that the initial sample is large and representative of the whole UK population and so supports direct inferences about the whole

population and subgroups of interest e.g. low income pregnant women[1]. However, non-response poses a threat to representativeness and the available sample sizes. Non-response could be initial non-response either at the household or individual level or subsequent attrition. Like most household surveys, Understanding Society uses weighting to make samples representative. A carefully designed set of weights are available in the public release data. Even so, as with any survey, a concern remains that weighting may not adequately correct for non-response, particularly if non-response causes missing data patterns that cannot be predicted from observable variables available to the data producer. Moreover, even for an adequately weighted sample, non-response may still threaten the sample sizes available for analysis.

Existing quality assessments confirm that, once the sample design is accounted for, the initial Understanding Society sample looks reassuringly similar to the census population, where both are designed to be representative of the UK population (Lynn and Borkowska (2011)). Lynn and Borkowska (2011) find only a suggestion that men, those residing in Greater London, and people with a severely limiting long-term illness are marginally under-represented. Their results on attrition are also reassuring. The effect of attrition over six waves leads to only minor under-representation of some groups. These are: the youngest age groups, men, non-whites, residents of Greater London and those on the lowest incomes. While the focus of the present report is on women and children – and not the full Understanding Society population – the above results inform our choice of demographics that receive our focus in this report.

The aims of this report are threefold. The first is to document the representativeness of Understanding Society subgroups relevant for pregnancy and child development; and show how the representativeness has evolved over time. It does this by comparing estimates from Understanding Society to those from a high quality cross-sectional survey – the Family Resources Survey (FRS). Our comparisons are cross-sectional in nature. This is relevant as many uses of Understanding Society are cross-sectional. Moreover, longitudinal measures such as transitions, are calculated as the difference between two cross-sections, and so it is important to get the latter right. Second, we document the available Understanding Society sample sizes of relevant subgroups. These are women of childbearing age and women with newborns, which we further subset according to measures of disadvantage. Third, we demonstrate the unique value of Understanding Society for research in this area. We report on the persistent poverty rates of our subgroups of women, something which has not been previously done for the UK, and is not possible without longitudinal data.

The report concludes with three main findings. First, the large Understanding Society sample delivers substantial subsamples of women of childbearing age and women with newborns. Even when subsetting these groups further, to the most disadvantaged, the sample sizes are adequate for common types of analysis performed by users. Second, women with newborns and women of

---

[1] An exception is new immigrants arriving since the start of the panel. A boost sample of this group was included at Wave 6 in the "Immigrant and Ethnic Minority Boost" sample.

childbearing age in Understanding Society line up closely to those in the FRS in terms of their demographic profile and we find little evidence that they deviate from the FRS overtime. We therefore conclude that Understanding Society started out and remains largely representative of the groups of interest. Third, Understanding Society is uniquely placed to address longitudinal research questions on women and their children.

The report proceeds as follows: the next section describes the data and methods. We then present our results in section three. The first subsection reports on the available Understanding Society sample sizes, a second subsection reports on the representativeness of Understanding Society women with newborns and a third subsection reports on the representativeness of women of childbearing age. Section four then shows our longitudinal analysis on persistent poverty and finally section five concludes.

2. Data and methods

In this section, we first give an overview of the Understanding Society data emphasising the relevant features of the sample design and set out the sample restrictions we impose. We then motivate our choice of comparator dataset – the Family Resources Survey – and give an overview of it. We describe its key features and point out where it differs from Understanding Society. Finally, we describe our methods.

i) Understanding Society

Understanding Society is a large household panel survey that began in 2009. As the study facilitates broad types of research, the sample design is somewhat complicated and consists of multiple subsamples. At Wave 1, samples were drawn to be representative of the Great Britain and Northern Ireland populations (GB and NI samples), and a further boost sample was drawn to enable research on ethnic minorities (EMB). To address attrition of ethnic minorities, and to capture new immigrants since Wave 1, a further boost sample was introduced at Wave 6 (IEMB). Further, the study incorporates the former British Household Panel Survey, for which 25 waves of data are now available. At the time of writing, eight Understanding Society waves of data (2009-17) have been publicly released and Waves 10 and 11 are in the field.

All of the initial Understanding Society respondents are interviewed annually as well as any other individuals they may be residing with at the time of their interview. Births to the original sample members also become permanent sample members and are followed over time. These following rules aim to mimic the demographic processes by which the population is reproduced and so Understanding Society should remain representative of the UK population over time, with the exception of new immigrants. This is important for our validations, as the Understanding Society population represents the same population as our comparator data set, once we account for new immigrants. We provide details of how we address new immigrants in the next subsection.

The large Understanding Society sample requires the fieldwork for each wave to be spread over 24 months. Note that this differs from our comparator data set, the FRS, where fieldwork occurs on a financial year basis (April-March). To make the reference periods comparable, we construct Understanding Society financial year samples by combining data from multiple waves i.e. months 4-15 from Wave n are combined with months 16-24 from Wave n-1 and months 1-3 from Wave n+1[2]. We include all available Understanding Society waves in our main analysis but note we must exclude data collected before April 2010 (Wave 1) and after March 2015 (Wave 8) where it is not possible to construct full financial year estimates. Our main validations therefore correspond to the financial

---

[2] Note, this requires pooling weights across waves and an appropriate rescaling of the weights so that later waves are not under-represented. We finally rescale each financial year weight to a mean of one as we do for the FRS below.

years 2010-15. Note, we exclude the IEMB subsample introduced at Wave 6 to keep a consistent treatment of new immigrants across all years of our analysis.

Note, inclusion into our main analysis sample is determined by having a non-zero value for the relevant Understanding Society cross-sectional weight. The publicly available weights give a positive value to original, temporary and permanent sample members[3].

As much research on mothers/babies focuses on inequalities, and the poorest are most likely to attrite, we have a particular interest in the Understanding Society coverage of the poorest in these groups. As Understanding Society includes a detailed set of individual income questions and all members of a household are, in principle, interviewed, it allows for the derivation of household net income variables. Indeed, the Understanding Society household net income variables, available in the public release data, have attempted to replicate those in our comparator dataset and are described in the next subsection.

As an aside to our main validations, we note that Understanding Society, collects data on a wide range of topics relevant to research on parents and their children. A wide range of data is collected on family life, both before, during and after pregnancy; family and household composition; income; employment; health; education; transport and neighbourhoods; civic engagement; attitudes; and behaviours. Specifically on `fertility, pregnancy and early life', we ask couples about fertility intentions; and once a woman is pregnant about conception, behaviours in pregnancy, child birth experience, birthweight and delivery, breastfeeding and sleeping. Questions are also asked on a number of validated child development scales for children under 1, ages 3, 5 and 8. In our Wave 12 grant, we included a child development post to lead on improving this area of the Study's data collection.

ii) FRS

Our comparator data set is the Family Resources Survey (FRS), a specialist income survey fielded annually since 1992. It is a household survey and collects detailed information on living standards including: income, housing tenure, caring needs and responsibilities and disability. It is conducted by the Office for National Statistics and the National Centre for Social Research, on behalf of the Department for Work and Pensions and is designated by the UK Statistics Authority as National Statistics. To match the period of our Understanding Society, we use data from the 2010-15 financial years. FRS is a suitable comparator data set for three main reasons.

---

[3] By design, the weights take a value of zero for some individuals. The Understanding Society weighting strategy is currently being revised by the data producers and in the revised cross-sectional weights the share of respondents with positive weights will increase. In the future, this will increase the weighted samples available for analysis. A given analysis will still represent the UK population, but the larger samples will imply greater statistical precision.

First, it is designed to be representative of the UK population. Therefore, with the exception of new immigrants, it represents the same population as Understanding Society in any given year. Differences in the timing of data collection (Understanding Society two-year wave vs. financial year in FRS) slightly complicate the matter further as discussed above.

Second, it includes reliable measures of interest, equivalent to our Understanding Society measures. These are: whether a woman has a new born, age of mother, household composition variables, housing tenure, region, and measures of household economic position[4]. The measures of household economic position (in particular household net income) are particularly relevant to our analysis and they are considered of high quality. We discuss these further below.

Third, it is large survey with around 20,000 households at the latest release. This means that our comparisons can be extended to analysis of subgroups of interest e.g. poor mothers.

The key differences with Understanding Society are that the FRS fieldwork period is on a financial year basis (rather than over two calendar years) and that new immigrants are covered by the FRS but not Understanding Society. To make the two surveys comparable, we remove new immigrants since 2009 from the FRS sample and, as set out above, construct Understanding Society samples on a financial year (rather than wave) basis.

In all of our analysis we focus on individuals aged 16+ and use the FRS weights to scale our results to the UK population.

Our main measure of economic position is based on household net income. Our FRS based household net income measure is the one used in the official publication on the UK income distribution ("Households Below Average Income" series). It has been carefully derived by the DWP who put extensive resources into the process. This includes extensive editing of the raw survey reports – based on DWP knowledge of the benefits system and their access to administrative records – and imputation of missing data. As the Understanding Society household net income measure aims to replicate the HBAI one, the two datasets are closely comparable in this dimension. In order to compare incomes for households of different sizes and compositions, we equivalise our income measure using the OECD-modified scale. To make living standards at different points in time comparable, all figures are expressed in 2016 prices using the bespoke monthly CPI price index used in the official UK income statistics and produced by the Office for National Statistics.

iii) Methods and variable definitions

---

[4] The final derived income variables are included in a separate dataset known as the HBAI datasets on which this report is based.

To compare means between the Understanding Society and FRS we perform standard t-tests under the assumption of constant variance. We appropriately weight each sample and weights for each survey are scaled to a mean of one. In calculating standard errors, we account for the stratification and clustering in the Understanding Society sample design[5].

The focus of this report is on women and pregnancy. We therefore present results separately for women with newborns and women of childbearing age. So that we have a consistent definition in both surveys, we define women with newborns as women with a child who is less than age one. This means that we exclude a handful of newborns entering Understanding Society who are already age one when we first observe them in an interview. We also exclude Understanding Society women in responding households but who did not provide a full interview (non-respondents and proxy interview cases), on the basis that many types of analysis are not possible without the content from an adult interview. Of course our weighting strategy takes account of this selection. Women of childbearing age are defined as women between the ages of 15 and 45. When comparing demographic variables across the two surveys, we pool across all financial years. For our measures of disadvantage, we present results on a financial year basis, in order to understand how the representativeness of Understanding Society may have changed overtime in this central dimension.

When presenting poverty rates we apply the standard UK Government definition, which is residing in a household where equivalised income is less than 60 percent of the median household income.

---

[5] We ignore the complex sample design of the Family Resource Survey as: i) this requires access to the resample datasets which we do not have and ii) the tables in Brewer et. al (2017) confirm that accounting for it has only a minor effect on estimated standard errors.

3. Results

This section presents the results. First, we begin by showing that the large Understanding Society sample size delivers large subsamples of new mothers and women of childbearing age and, moreover, it is sufficiently large to enable further subgroup analysis of these groups. The next two subsections then present our validations, first for new mothers and then for women of childbearing age. They show that both groups look very comparable in socio-economic characteristics to their FRS counterparts. We finish with an illustrative example, in the final subsection, that demonstrates the power of Understanding Society in understanding the evolution of living-standards of new mothers. We perform a longitudinal analysis that documents trends in persistent poverty of new mothers, something not possible with cross-sectional data such as the FRS.

i) Sample sizes available for analysis

Understanding the socio-economic performance of disadvantaged mothers and their children forms an important social science research agenda. At the same time, there is a concern that these groups might be the hardest to reach in social-science surveys and, for panel surveys like Understanding Society, be the most likely to drop-out. One possible effect is that there will be inadequate sample sizes with which an interested researcher can perform analysis. In this subsection we tackle this concern head-on. We demonstrate the feasibility of Understanding Society as a basis for research on mothers and their children, where the focus is on the available Understanding Society sample sizes. We show that the large Understanding Society sample delivers large subsamples of mothers and women of childbearing age, moreover, when we look at the most disadvantaged in each group, Understanding Society provides adequate sample sizes for common types of analysis.

Unlike with cross-sectional data, there is no one definition of sample size in the longitudinal setting. A given sample of individuals will deliver analysis samples of different sizes, depending on the research question and how the researcher needs to treat the repeated observations on the same individual. Therefore, in what follows, we restrict our focus to three definitions of sample size that we think are relevant for the majority of analysis types. The first we call "total analysis sample" which counts repeated observations on the same person independently and so the sample is straightforwardly the sum of the number of units per wave (Waves 1-8). This would be relevant, say, if a researcher wanted to analyse a large sample of new mothers cross-sectionally and was willing to pool new mothers from all waves. Second, we break down the "total analysis sample" by wave. This number would be relevant if a researcher wanted to calculate trends over-time, say the trend in poverty of new mothers, or if they wanted to restrict analysis to a particular set of time periods (waves). Third, we report the number of unique persons available for analysis over all eight waves. This would be relevant, say, for studying the living standards of new mothers following child-birth e.g. when calculating the number of waves that a given new mother spent in poverty and each mother contributes exactly one observation to the analysis sample. Of course, in many types of longitudinal analysis, each person may contribute more than one observation e.g. if pooling wave-to-

wave transitions. Therefore, the numbers we report are indicative of the sample sizes available for broad types of longitudinal analysis but would have to be tailored to fit the exact research question understudy.

Table 1 focuses on sample sizes for two groups of substantial research and policy interest: single mothers and women with newborns (age<1). Single parent rates are high in the UK and those with young children are central in various literatures including on early years interventions, and the health of newborns and their mothers. We see in the table substantial sample sizes for each group. A total analysis sample of 15,699 single mothers is available and even for the smaller group of mothers with newborns the sample size is a sizeable 5,810. These are notably more than the equivalents from the FRS, for example, which is considered to be a large survey (equivalent FRS sample sizes are 13,724 and 3,737, respectively).

Understanding Society with its detailed measures of living standards is well suited to performing analysis of the most disadvantaged groups, for example, those that are poor. Columns three and four of the table further subset our two groups of Mothers to look at the available sample sizes of the poorest among them. For single mothers who are poor we see a sample size of 5,290 and for mothers with newborns who are poor we see a sample size of 1,399. These samples sizes can still be considered fairly large and open up the possibility for subgroups analysis that would not have been possible in the former BHPS, say. The final column of the table shows that even when focusing on single mothers with newborns, a much smaller group in the population, there is still a reasonable analysis sample of 931.

Table 1: Understanding Society sample sizes Waves 1-8: Mothers

	single		with newborns		single and poor		with newborns and poor		single and with newborns	
Total analysis sample:	15,699	(14,278)	5,810	(5,242)	5,290	(4,767)	1,399	(1,257)	931	(828)
By wave:										
1	2,754	(2,727)	1,116	(1,102)	1,060	(1,049)	287	(286)	227	(224)
2	2,483	(2,446)	886	(866)	829	(819)	204	(199)	117	(116)
3	2,173	(1,982)	845	(748)	706	(635)	201	(170)	144	(124)
4	1,925	(1,688)	728	(621)	600	(513)	170	(143)	128	(106)
5	1,719	(1,453)	655	(556)	529	(428)	162	(127)	94	(76)
6	1,773	(1,483)	601	(507)	608	(498)	129	(118)	81	(67)
7	1,542	(1,233)	545	(425)	504	(395)	126	(100)	73	(54)
8	1,330	(1,266)	434	(417)	454	(430)	120	(114)	67	(61)
Unique cases:	4,879	(4,681)	4445	(4,047)	2,564	(2,392)	1204	(1,091)	795	(720)

Notes: Total analysis sample pools all available waves (Waves 1-8) where an adult respondent completed a full interview. Sample sizes for our analysis samples as described in section 2 (i.e. financial years 2010-15) are in parenthesis.

A large stream of research around mothers and their children relates to the age of mothers. To give just two examples, there is a large interest in teen pregnancy and the socio-economic outcomes of such teens and their children; and in understanding fertility patterns of older women in a context of delayed childbearing. Figure 1 explores possibilities for this type of research by presenting the pooled sample sizes for six different age groups for our single mothers and women with newborns. The figure shows the sample sizes separately for ages: 16-20; 20-24; 25-29; 30-34; 35-39; and 40+. Each of the groups provides adequate samples for analysis and some of them are particularly large. For example, there are a considerable 7,218 single mothers over age 40 and 1899 women with newborns age 30-34. The group of teens is the smallest for both subgroups, but even so, there are sufficient observations to provide pooled analysis of these smaller groups with 227 single mothers and 165 women with newborns.

Note, the sample sizes quoted above exclude teen mothers who are less than age 16. While we can identify such cases in our data, we choose to exclude them, as only those aged at least 16 provide data in an adult interview in Understanding Society. Second, the group of such people is very small

indeed – according to the Office for National Statistics, in England and Wales, there were only 3.8 conceptions per thousand women age 13-15 in 2015 and 60% lead to an abortion (ONS 2017).

Table 2 extends the results to the larger group of all women of childbearing age. The sample sizes are substantial. There are 93,120 women of childbearing age and 18,103 of them are poor. For example, the FRS equivalents are 50,812 and 8,803. When looking at the unique cases, we see that there are still substantial sample sizes with 25,205 women of childbearing age and 9,188 who are of childbearing age and poor.
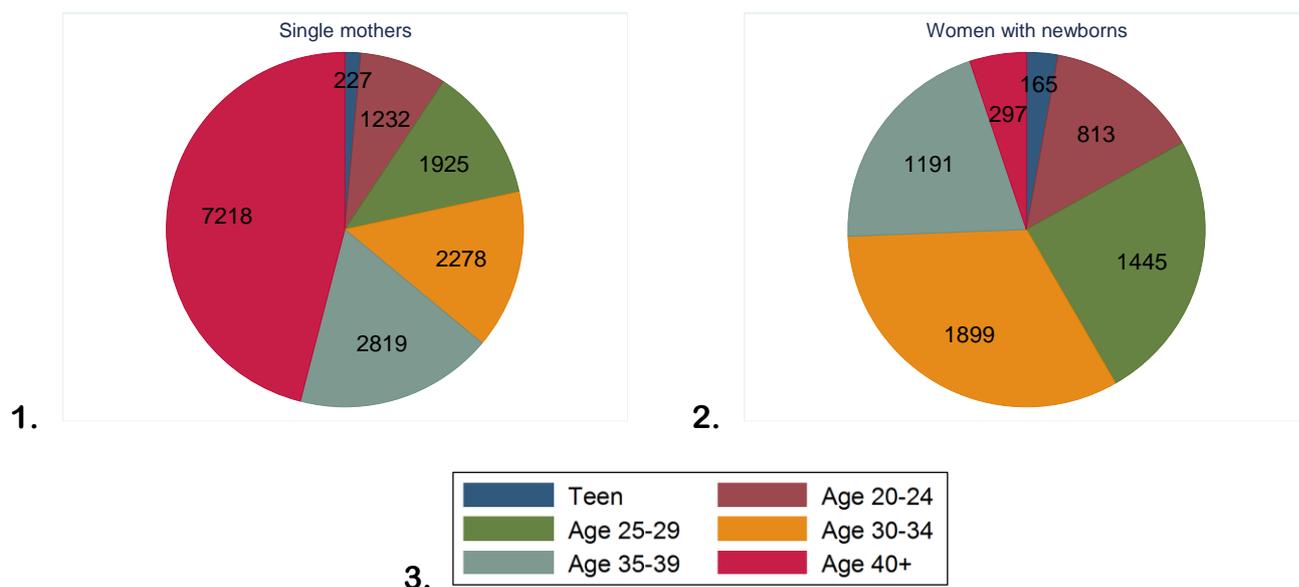
Figure 1: Pooled analysis sample by age

1. Single mothers
2. Women with newborns
3. Legend: Teen, Age 20-24, Age 25-29, Age 30-34, Age 35-39, Age 40+

Table 2: Understanding Society sample sizes Waves 1-8: Women of childbearing age

	Childbearing age		Childbearing age and poor	
Total analysis sample	93,120	(85,391)	18,103	(16,387)
By wave:				
1	14,151	(14,044)	3,356	(3,326)
2	14,196	(13,945)	2,683	(2,644)
3	12,613	(11,640)	2,330	(2,106)
4	11,560	(10,240)	2,105	(1,811)
5	10,671	(9,173)	1,868	(1,527)
6	10,939	(9,397)	2,125	(1,810)
7	9,960	(8,267)	1,877	(1,489)
8	9,030	(8,685)	1,759	(1,674)
Unique cases:	25,205	(24,421)	9,188	(8,601)

Notes: Total analysis sample pools all available waves (Waves 1-8) where an adult respondent completed a full interview. Sample sizes for our analysis samples as described in section 2 (i.e. financial years 2010-15) are in parenthesis.

While this section has demonstrated that sample size limitations in Understanding Society pose little threat to common types of analysis, it has not addressed the issue of representativeness. That is the topic of the next two subsections.

ii) Representativeness: Women with newborns

This section presents our Understanding Society vs. FRS comparisons for women with newborns. Our aim is to examine whether Understanding Society women with newborns are representative of those in the wider UK population. We do this by documenting how they differ from the comparable FRS group. We first present comparisons pooling all our financial year subsamples. We show that the two surveys give a remarkably similar picture of the broad characteristics of new mothers, which we interpret as showing that Understanding Society is representative. We then focus on the living standards of new mothers and examine how well Understanding Society tracks the FRS over time in this dimension. The surveys could deviate if Understanding Society became less representative over time. At Wave 1, Understanding Society mothers look very similar to the FRS ones in terms of age, household size, marital status, ethnicity and region. Further, we find little evidence of difference or divergence in their living standards over time. We conclude that Understanding Society starts out and remains representative and is thus a valuable data source for research on new mothers and babies.

Table 3 presents a descriptive comparison of Understanding Society and FRS women with newborns. For a set of demographics collected in both surveys the table reports the weighted means for each survey (columns 1 and 2), the difference in means and whether the difference is statistically significant (column 3). The final three columns present the standard errors (SE) and the respective sample sizes (N_ UKHLS and N_FRS).

The majority of the differences are not statistically significant and when differences do occur, they are small in magnitude. There is no statistical difference in the share who are teens, married, mixed ethnic group, black or from regions excluding Norther Ireland and London. We see small differences for age (30.31 vs 29.8); household size (4.07 vs. 3.81), Asian (0.067 vs. 0.083), other ethnic groups (0.011 vs 0.027) and from residing in Northern Ireland (0.029 vs. 0.036). Slightly larger differences, but still small in practical terms, are seen in the share white (0.88 vs. 0.84), and the share residing in London (0.12 vs. 0.16). Overall, the women from the two surveys look remarkably like each other in terms of their demographic profile.

Figure 2 shows how the Understanding Society poverty rate compares to the FRS ones by financial year. The top panel shows the poverty rates and the lower panel the difference with confidence intervals. We see that mothers with a newborn experience very similar poverty rates according to both surveys. For example, in 2010 we see a poverty rate of 22.3 in Understanding Society and 21.4 in FRS. Overtime, the difference in means between the two surveys fluctuates around zero and is never statistically significant. It is the FRS series that appears the more noisy of the two, and this

reflects the advantage that Understanding Society has in terms of sample size. Put together, Understanding Society mothers look very much like the FRS ones in terms of their experience of poverty.

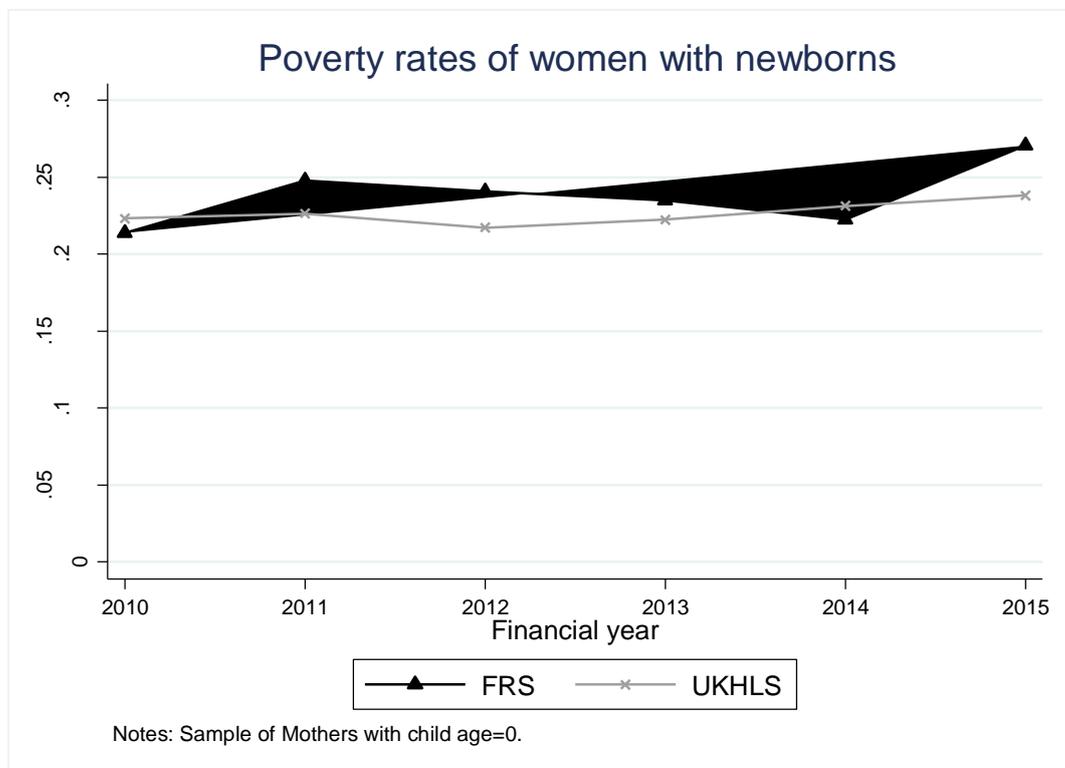**Table 3: Comparison of descriptive statistics for women with newborns**

	UKHLS	FRS	Mean Diff	SE	N_ UKHLS	N_FRS
Age	30.31	29.80	0.512**	0.172	3708	3726
Teen	0.0275	0.0357	-0.0082	0.0049	3708	3726
Household size	4.065	3.805	0.260***	0.0341	3708	3726
Married	0.537	0.539	-0.0019	0.0150	3708	3726
Ethnicity:						
White	0.875	0.836	0.0393***	0.0099	3619	3720
Mixed ethnic group	0.0159	0.0167	-0.0008	0.0035	3619	3720
Asian	0.0670	0.0831	-0.0161*	0.0075	3619	3720
Black	0.0309	0.0374	-0.0065	0.0048	3619	3720
Other	0.0114	0.0271	-0.0158***	0.0041	3619	3720
Region:						
North East	0.0365	0.0371	-0.0006	0.0058	3704	3726
North West	0.113	0.109	0.0042	0.0103	3704	3726
Yorkshire and the Humber	0.0959	0.0899	0.0060	0.0089	3704	3726
East Midlands	0.0765	0.0699	0.0066	0.0083	3704	3726
West Midlands	0.0882	0.0892	-0.0009	0.0088	3704	3726
East of England	0.0946	0.0816	0.0130	0.0093	3704	3726
London	0.122	0.162	-0.0391***	0.0108	3704	3726
South East	0.145	0.138	0.0071	0.0105	3704	3726
South West	0.0760	0.0724	0.0035	0.0086	3704	3726
Wales	0.0500	0.0392	0.0108	0.0058	3704	3726
Scotland	0.0726	0.0760	-0.0035	0.0075	3704	3726
Northern Ireland	0.0290	0.0361	-0.0072**	0.0026	3704	3726

Notes: Sample of Mothers with children age=0. Pools weighted data for financial years 2010-2015.
* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Figure 2: Poverty rates of women with newborns

4.   Figure 2A

Poverty rates of women with newborns

Notes: Sample of Mothers with child age=0.

5.

6.   Figure 2B

Difference in estimated poverty rates (UKHLS-FRS)

Notes: Coefficients from a regression of poverty indicator on being in the UKHLS sample
and a constant. Sample sizes are: 1800; 1400; 1253; 1154; 964; 863.
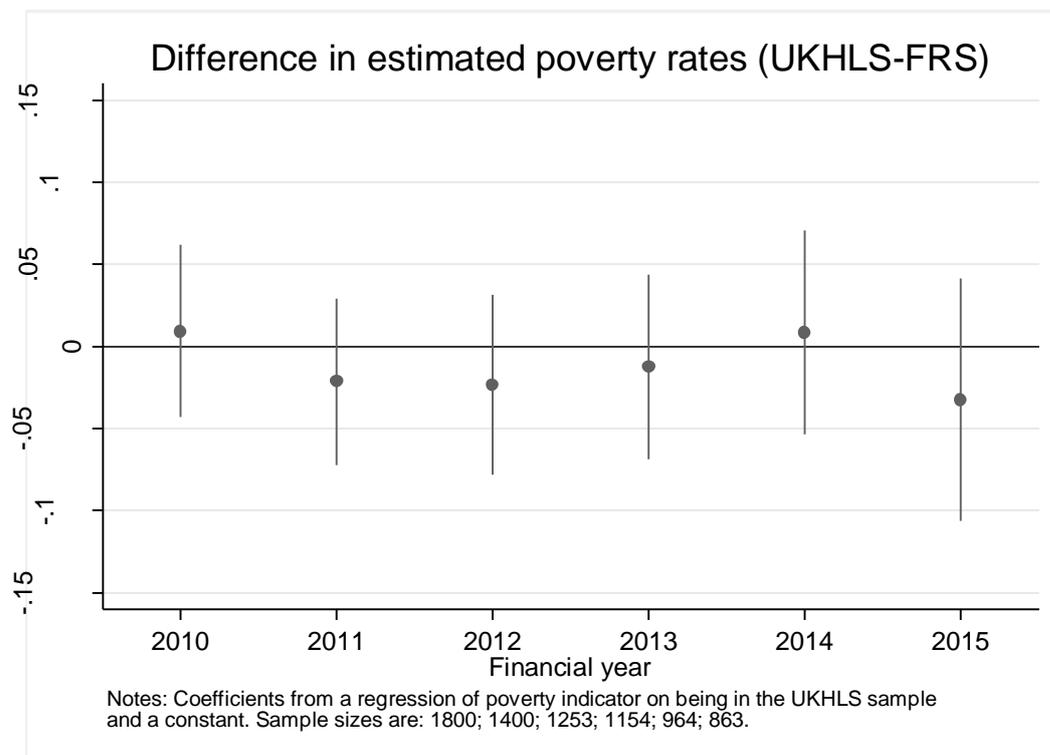
7.

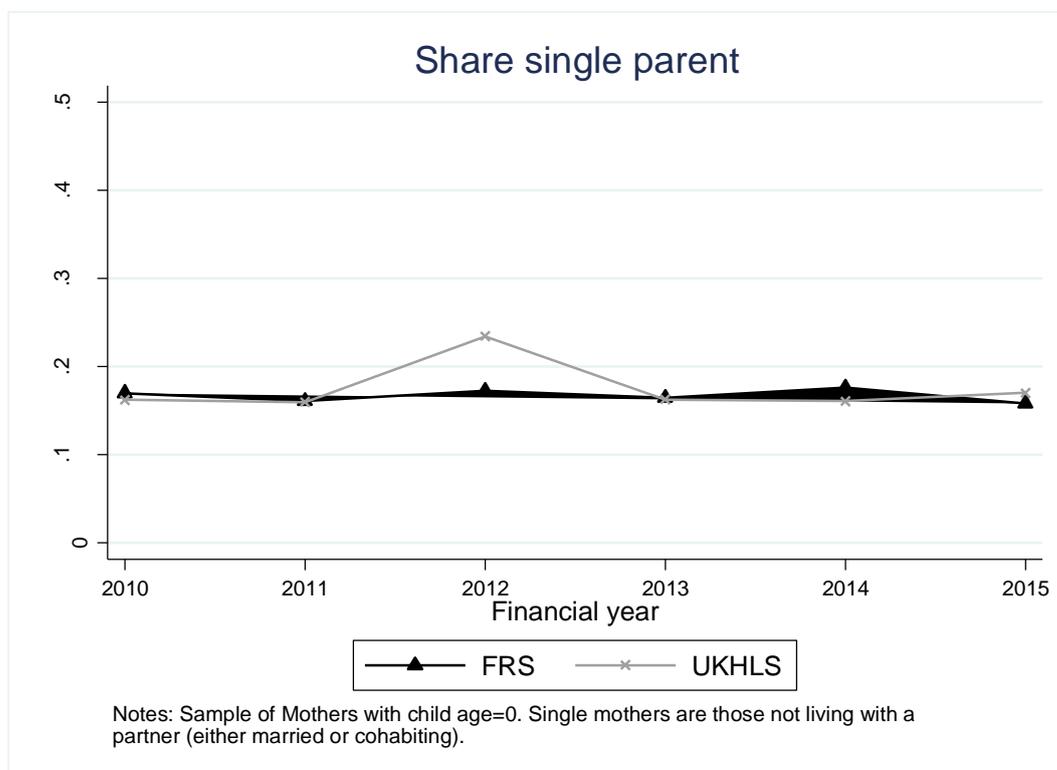To further check that our Understanding Society sample accurately captures the most disadvantaged women with newborns, we look at the share of them which are lone mothers. Lone mothers are a disadvantaged group of much research and policy interest i.e. both mothers and their children go on to have worse socio-economic outcomes and there are significant research agendas around this. We want to check that Understanding Society is representative of this group and that continues to be so over time. Figure 3 plots trends in lone parenthood shares. The top panel shows the rates and the lower panel the difference between the surveys with confidence intervals. The difference in the lone parent share is always small and statistically insignificant, excluding a one-off blip in 2012. This result lines up with what we saw for the poverty rates, and indicates that Understanding Society is representative of this group of disadvantaged mothers.

A different dimension of living standards is the type of housing occupied. Figure 4 shows for our sample of women with newborns their type of housing tenure and how this changes over time for each survey. Again, if Understanding Society became less representative of disadvantaged mothers, then we would expect it to diverge from the FRS over time. We measure housing tenure according to five categories: owned outright, owed with a mortgage, private rented and other, local authority rent, housing association rent. The latter two categories are associated with low cost housing and disadvantage. We see that relative to FRS, Understanding Society tends to underestimate the share of mothers that either own or own with a mortgage their accommodation. Similarly, we see that Understanding Society contains a high share of mothers living in LA/HA accommodation, for example 18.5 vs 23.3 percent in 2010. Importantly, though the difference has remained rather stable over time. So again we find no evidence that Understanding Society has become less representative over time, relative to the FRS.

This section has focused on the representativeness of women with newborns. A larger group of women that are of much research interest are women of childbearing age. We demonstrate Understanding Society representativeness of this group in the next subsection.
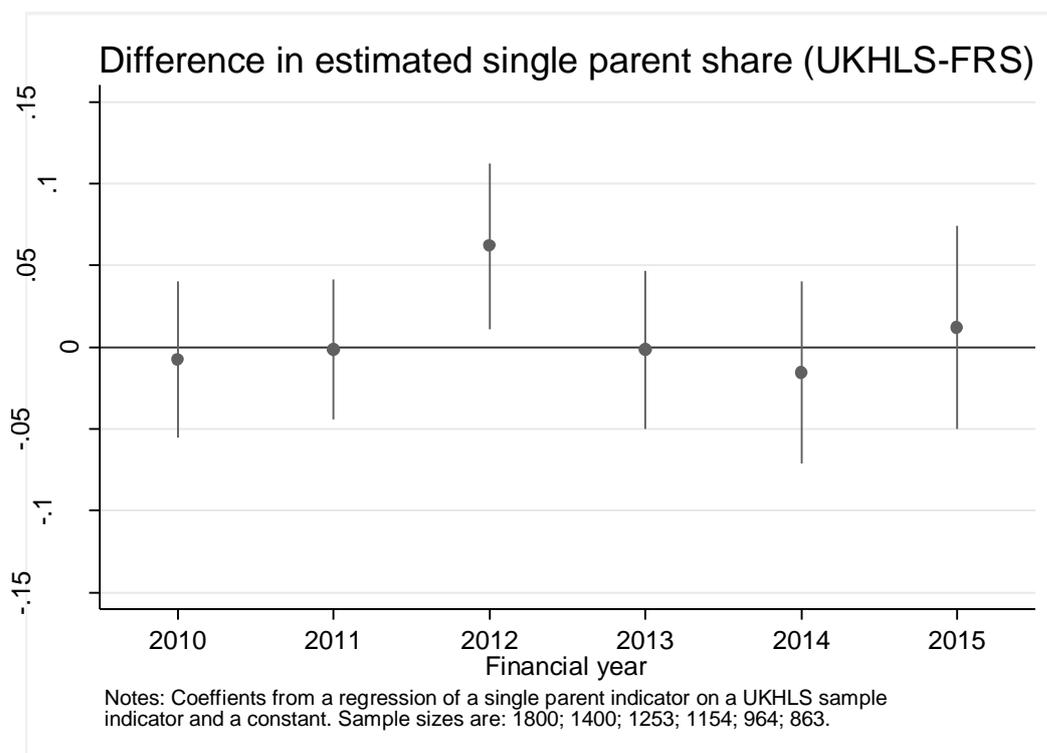
Figure 3: Share of mothers with newborns that are single parents

8. Figure 3A

Share single parent

Notes: Sample of Mothers with child age=0. Single mothers are those not living with a partner (either married or cohabiting).

9.

10. Figure 3B

Difference in estimated single parent share (UKHLS-FRS)

Notes: Coeffients from a regression of a single parent indicator on a UKHLS sample indicator and a constant. Sample sizes are: 1800; 1400; 1253; 1154; 964; 863.

11.

Figure 4: Housing tenure type for Mothers with newborns

Housing tenure by financial year

Notes: Sample of Mothers with child age=0.
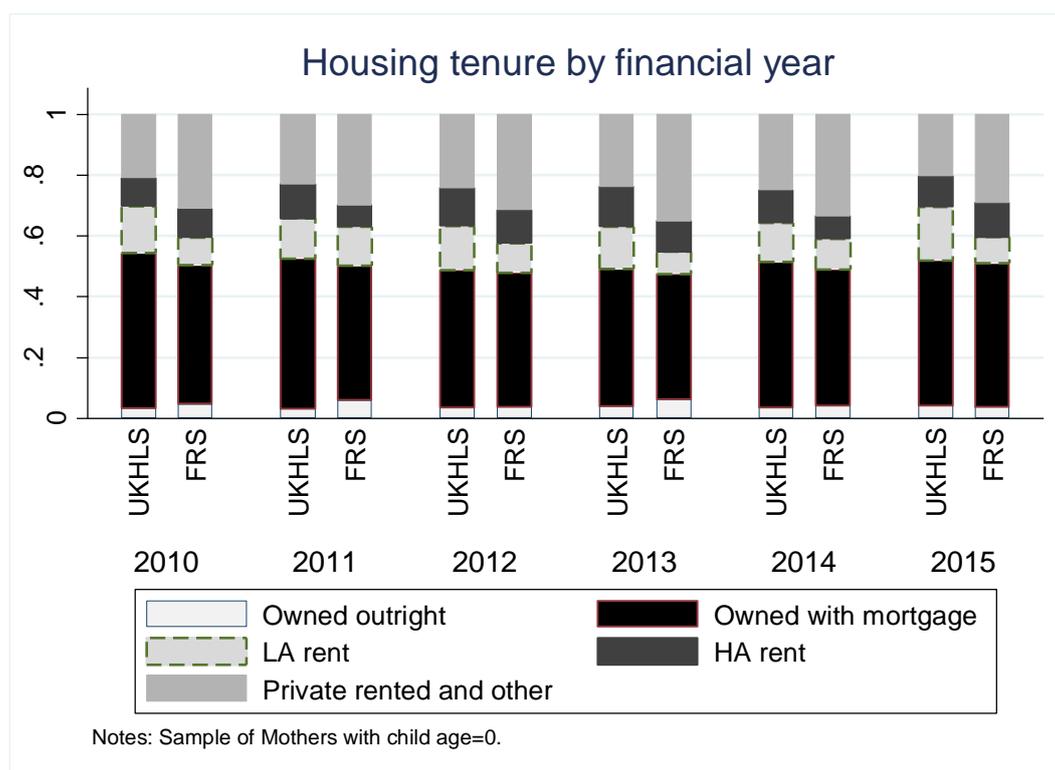
ii) Representativeness: Women of child-bearing age

This section repeats our Understanding Society vs. FRS comparisons but for women of childbearing age. As was the case with women with newborns, our aim is to examine whether the group are representative of the corresponding population group and so to document if and how they are observationally different from the FRS ones. In what follows, we see that women of childbearing age have similar demographic profiles in both surveys. Moreover, we see that their living standards are similar in both surveys and we find no evidence of divergence over time. We thus find no evidence that Understanding Society becomes less representative as it ages. We conclude, as with mothers with newborns, that Understanding Society starts out and remains representative and is thus a valuable data source for research on women of childbearing age.

Table 4 presents a descriptive comparison of Understanding Society and FRS women of childbearing age. For a set of demographics collected in both surveys the table reports the weighted means for each survey (columns 1 and 2), the difference in means and whether the difference is statistically significant (column 3). The final three columns present the standard errors (SE) and the respective sample sizes (N_ Understanding Society and N_FRS).

The majority of the differences are not statistically significant and when differences do occur, they are small in magnitude. We see small differences for teen, household size, mixed ethnic group,

Asian, black, East Midlands, Northern Ireland. Slightly larger differences, but still small in practical terms, are seen in the share married, white, other ethnic groups and residing in London. This repeats the pattern we saw above for women with newborns: women of childbearing age in the two surveys look remarkably similar (when appropriately weighted) in their demographic profile.

Table 4: Comparison of descriptive statistics for women of childbearing age

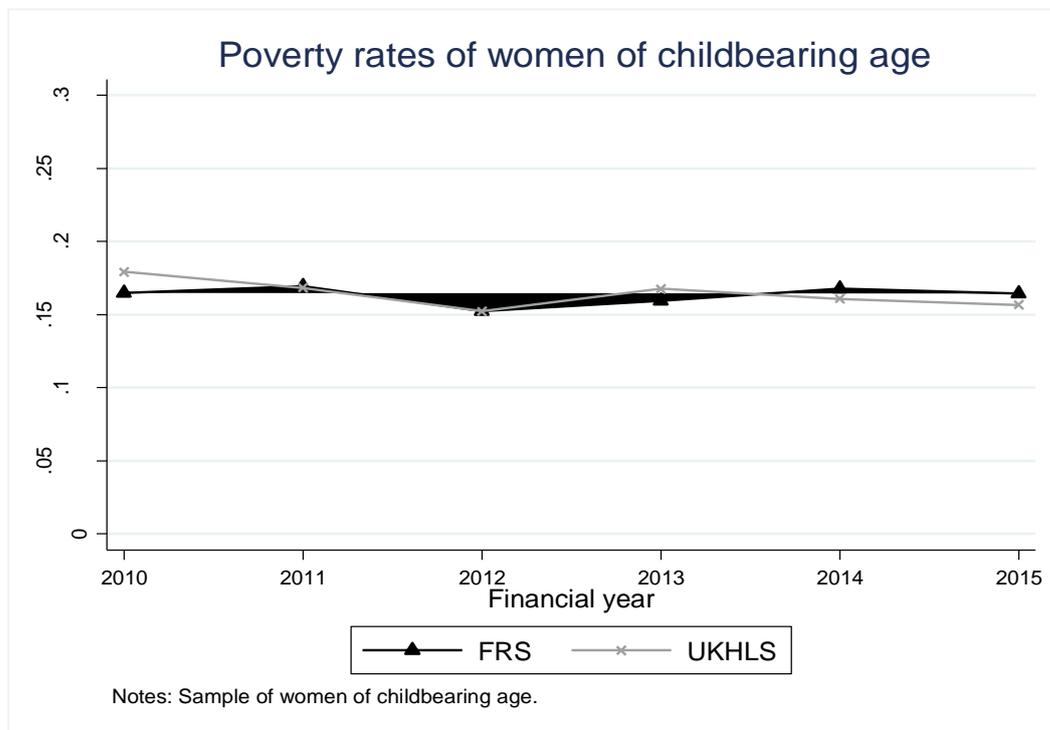|	UKHLS	FRS	Mean Diff	SE	N_ UKHLS	N_FRS
Age	30.49	30.39	0.0996	0.0883	61086	50688
Teen	0.137	0.119	0.0182***	0.0032	61086	50688
Household size	3.447	3.298	0.150***	0.0166	61086	50688
Married	0.350	0.372	-0.0219***	0.0053	61086	50688
Ethnicity:						
White	0.879	0.851	0.0272***	0.0037	58654	46723
Mixed ethnic group	0.0187	0.0160	0.0028	0.0016	58654	46723
Asian	0.0656	0.0729	-0.0073*	0.0028	58654	46723
Black	0.0287	0.0333	-0.0046**	0.0017	58654	46723
Other	0.00834	0.0265	-0.0181***	0.0014	58654	46723
Region:						
North East	0.0402	0.0398	0.0004	0.0025	61046	50688
North West	0.110	0.111	-0.0008	0.0041	61046	50688
Yorkshire and the Humber	0.0883	0.0831	0.0052	0.0037	61046	50688
East Midlands	0.0775	0.0697	0.0078*	0.0037	61046	50688
West Midlands	0.0856	0.0868	-0.0013	0.0036	61046	50688
East of England	0.0935	0.0894	0.0042	0.0040	61046	50688
London	0.132	0.152	-0.0199***	0.0050	61046	50688
South East	0.131	0.132	-0.0009	0.0043	61046	50688
South West	0.0779	0.0770	0.0009	0.0034	61046	50688
Wales	0.0497	0.0460	0.0037	0.0028	61046	50688
Scotland	0.0816	0.0832	-0.0017	0.0033	61046	50688
Northern Ireland	0.0322	0.0298	0.0024*	0.0011	61046	50688

Notes: Sample of women of childbearing age. Pools financial years 2010-2015.

* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Figure 5 shows how the Understanding Society poverty rate compares to the FRS ones by financial year. The top panel shows the poverty rates and the lower panel the difference with confidence intervals. We see that women of childbearing age experience very similar poverty rates according to both surveys. The difference between the two surveys is always small, statistically insignificant and does not systematically change over time. We conclude that the Understanding Society women look and remain very much like their FRS counterparts in terms of poverty.
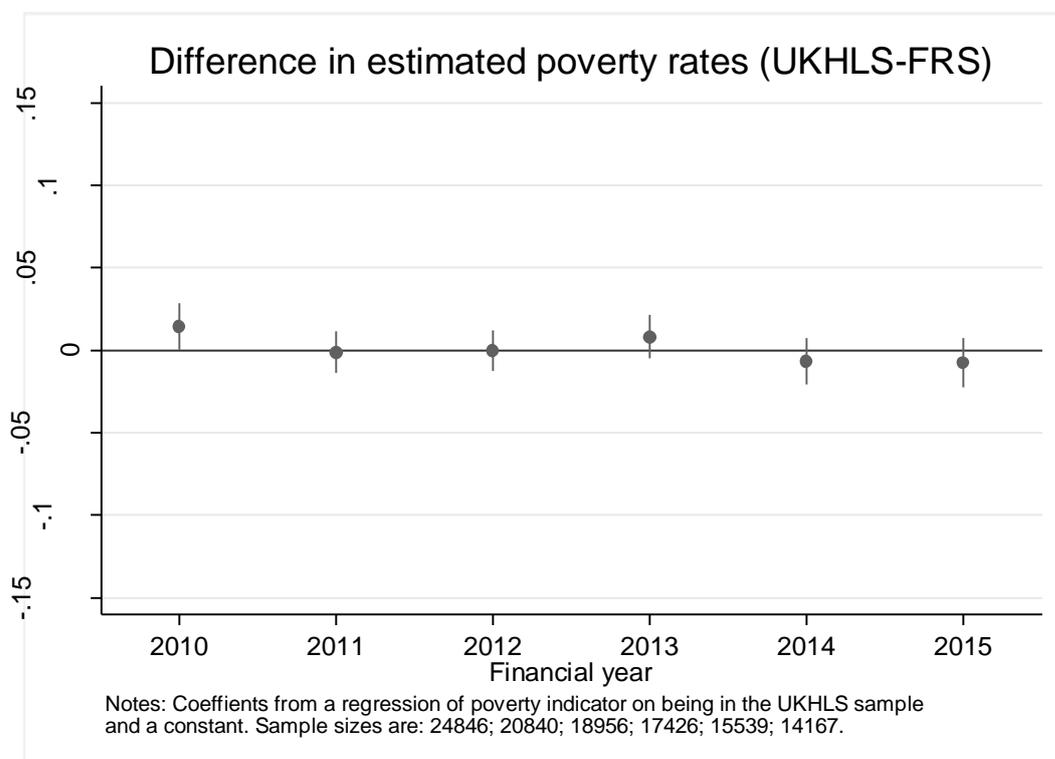
Figure 5: Poverty rates of women of childbearing age

12. Figure 5A

Poverty rates of women of childbearing age

Notes: Sample of women of childbearing age.

13.

14. Figure 5B

Difference in estimated poverty rates (UKHLS-FRS)

Notes: Coeffients from a regression of poverty indicator on being in the UKHLS sample and a constant. Sample sizes are: 24846; 20840; 18956; 17426; 15539; 14167.
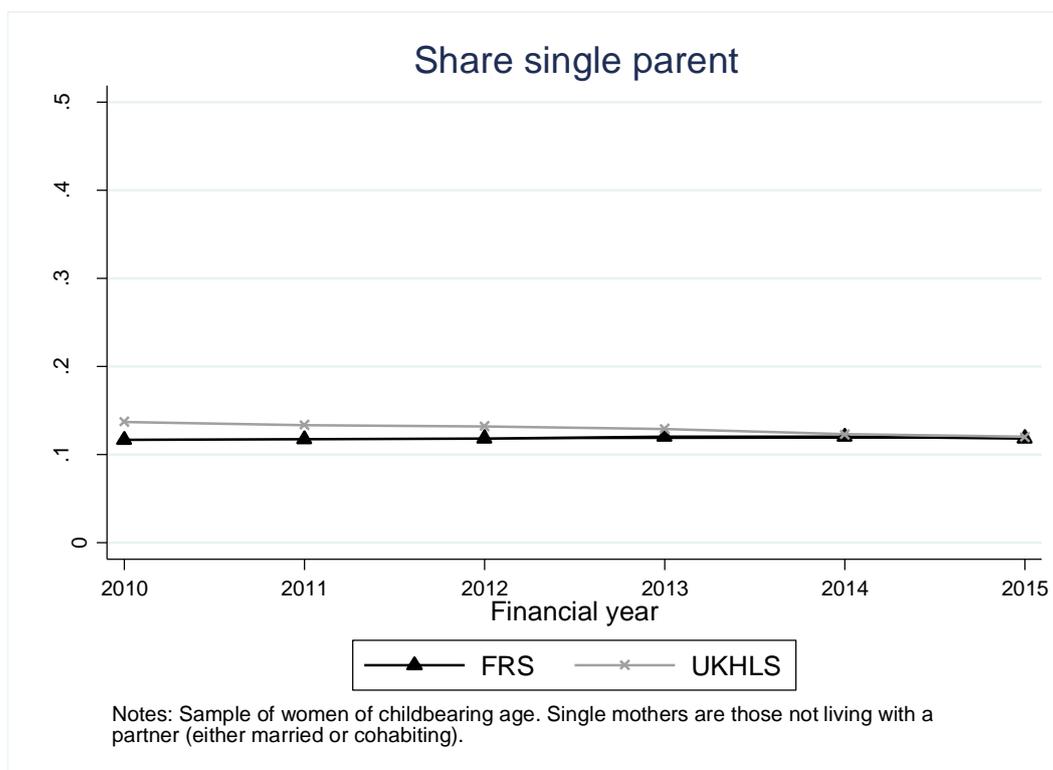
15.

Figure 6 shows our comparison for the share that are lone mothers. As lone mothers and their children are a disadvantage group, we want to confirm that they are accurately represented in Understanding Society. The top panel shows the rates and the lower panel the difference between the surveys with confidence intervals. The lone parent share is very similar in both surveys, although Understanding Society has a slightly higher share of this disadvantaged group than the FRS. For example, in 2010 13.7 percent of women of childbearing age are lone parents in Understanding Society but the corresponding FRS figure is 11.7 percent. The overestimation of Understanding Society looks to be systematic as the gap between the surveys is stable over time. But to emphasise again, the differences are fairly small in magnitude and one would draw the same substantial conclusions about the extent of lone parenthood from both surveys. The figure thus presents more evidence that the Understanding Society accurately covers the most disadvantaged women of childbearing age.

We complete this subsection with our final comparison of living standards which is of the type of housing occupied. Figure 7 shows for our sample of women of childbearing age their type of housing tenure and how this changes over time for each survey. Again, if both surveys represent the population of interest, then we expect any differences to be small and stable over time. The figure shows that relative to FRS, Understanding Society tends to underestimate the share of women that either own or own with a mortgage their accommodation. Similarly, we see that Understanding Society contains a high share of women living in LA/HA accommodation, for example 18.5 vs 23.3 percent in 2010. Importantly, though the difference has remained rather stable over time. So again this points to Understanding Society providing excellent coverage of women of childbearing age.

This section has confirmed that Understanding Society women of childbearing age, like women with newborns above, look and remain very similar to the corresponding ones from what is regarded as a high quality cross-sectional survey. In the next subsection, we go beyond demonstrating quality in the cross-sectional dimension and illustrate the power of Understanding Society as a basis for important longitudinal analysis.
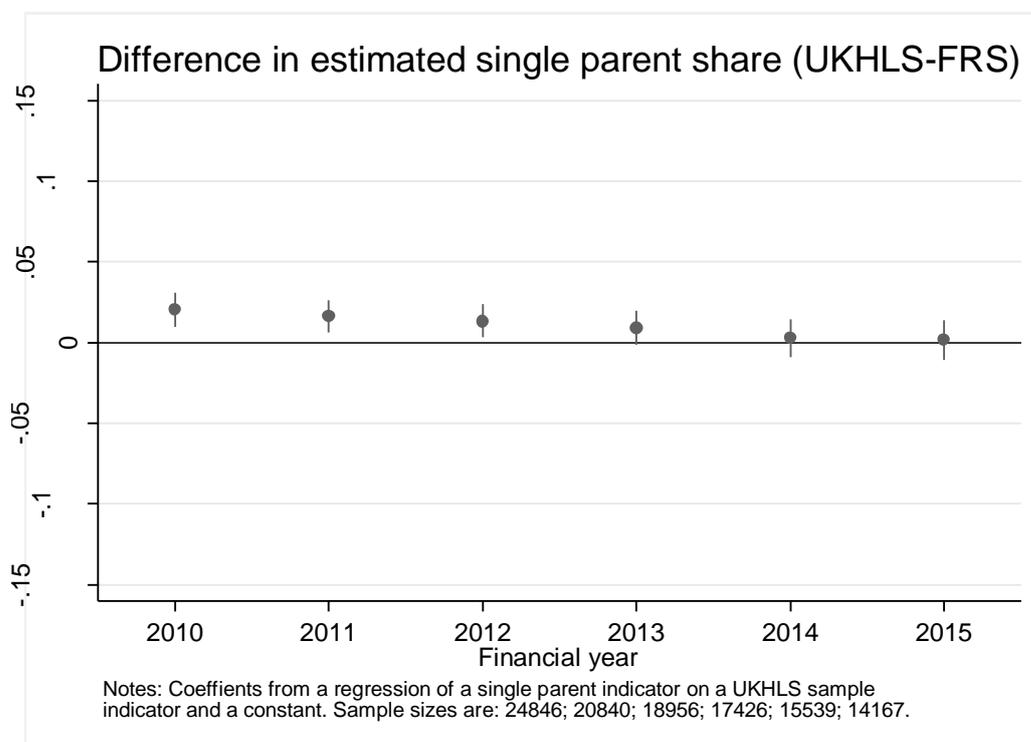
Figure 6: Share of women of childbearing age who are single parents

16. Figure 6A

Share single parent

Notes: Sample of women of childbearing age. Single mothers are those not living with a partner (either married or cohabiting).

17.

18. Figure 6B

Difference in estimated single parent share (UKHLS-FRS)

Notes: Coeffients from a regression of a single parent indicator on a UKHLS sample indicator and a constant. Sample sizes are: 24846; 20840; 18956; 17426; 15539; 14167.

19.

Figure 7: Housing tenure for women of childbearing age

Housing tenure by financial year

Notes: Sample of women of childbearing age.
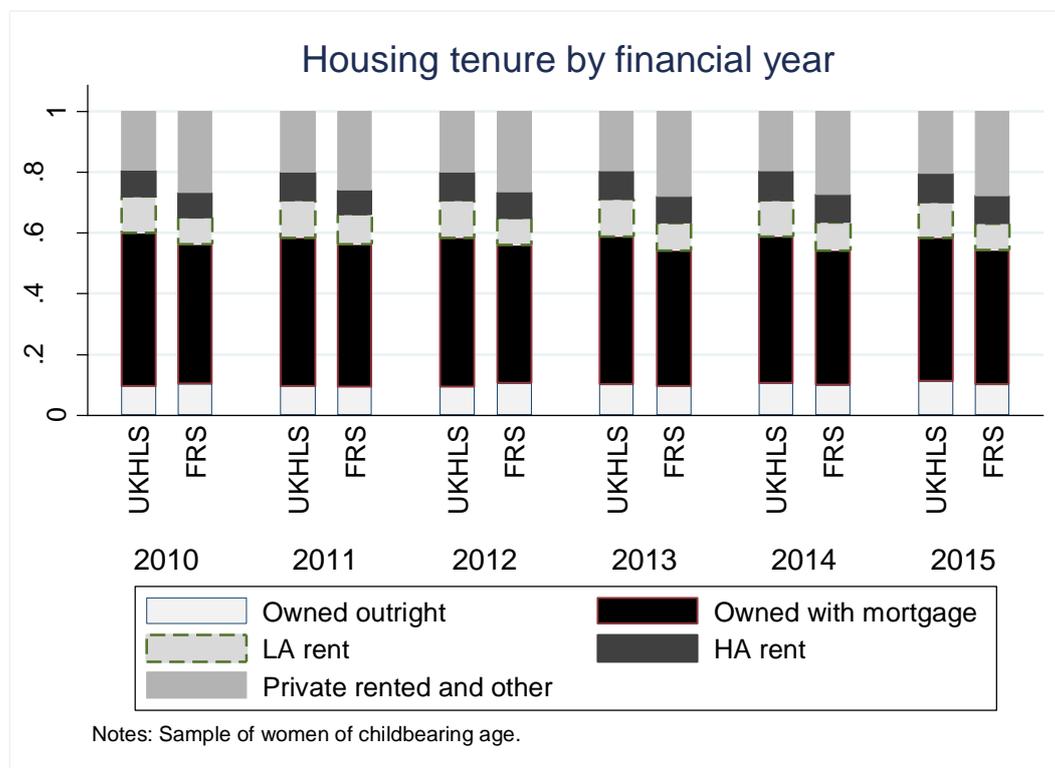
4. Value of Study for research in this area: persistent poverty rates

The previous subsections have shown that Understanding Society has large samples of both women with newborns and women of childbearing age. Moreover, the samples were shown to representative in the sense that they are closely comparable in demographics to those from what is regarded as a high-quality cross-sectional survey. In this subsection, we move beyond a purely cross-sectional analysis to demonstrate the power of Understanding Society in answering longitudinal research questions concerning women and their children. We extend the analysis of living standards of the previous subsections to report on persistent poverty rates for women with newborns and women of childbearing age.

A long spell in poverty is presumably worse than a short one. For example, during short spells of low income, a family might draw down savings or borrow from relatives to fund consumption. This is more difficult during longer poverty spells. Therefore, the cumulative effect on well-being of spending multiple years in poverty can be assumed to be greater than for a period of short duration. Moreover, in the context of women with children, long spells of poverty are a concern for future life outcomes of the children.

Historically, the data available in the UK has not allowed a breakdown of persistent poverty for smaller subgroups, such as mother with newborns. However, the large sample sizes of Understanding Society now make such an analysis of persistent poverty possible. The backdrop to which we answer this question is one of austerity and cuts to welfare income in the UK, which gives added urgency to this question.

Table 5 shows the number of years each women spent in poverty: women with newborns and women of childbearing age. A substantial 16.6 and 15.2 percent, respectively, spent at least one period in poverty. We see that many of our sample spent multiple years in poverty. For example, 1.6 and 1.1 percent of our women spent all eight years we observed them in poverty, whereas, around 10 percent of each group spent at least five years out of eight in poverty.

Figure 8 explores trends in persistent poverty rates for our two groups, where persistent poverty is defined as living in poverty for at least two of the last three years. Women with newborns have higher persistent poverty rates than the sample of women of childbearing age and so they represent a more deprived subset of this group. For example in 2011, the persistent poverty rate was 22.3 for women with newborns but only 15.5 for women of childbearing age. The trends in persistent poverty appear to be stable over time and so there is no indication that the problem of persistent poverty has lessened as we move away from the 2008 financial crisis, say.

Overall, the Understanding Society data with its repeated observations on the same individual, has given us a much richer understanding of patterns of low income than possible with the cross-
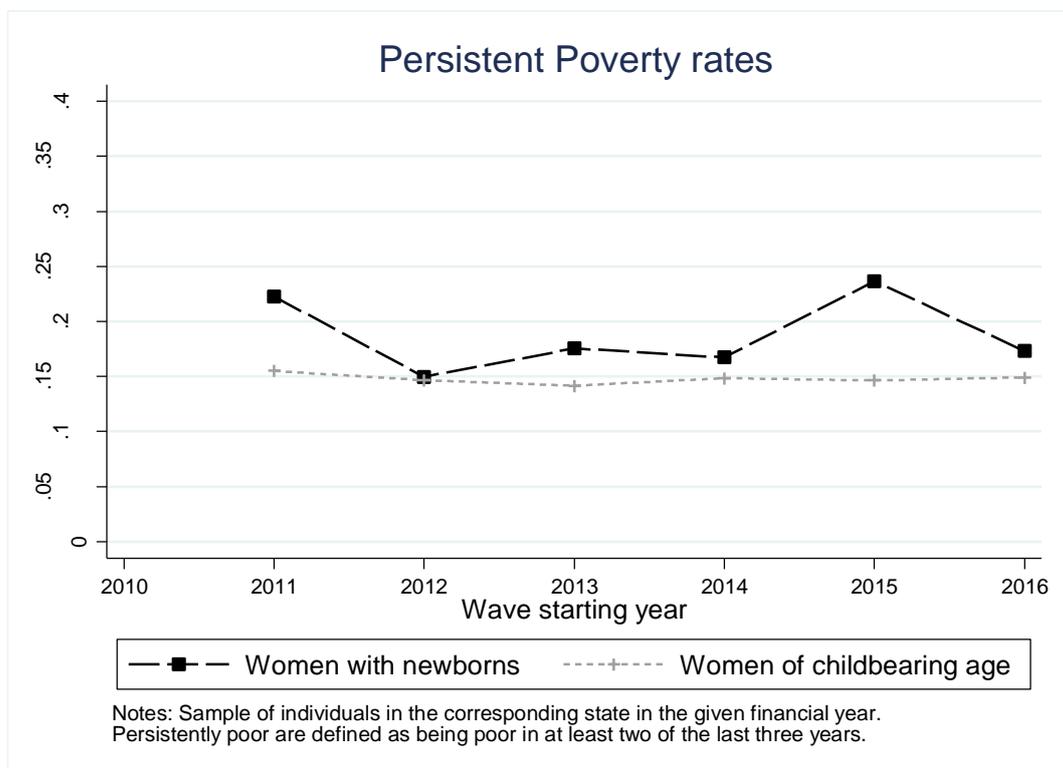
sectional data alone. The analysis has revealed the worrying pattern that there are a substantial number of women with newborns who have been in poverty for multiple years. This is a worrying public policy issue and leaves space for much more analysis of the Understanding Society data to understand the problem and consequences more deeply, for example, by exploring the role of unsecure employment, welfare cuts, living arrangements, marriage etc. in explaining persistent poverty but is beyond the scope of this report.

Table 5: Number of years spent in poverty

	Percent	
	with newborns	childbearing age
Number of years in poverty:		
0	48.88	55.14
1	16.58	15.23
2	9.57	9.5
3	6.69	5.95
4	5.97	4.82
5	4.87	3.33
6	3.21	2.97
7	2.62	1.93
8	1.62	1.12
Number of individuals:	1340	4151

Notes: Sample of individuals who were in the corresponding state (with newborns or of childbearing age) in at least one year.

Figure 8: Trends in persistent poverty

Persistent Poverty rates

Notes: Sample of individuals in the corresponding state in the given financial year.
Persistently poor are defined as being poor in at least two of the last three years.

5. Conclusions and recommendations

There are important research agendas around pregnancy, mothers and their newborns. Understanding Society is an excellent and important data source for researchers addressing questions in these areas. With eight waves of data now available, and as part of our ongoing process of quality control, this report has checked on the continuing representativeness of the mothers and mothers to be. The report concludes with three main findings.

First, the Understanding Society sample sizes of mothers with newborns and mothers of childbearing age are large. Large enough to allow for common types of analysis on small subgroups of interest such as the most disadvantaged mothers. For example, the pooled sample of single mothers is over 15,000, it is nearly 6,000 for women with newborns and over 5,000 are poor single mothers. Even when focusing on the number of unique cases i.e. not pooling, the sample sizes are still substantial.

Second, our Understanding Society samples line up closely to their FRS counterparts in terms of their demographic profile. Further, when we look at measures of deprivation (poverty rates, share single mothers, living in low-cost housing) the two surveys again give very similar estimates that remain close over time. These facts provide strong evidence that our Understanding Society groups of interest start out representative of their population counterparts and remain so as the panel ages. A caveat should however be noted. The FRS is not a gold-standard benchmark but rather also a survey. It suffers from non-response (56 percent at the latest survey), like Understanding Society, which weighting seeks to address. As such, our validations reflect two independent attempts to estimate the same population parameters. We interpret the fact that Understanding Society tracks this leading cross-section survey as indicative that it remains of high quality.

Third, Understanding Society opens up longitudinal research possibilities that are simply not possible with cross-sectional data such as the FRS. This report has given an illustration of one such type of longitudinal analysis. It has examined the persistent poverty rates of women of childbearing age and women with newborns. First, the analysis concludes that women with newborns experience higher poverty rates than the overall sample of women of childbearing age. Second, a substantial share of both groups live in poverty for multiple years and this trend has been stable since the financial crisis. While the first fact could have also been established from cross-sectional data, the latter could not be identified with cross-sectional data alone. This is only one such example of a type of longitudinal analysis that is possible with Understanding Society. Creative researchers will no doubt contribute many more over the years to come (many large research agendas are clearly longitudinal in nature and so could exploit Understanding Society e.g. child ageing and development, living standards through pregnancy and early childhood, fertility timing and the evolution of health through pregnancy). In sum, the longitudinal dimension of Understanding Society makes it uniquely placed to enrich our understanding of the experience of parents and their children. This makes it well placed to be the data source of highly important contributions to both public policy and research.

References

Mike Brewer, Olga Gdula and Robert Joyce (2017) "Refining the bootstrap methodology for HBAI statistics", mimeo, Institute for Fiscal Studies

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

Peter Lynn and Magda Borkowska (2018), "Some Indicators of Sample Representativeness and Attrition Bias for BHPS and Understanding Society", No. 2018-01, Understanding Society at the Institute for Social and Economic Research.

Office for National Statistics (2017), "Conceptions in England and Wales: 2015", Newport: Office for National Statistics,
https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/conceptionandfertilityrates/bulletins/conceptionstatistics/2015