

**Understanding Society: *UK Household Longitudinal Study:*
*User Guide to ethnicity and immigration research***

USER GUIDE

12th Edition, October 2024

Contents

1. Overview of the study	2
1.1 Main survey.....	2
1.2 Changes to the main survey due to the COVID-19 pandemic	3
1.3 The COVID-19 survey (April 2020 – September 2021).....	4
2. Why study ethnicity and immigration in a longitudinal household panel survey?	5
3. Sample design for surveying ethnic minorities and immigrants.....	11
3.1 Understanding society sample design	11
3.2 Ethnic Minority Boost sample.....	12
3.3 The Immigrant and Ethnic Minority Boost Sample	14
3.4 Fieldwork, survey instruments, interview mode	16
3.5 Sample status and following rules	18
4. Specific content.....	20
4.1 The extra five minutes sample.....	20
4.2 Extra five minutes questions.....	21
4.3 New IEMB questions	23
5. Response and attrition	24
6. Data files and question coverage	28
6.1 Data structure, naming conventions, key variables.....	28
6.2 IEMB vs GPS-EMB-BHPS sample Questionnaires in Wave 6	30
6.3 Key variables for ethnicity and immigration research	30
6.4 Notes on educational qualification variables.....	33
6.5 Notes on determining country of birth for children in Understanding Society.....	34
7. Issues for analysis: sample design, non-response, weighting.....	35
7.1 Why use weights?	35
7.2 Available weights	37
7.3 Analysis and Sample design variables.....	37
8. Approach to translation and language	39
9. Help and support.....	41
9.1. Understanding Society resources	41
9.2. Resources for ethnicity and immigration research.....	41
10. Citation information and ethical approval statement	42
10.1 Data citation.....	42
10.2 Citation for this Guide:.....	42

10.3 Ethics	42
11. References	43
12. FIND OUT MORE.....	44
Appendix I: Screening questions.....	45
Appendix II: List of variables corresponding to the extra five minutes questions, Waves 1-14	47
Appendix III: List of variables corresponding to the new questions asked of the IEMB sample in Wave 6	50
Appendix IV: Names of occupation (SOC) variables	51
Appendix V: List of questions to measure ethnicity	52
Appendix VI: Syntax for derived variables mintent_dv mintwhen_dv mreason?_dv.....	57
Appendix VII: Fieldwork periods for different samples	59

1. Overview of the study

1.1 Main survey

This guide provides an introduction to carrying out ethnicity and immigration related research with *Understanding Society*. It is intended to complement other [documentation](#) providing support and guidance on the study.

Understanding Society, the UK Household Longitudinal Study (UKHLS), is a large longitudinal survey of households in the United Kingdom (England, Scotland, Wales and Northern Ireland). Members of households recruited at the first round of data collection are visited one year later to collect information on changes to their household and individual circumstances.

Since the survey started in 2009, all eligible adult respondents in participating households are interviewed annually. Each annual interview is referred to as a “wave”. As interviewing takes place over two years each wave covers an overlapping two-year period. Adult respondents (16+ year olds) were originally interviewed face-to-face (with a self-completion component) with approx. 500 households being interviewed by telephone, but from Wave 8 onwards an increasing proportion complete the interview by themselves online. During the Covid-19 pandemic, all interviews were conducted online and face-to-face interviewing as applicable was resumed in 2022 (see the next sub-section for further details on changes during the pandemic). Children aged 10-15 complete a separate self-completion questionnaire, becoming eligible for the main questionnaire when they reach 16. Over time a detailed picture of individuals’ lives builds up through the annual interviews. In addition, each annual survey provides information about changes and continuity in respondents’ lives since the last interview, changes which can be related to their characteristics and to other events that have taken place in their lives.

Data from each wave of the study is available from the UK Data Service (see the [Data Access page](#) for further details). Analysts should consult the [Understanding Society Main Survey User Guide](#) for further information on the overall study.

The purpose of *Understanding Society* is to provide high quality longitudinal data on the population, covering subjects such as health, work, education, income, family, and social life to help understand the long-term effects of social and economic change, as well as policy interventions designed to impact upon the general well-being of the UK population.

Enabling research on ethnicity and immigration and comparisons of the experience of the UK’s ethnic groups was one of the original aims of *Understanding Society*. This was reflected in three core aspects of the study. First, the main sample was supplemented by an [ethnic minority boost sample](#) of over 4,000 households. Along with the ethnic minorities in the main sample, this boost sample enables detailed analysis of and comparison across specific ethnic groups. We discuss the design and implications for analysis of the inclusion of the

boost sample further below. A further [immigrant and ethnic minority boost sample](#) of 2,922 households was added at Wave 6. Again, we discuss the design and implications for analysis of this additional refreshment sample, below.

Second, ethnicity was considered to be a core topic area for this survey (in addition to the usual topic areas such as socio-demographic characteristics, partnership, fertility, employment, income, wealth, health and education). As a result, questions that were salient for research into ethnicity and comparisons across ethnic groups were included in the questionnaire. Questionnaire content was subject to an extensive consultation process. For example, questions on country of birth of the respondent, their parents and grandparents were included in the first wave of the survey.

There were other questions that the consultation process identified as key to ethnicity related research but less relevant for the majority ethnic group (except as a comparison group). For example, remittance behaviour, and experiences of (racially motivated) harassment and discrimination. So, the third strategy was to allocate an additional five minutes of question time to a sub-sample of respondents comprising ethnic minorities and a comparison sample. This sample is referred to as the [extra five minutes sample](#). From Wave 7 it has been extended to include the new immigrant and ethnic minority boost sample and all immigrants surveyed in the main sample. The coverage and implications for analysis of the extra questions asked of the extra five minutes sample are discussed further in Section 4.

1.2 Changes to the main survey due to the COVID-19 pandemic

In 2020 when the Covid19 pandemic started the survey underwent a few changes. We summarise the key changes here. For further details please see the [Changes due the COVID-19 pandemic document](#).

- All interviews were conducted via web from April 2020 onwards and if they didn't complete the online questionnaire, they were followed up by interviewers over the telephone. At that time the fieldwork for Wave 11 Year 2 and Wave 12 Year 1 (a few re-issued Wave 10 cases) was ongoing and so this change affected those cases. This change in fieldwork, mode transition, could have implications for measurement, particularly measurement of change. Fieldwork did not fully return to face-to-face interviewing (for those subject to this mode) until after April 2022.
- The overall response rate for the 2020 samples was lower by 1.1 percentage points as compared to 2019 but this difference was higher in some sub-groups including ethnic minorities.
- The questionnaire content for Waves 11 and 12 changed midway through the fieldwork in response to the pandemic, to accommodate COVID-19 related questions. One set of questions were about the disease (such as Coronavirus symptoms, testing and results, whether respondents were hospitalised or shielding and long-term health conditions) and another set of questions explored the "furlough" scheme and whether respondents were eligible and had applied for the HM Revenue and Customs Self-

Employment Income Support Scheme grant, how much was awarded and the period of grant cover, start and end month and year.

- If you are interested in analysing changes during the pandemic (2020) as compared to the pre-pandemic period (2019), please look at the analysis guidance in the [Changes due the COVID-19 pandemic document](#).
- Weights were produced to account for changes in attrition and non-response patterns. As always, users are advised to use the appropriate weights for their analysis from the basket of weights provided, as these weights account for sample design and non-random attrition.

1.3 The COVID-19 survey (April 2020 – September 2021)

In April 2020, all 16+ year old sample members in households where at least one interview was completed in the last two waves were invited to participate in a short (20min) online survey about their experiences during the Covid-19 pandemic. They were interviewed again in May, June and July 2020. After that survey became less frequent, with further follow-ups in September 2020, November 2020, January 2021, March 2021 and September 2021. For more information about this survey please see the [COVID-19 survey documentation page](#).

2. Why study ethnicity and immigration in a longitudinal household panel survey?

Understanding Society brings an unprecedented opportunity to study issues of continuity and change within and across the UK's ethnic and immigrant groups. It enables analysis of inequalities, and of transitions over time and generations. It allows researchers to investigate whether, how and why individuals from different ethnic backgrounds have different life outcomes. Such analyses can help policy makers to better understand the complexity of the current population and to identify those factors that drive inequalities.

Studying ethnicity in survey research brings a number of challenges. The concept of ethnicity is contested and its definition and operationalisation varies across disciplines. Moreover, there are distinctions in the extent to which researchers and analysts adopt an understanding of ethnicity as inherently meaningful or regard ethnic groups as proxying for other factors, or collections of factors, that are (currently) unmeasured. See further the discussion in Burton et al. 2010.

The concept of ethnicity can be approached in a number of ways. For example, it can be understood as commonality within a group or as differences from 'other' groups. Its multifaceted nature makes measurement using a single measure challenging, and yet that is often what is available to researchers. Researchers bring their own agendas to the topic. For example while social psychologists may be interested in understanding individuals' identity, sociologists may be interested in social stratification and the role of ethnicity within that. Other social and health scientists may be primarily interested in the role of immigration in shaping individuals' outcomes and trajectories in the country of destination.

As a result, *Understanding Society* incorporates an exceptionally rich and varied range of measures to capture different dimensions of ethnicity, immigrant status and ethnic identity, namely:

- 1) [ONS 2011 Census, single category question on ethnic group](#)
(See further, the detail on the [ONS website](#))

Developing an ethnic self-identification (categorical) question is a major challenge. In addition to standard survey design issues such as question wording, the design of such a question requires a fine balancing act between trying to get consistent, reliable measures of ethnic group and capturing people's perception of their own ethnic identity. The response categories need to be meaningful, acceptable and consistent, so that people will be able to respond to them in expected, and predictable, ways. The ONS question represents a national-level attempt to provide such a simple single-category solution. While it is subject to some critique, it is widely used across government and other surveys as well as administrative sources, and hence provides an important point of reference for comparison. See further the documents relating to the harmonised recommended questions on the [ONS website](#).

This question was asked of the whole adult sample at Wave 1, and continues to be asked whenever someone is interviewed for the first time. Note the response codes in the overarching “*Black/African/Caribbean/Black British*” category for the sub-categories Caribbean and African are 14 and 15 in Understanding Society and 15 and 14 in the ONS 2011 Census harmonised versions. Please refer to the showcard for each wave to see the exact response options that were available to the respondents. Telephone respondents were given the telephone version of the question – where the broad groups were first read out. Depending on the overarching category chosen, the sub-categories were then read out.

2) National identity, multiple response

As with the implementation of the ONS ethnic group question in the 2011 Census, *Understanding Society* includes the multiple response [national identity question](#) that was asked directly before the ethnic group question. The inclusion and positioning of the national identity question in the 2011 Census was designed to enable respondents to express national identity separately from ethnic group. For example, someone could choose their national identity as “Welsh” and “British” and their ethnic group as “Black or Black British: Caribbean”.

This question was asked of the whole sample at Wave 1, and, as with the ethnic group question, whenever someone is interviewed for the first time.

3) Country of birth (across multiple generations), and parental ethnic group

Immigrant status is of core interest to a wide range of researchers. Thus, it was vital to include the respondent’s country of birth and their date of arrival in the UK. However, immigrant *generation* is also of substantial interest, hence the countries of birth of the respondent’s parents and grandparents were also asked. This not only allows outcomes across immigrant generation to be tracked, it also enables researchers to relate ethnic identity to associations with countries outside the UK. For example, it is possible to compare individuals who report their ethnic group as “Asian or Asian British: Indian” and their country of birth as India with those reporting the same ethnic group but born in Uganda. Having measures of country of birth for multiple generations also facilitates the construction of measures of ethnicity that combine information from across these different sources on heritage.

This question was asked of the whole sample at Wave 1, and, as with the ethnic group question, whenever someone is interviewed for the first time.

4) Parental ethnic group

It was also important to consider transmission of ethnic group identification. Parents and children may not consider their ethnic groups to be the same. This may be more relevant for children of mixed parentage. To empirically measure this, respondents were

asked about their father's and (if different from their father's) their mother's parents' (father's and – if different – mother's) ethnic group.

These questions were asked of the whole sample, in Wave 1 only.

5) Strength of British identity and of identification with parental ethnicity

There is an extensive and developing literature on the identification of minority groups with both majority or national and minority identities, and how this relates to constructs of ethnic group and individuals' relationship to the majority society. For those interested in such issues, there are measures of strength of British identity and strength of identification with parents' ethnicity.

Strength of identification with parents' ethnic group(s) was only asked of those who reported their parents' ethnic group and reported their own ethnic group as anything other than "White: British/English/Scottish/Welsh/Northern Irish

This question was only asked of the extra five minutes sample in Wave 1 and Wave 3.

6) Measures of identity, belonging and pride in domains of ethnic identity

To better understand how identity is expressed and experienced at different points in the life course and in relation to ethnicity or immigration status, a suite of questions is included in *Understanding Society*, which complements the question on ethnic group. This stems from the fact that ethnic group may potentially primarily reflect ascribed rather than 'owned' ethnicity. This additional suite of questions taps into different psychological dimensions of identity (pride, belonging) across a series of domains (country of birth, ethnic group etc.).

These questions were asked of the extra five minutes sample in Wave 2 and are repeated every six waves. A subsample of the extra five minutes sample consisting of 16-19-year olds and new immigrants – that is those who arrived within the three years prior to being interviewed in Wave 1 – are asked at more frequent, three-yearly intervals.

7) Ethnicity of associates

Social understandings of ethnicity and its behavioural and social conception or implications are assisted by measures that identify ethnicity of friends and networks. Such measures, that also identify other characteristics of friendship networks, are therefore also incorporated in the study.

These questions were asked in Wave 3, with some questions being asked of the whole sample, and others of the extra five minutes sample only. This module is asked at 3-year intervals.

8) Harassment and discrimination

The salience of ethnicity and ethnic group identity has been linked to the boundaries that are set up to exclude those regarded as 'other'. Hence measurement of experience of discrimination or harassment can provide insight into how these experiences impact the lives of ethnic minorities in the UK.

These questions are asked of the Extra Five Minutes sample every other wave, starting in Wave 1.

9) Religion

Measures that may be associated with particular ethnic or national origins or be used in construction with them are also of value in enabling researchers to operationalise ethnic group constructs that are of most interest to them. One such measure is religion. Many researchers prefer using ethno-religious groups to ethnic groups as analytic categories because single ethnic group categories as defined by the Census 2011 ethnic group question may cover relevant sub-populations who can most easily be distinguished by their religious affiliation. For example, individuals reporting their ethnic group as "Asian or Asian British: Indian" and their religion as "Hindu" may be different in terms of background, behaviour and life chances to those reporting the same ethnic group but a different religion such as "Sikh" or "Muslim".

In this module of questions, respondents are asked about their current religion and those who say they do not have a current religion are asked about the religion they were brought up in. These questions were asked of the whole sample in Wave 1 and are asked whenever someone is interviewed for the first time.

10) Language

Language is also embedded in constructions of ethnicity and minority status and so a question on childhood language was asked of the whole sample at Wave 2. In Waves 8 and 9, this question was asked of new entrants and for continuing members if this information was missing.

Additional questions on whether English language is their first language and if not, questions on English language proficiency were asked in the initial wave. But the questions about English language proficiency are not considered to be a component of ethnicity. In Waves 5 and 10, if this information was missing for the extra five minutes sample, then they were asked again.

11) Ethnic minority boost screening question

The composition of the ethnic minority boost (EMB) sample for *Understanding Society* was intended to maximise the analytical potential of the sample for comparisons of ethnic groups and within minority groups. The screening question was therefore intended to capture a number of specific groups that have some internal coherence and

are sufficiently numerous for meaningful analysis. Thus, its operationalization was driven by pragmatic rather than analytic concerns. It is not intended to reflect ethnic identity but is nevertheless available for researchers if they wish to understand how respondent households came to be included in the sample. This was recorded as a series of household level variables collected at the time of screening.

In addition to the information collected at the screening stage, the same ethnic group information was also collected for every 10+ year-old enumerated in the household. This was collected as part of the household grid (to help in producing design weights) and so it is not self-reported.

This was only asked of the EMB sample in Wave 1. The exact wording is shown in [Appendix I](#).

12) Immigrant and ethnic minority boost screening question

As with the EMB, a screening question was used to assess eligibility for the immigrant and ethnic minority boost sample (IEMB). It used comparable ethnic group categories to the EMB screening question; but it also identified those who were not from one of these groups and were not born in the UK. Once again, its use was practical rather than analytical: to maximise the inclusion of immigrants and ethnic minorities according to the design. This was recorded as a series of household level variables collected at the time of screening.

In addition to the information collected at the screening stage, the same information (ethnic group and whether born in the UK) was also collected for every individual enumerated in the household. This was collected as part of the household grid (to help in producing design weights) and so these are not self-reported. This screening question was only asked of the IEMB sample at Wave 6. The exact wording is shown in [Appendix I](#).

13) Citizenship

Respondents who were not born in the UK were also asked questions about their citizenship in Wave 1. They could respond that the country of their citizenship was UK, and/or up to two other countries. All new entrants who were not born in the UK were also asked these questions in Waves 2, 3 and 4. From Wave 5 onwards, in addition to new entrants not born in the UK, continuing non-UK born members who had not reported UK citizenship were asked this question again. The IEMB sample in Wave 6 who were non-UK citizens were also asked whether they had indefinite leave to remain and if they were intending to apply for UK citizenship.

Information on these variables, including variable names and at which waves they were asked are provided in [Appendix I](#), [Appendix II](#), [Appendix III](#) and [Appendix V, and the rotation of the extra five minutes questions in Tables 4.1a and 4.1b](#).

[Full questionnaires](#) for each of the waves completed and the one currently in the field are available on the Understanding Society website. The [long-term content plan](#) is also available on the website. This includes a list of all the questionnaire modules and the frequencies at which they are asked.

Overall, *Understanding Society* furnishes researchers with a large repertoire of ethnicity-related questions to enable them to carry out analysis on topics across the survey according to the framework of their discipline and research questions.

[Ethnic group of interviewers](#)

This information is also recorded and provided by the fieldwork agencies and made available in the data file on interviewer characteristics. This data file is available separately and can be accessed under Special License from the [UK Data Service \(SN8579\)](#).

3. Sample design for surveying ethnic minorities and immigrants

3.1 Understanding society sample design

Understanding the sample design is important for any research using *Understanding Society* data including ethnicity research. The *Understanding Society* sample has multiple components. The design of all components is described in more detail in three [Understanding Society working papers](#) (Lynn et al. 2016; Lynn 2009, Berthoud et al 2009).

- The **General Population Sample (GPS)** consists of two separate samples of residential addresses: one for Great Britain (that is, England, Scotland and Wales) and another one for Northern Ireland. The **Great Britain sample** is a proportionately stratified (equal probability), clustered sample of addresses selected from the Postcode Address File. The **Northern Ireland sample** has an unclustered systematic random sample of addresses selected from the Land and Property Services Agency list of domestic addresses, *with selection probability which is twice from that of the Great Britain sample*. Interviewing for the GB part of the GPS takes place over 24 months, while the Northern Ireland part is interviewed over the first 12 months. Each monthly sample of the GB part of the GPS is a random sub-sample but that is not the case for the other samples. At Wave 1 the GPS had a sample size of approximately 26,000 participating households.
- The **former BHPS sample** became a part of the UKHLS at Wave 2. It has all members from the BHPS sample who were still active at Wave 18 of the BHPS and who had not refused consent to be issued as part of the *Understanding Society* sample. It should be noted that the BHPS sample itself has multiple components: a nationally representative sample of addresses in Great Britain (south of the Caledonian Canal) in 1990, boost samples of Scotland and Wales added in 1999, and a boost sample of Northern Ireland addresses added in 2001. For further details of the BHPS sample, see section IV of the [BHPS User Manual](#) (Taylor et al 2018). The BHPS sample data collection takes place over the first 12 months for each wave. At Wave 2 when the BHPS sample was added to *Understanding Society*, the BHPS sample had a sample size of approximately 6,600 participating households.
- **Ethnic Minority Boost (EMB) sample**, like the main sample, is collected over a 24-month survey period for each wave. [The design is described further below](#). At Wave 1 the EMB comprised approximately 4,000 participating households
- **Immigrant and Ethnic Minority Boost (IEMB) sample** was added in Wave 6 and was collected over the second year of Wave 6 fieldwork, that is, 2015. As the survey interviews individuals annually, from Wave 7 onwards this sample will be interviewed in the second year of the 24-month fieldwork period. [The design is described further below](#). At Wave 6 the IEMB sample comprised approximately 2,900 participating households.
- A **General Population Sample boost (GPS2)** of approximately 5,700 households was added in Wave 14 (2022/23). The sample was clustered in Great Britain but unclustered in Northern Ireland sample (with the sample selected with equal selection probability as the Great Britain part).

For all sample components, once addresses were selected, up to 3 dwelling units at each address were randomly selected and at each dwelling unit up to 3 households were randomly selected. Dwelling units are defined in [interviewer instructions](#) as “a living space with its own front door – this can be either a street door or a door within a house or block of flats. Usually there is only one dwelling unit at an address”. Households are defined as the person or people occupying particular shared spaces, namely, “a **household** is when one person or a group of people have the accommodation as their only or main residence and share at least one meal a day **or** share the living accommodation” Around 95% of addresses have one DU and one household – multiple DUs/households tend to be concentrated in urban areas.

3.2 Ethnic Minority Boost sample

The Ethnic Minority Boost sample was designed to provide at least 1,000 adult interviews from each of five target ethnic minority groups: Indian, Pakistani, Bangladeshi, Caribbean, and African, as well as somewhat less comprehensive coverage of other minority ethnic groups within the sampled areas.

The EMB sample design comprised of screening addresses in high ethnic minority concentration areas in Great Britain. **Note that Northern Ireland does not contribute to the EMB sample.** The initial step was identifying postal sectors with relatively high proportions of relevant ethnic minority groups, based on 2001 Census data and more recent Annual Population Survey data. This identified 3,145 sectors or approximately 35% of the sectors in Great Britain. It covers between 82% and 93% of the population of the five target ethnic groups in the UK.

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

The number of addresses selected per postal sector ranged from 15 to 103. Sampling fractions varied across the sectors in a way designed to deliver target numbers of respondents in each target ethnic minority group with adequate statistical efficiency (see Berthoud et al., 2009 for more details).

Once the dwelling unit and household was selected, the final stage of sampling was screening conducted by the interviewers. [The Understanding Society Wave 1 Project Instructions](#) for interviewers outlines the procedures. As a result the [Address Record Forms for the EMB sample](#) were different from those used with the General Population Sample. Within each household in each selected address, rather than all resident persons becoming sample members, there were two additional steps:

- 1) A “screen” was carried out to identify whether there were any persons who were from any of the specified minority ethnic groups in the household. The household member answering the screening question was asked **“Does anyone living at this address come from, or have parents or grandparents from any of the following ethnic groups?”** The ethnic groups that they could choose from were Indian, Mixed Indian, Pakistani, Bangladeshi, Sri Lankan, Caribbean/West Indian, Mixed Caribbean/West Indian, North African, Black African, African Asian, Chinese, Far Eastern, Turkish, Middle Eastern / Iranian AND None of these. The screening question is provided in the [Appendix I](#).
- 2) If the response was anything other than “None of these” then the household had a positive selection probability (“sampling fraction”). At the outset the sampling fraction was not 100 per cent for all ethnic groups but differed across ethnic groups according to the sample design stage (see Berthoud et al 2009). In other words, not every household that included someone with an ethnic minority background was chosen. A review after six months concluded that sampling fractions for all groups had to be increased and additional addresses had to be screened for Bangladeshi members to reach the target. As a result, for most groups the sampling fractions were increased to 100 per cent. The only exception was Indians who would otherwise have been sampled in such large numbers that they would have dominated the overall boost sample, and hence undermined attainment of the target sizes for the other target groups. Table 3.1a shows the selection fractions by ethnic groups.

Table 3.1a: Secondary screening retention fractions

Main category	Sub-group	Fraction for assignments issued	
		2009	2010
Indian	Indian (non-mixed)	50%	65%
	Mixed Indian	100%	100%
Pakistani	Pakistani	54%	100%
Bangladeshi	Bangladeshi	100%	100%
Caribbean	Caribbean (non-mixed)	91%	100%
	Mixed Caribbean	100%	100%
African	African	82%	100%
Other included groups	Sri Lankan	100%	100%
	Chinese	100%	100%
	Turkish	100%	100%
	Far eastern	30%	100%

Middle eastern	30%	100%
Source: Berthoud et al 2009		

The distribution of ethnic groups reported by households in response to the “screener” is shown in Table 3.1b. As some households include members from different ethnic groups, these numbers will add up to more than all the households in the EMB sample (screened in). The overall sampling fractions combine a) the probability of sampling the sector, b) the fraction of addresses selected within the sector, and c) the probability of a household being retained following the application of the random selection mechanism described above.

Table 3.1b: Number of households reporting at least one household member (or their parents or grandparents) come from different ethnic group^a

Indian	2303	North African	187
Mixed Indian	234	Black African	1730
Pakistani	1645	African Asian	93
Bangladeshi	955	Chinese	451
Sri Lankan	265	Far Eastern	402
Caribbean/West Indian	1785	Turkish	224
Mixed Caribbean/West Indian	491	Middle Eastern Iranian	463

^a Note some households reported members from more than one ethnic group and so these numbers will not add up to all the households in the EMB sample (screened in)

3.3 The Immigrant and Ethnic Minority Boost Sample

The IEMB was intended to refresh the EMB sample by providing additional respondents from the five target ethnic minority groups in the EMB (Indian, Pakistan, Bangladesh, Caribbean, and African) and additionally to provide a sample of immigrants (i.e., non-UK born) from groups other than these five ethnic minority groups. The target was to achieve around 2,500 respondents from the five target ethnic minority groups and 2000 immigrants, with some overlap across these two target populations. Going forward the IEMB enables UKHLS to continue to offer a representative sample of the five ethnic minority target groups with sufficient large numbers for distinct analysis, as well as a representative sample of immigrants with some additional dedicated content and sufficient numbers for analysis of the UK’s immigrant population.

Publicly-available data from the 2011 UK Population Census at LSOA and SDZ level on ethnic group and country of birth was used to define the sampling strata. As with the EMB the design and the construction of the strata was intended to optimise the trade-off between sample size and statistical efficiency for the available budget. That is, skewing the sample towards higher density areas increases the cost-effectiveness of obtaining a sample of a given size but with a loss of statistical efficiency. The design was complicated not only by the fact that the design was intended to reach the five target ethnic minority groups but that it was also intended to reach immigrants from other groups. Hence the final design involved five strata. It was necessary to create a first stratum containing relatively high concentrations of black Africans because the ratio of desired sample size to population size

was higher for black Africans than for any of the other target groups and because this group is less concentrated geographically than the other target groups. A second stratum consisted of all other areas with a relatively high concentration of UK-born ethnic minorities from the five target groups, while a third stratum contained areas with high concentrations of foreign-born who were not from the five target groups. It was necessary to separate out such areas as the requirement to over-sample immigrants would otherwise have led to an inefficiently large proportion of the ethnic minority sample being immigrants and of the immigrant sample being from the target ethnic minority groups. A fourth stratum consisted of all other areas with high concentrations of either the five target groups (regardless of whether or not they were UK-born) or immigrants (regardless of whether or not they were from the five target groups). All other areas were consigned to a fifth stratum which would not be sampled for the boost sample. This fifth (non-sampled) stratum covered two-thirds of the overall UK population but only 26 per cent of non-UK born and only between 9-16 per cent of the populations of the five target ethnic minority groups. See further Lynn et al. 2017 for more details on the design and fieldwork for the IEMB.

The design is therefore efficient as fieldwork could be restricted to just one-third of the territory, while still covering a high proportion of each of the sub-populations of interest. While it did not provide full coverage of ethnic minorities and, more so, immigrants living in less dense areas, these are included in the design of the general population sample, and hence it is possible to carry out population analysis of all immigrants or of all ethnic minorities in UKHLS, using the appropriate weights. (See further information on weights, below). This is also why it is not possible to analyse the IEMB (or the EMB) independently – and therefore no weights are supplied that would enable such analysis.

As with the EMB, the IEMB involved a two-stage process of screening, followed by interviewing of those households identified as eligible by the screen. The screening questionnaire was administered using a screening card presented to respondents on the doorstep, and which included relevant answer categories for inclusion. The first screening question was designed to identify immigrants (not born in the UK), while the second one asked about ethnic minority background. If the respondent answered ‘yes’ to either of the two screening questions, the households was deemed to be eligible. The screening question is provided in [Appendix I](#). The screening card was only available in English, although a translation card to identify the language spoken was used in households where none of the members spoke English. In such cases, a bilingual interviewer or an agency translator could be sent to that address to conduct screening and interviewing. Unlike with the EMB there were no secondary screening fractions: those households identified at the point of screening were all eligible for interview.

Further information on the design and implementation of the IEMB and how it maintains the representativeness of the study’s minority and immigrant populations going forward can be found in Lynn et al. (2017). In Wave 6, the achieved sample sizes for the IEMB were:

- 2,922 responding households
- 8,517 individuals (of all ages) enumerated as living in these 2,922 households
- 4,656 adult interviews (includes 198 proxy interviews)

Please note, the youth questionnaire was not administered to the IEMB sample in Wave 6.

3.4 Fieldwork, survey instruments, interview mode

Here is a summary of the fieldwork process, interview mode and survey instruments to give an overview of how information is collected. For further details on interview mode, fieldwork procedures, incentives and consent and changes over time see the Main User Guide section on [fieldwork procedures](#) and [interview process](#).

Interview mode

Since Wave 1 interviews are primarily conducted face-to-face via computer aided personal interview (CAPI). As discussed earlier the surviving BHPS sample were incorporated into *Understanding Society* from Wave 2 onwards. Around 500 BHPS households had already opted for telephone interviews, and they continue to be interviewed by telephone in *Understanding Society*. From Wave 3 onwards, at the end of the fieldwork period for each monthly/quarterly sample, those who had not responded were contacted via telephone and offered the option of completing the interviews by telephone. There is no difference in the content or format of questions asked except they are not administered the self-completion questionnaire. In Wave 7, web interviews were introduced for the first time and offered to a substantial proportion of the sample from Wave 8 onwards. There is no difference in the face-to-face and web questionnaires. For further details about the mode experiment, mixed modes design and analysis guidance see the Main Survey User Guide Sections on [Study design](#), the [Interview Process](#) and [Analysis advice for mixed modes data](#).

Due to the COVID-19 pandemic, all face-to-face interviews was suspended from April 2020 and the sample members were invited to take part online or by telephone. Face-to-face interviews were not fully re-instated until April 2022.

Fieldwork process and survey instruments

Once households are selected, interviewers attempt to ask an adult in the household the names of everyone who is part of the household and some basic information (such as date of birth or age, sex and marital status) about them. From Wave 2 onwards, interviewers find out who has left the household (and why), who has joined the household and whether any of the basic information has changed. This information is recorded in the **household coversheet (comprises enumeration and household grid)**. To find out who is eligible for an interview each year see the discussion on following rules in section 3.5.

Next, interviewers attempt to conduct a household interview (using the **household questionnaire**) with an adult who is most likely to be informed about the household (generally this is the person who owns or rents the property and, if more than one, the

eldest). The household questionnaire includes questions on ownership of the property, rent and mortgage information, expenditures, fuel consumption, ownership of consumer durables etc.

Interviewers attempt to interview every adult (defined as 16+ year old) household member and ask them questions from the **main adult individual questionnaire**. If a household is part of the extra five minutes sample then the adult members in these households are also asked the extra five minutes questions.

These individuals who complete the adult individual questionnaire are also asked to complete a **self-completion questionnaire**, i.e., questions that they answer by themselves (without interviewer involvement). This questionnaire generally consists of sensitive questions that the respondent may not be comfortable telling the interviewer. In Waves 1 and 2, interviewers handed over the paper questionnaire to the respondent. From Wave 3 onwards, this was conducted using computers (Computer Assisted Self Interview, CASI) and so the interviewers turned round the computer to the respondents.

Sometimes a person who is not present gives permission for their information to be collected on their behalf from someone else, usually their spouse or adult children. The person being interviewed is referred to as the “proxy informant” and the person on whose behalf they are being interviewed is referred to as the “proxy respondent”. Note this questionnaire is shorter and comprises only factual questions. This is known as the **proxy questionnaire**. The interview outcome variable can be used to identify proxy respondents. A variable identifying the proxy informant and their relationship to the proxy respondent is also made available.

Young persons between the ages of 10 and 15 years are also asked to complete a paper self-completion questionnaire (**youth questionnaire**).

Information about children younger than 10 years is asked of their parents or guardians and included in the main adult individual questionnaire.

All samples (GPS, EMBS, IEMBS and BHPS) receive the same questionnaires, but some additional questions are asked of a sub-sample referred to as the extra five minutes sample (see Section 4 for further details).

As the IEMB sample was interviewed during year 2 of the 6th wave of the main survey, IEMB respondents received the Wave 6 questionnaire with some differences. Some new questions were included in the IEMB such as migration intentions, reasons for migration, their first job when they came to the UK, their parents’ first job when they came to the UK and so on (see Section 4.3). Due to time constraints some questions from the Wave 6 questionnaire were excluded such as the Life Satisfaction question. The youth questionnaire was also excluded. This had implications for the weights provided in Wave 6. For a detailed discussion see Section 6.2. Note that from Wave 7 onwards, IEMB sample members receive

the same questions as all other samples. Additionally, they also receive the extra five minutes questions. See discussion in Section 4.

Interviewer characteristics

In addition to the information collected directly from respondents, information about the interview process is also collected and reported by the interviewer such as whether the respondent was cooperative. Information about the interviewers is provided by the fieldwork agency. This information is recorded in a separate data file, **xivdata**, which is available in a separate dataset that can be accessed under Special Licence [from the UK Data Service \(SN 8579\)](#). Details include information on their year of birth, sex, the year they started working for the fieldwork agency and their ethnic group. The data file can be merged onto **w_callrec**, **w_hhresp** and **w_hhsamp** using the variable **w_intnum**.

Implications of the 12- and 24-month fieldwork design

In the last section we discussed how some samples are interviewed over a 24-month fieldwork period while others over a 12-month period. Does this mean some people are interviewed every year while others are interviewed every two years? **NO**. Even if some samples are interviewed over two years while others over one year, the gap between two interviews for each household is *always* approximately one year. So, when measuring change across waves, for every household, the change is measured over the same period. For example, if a household was first interviewed in January 2009, this household (and all the individuals within that household) will be interviewed, as part of its second wave interview, around January 2010 (see Figure 3.1). Suppose another household was interviewed for the first time in January 2010, that household's second wave interview will be around January 2011. When using the data, unless the exact year of interview is important to the research question this issue of overlapping fieldwork period can be ignored. Variables identifying the interview date as well as the monthly sample each household is assigned to are made available to users.

Person ID	Interview date in Wave 1	Interview date in Wave 2	Wages in Wave 1	Wages in Wave 2	Annual change
1001	24 January 2009	1 February 2010	2500	2600	100
1002	20 January 2010	9 January 2011	4000	4400	400

Figure 3.1 Measuring annual change

See [Appendix VII](#) for an illustration of the overlapping fieldwork period across different samples.

3.5 Sample status and following rules

Following rules determine whether specific household members should be asked to be interviewed in the next wave. The purpose of these following rules is to make sure that the study continues to be representative of the original sample selected in 2009/10 and their

descendants and information on the household context is always collected. For ease of implementation of these following rules, all household members are classified into three sample statuses: Original Sample Members (OSMs), Temporary Sample Members (TSMs), and Permanent Sample members (PSMs).

OSMs and PSMs, of all ages, are followed and remain eligible for interview as long as they are resident within the UK, potentially for the duration of their life or the life of the survey. TSMs remain eligible for interview as long as they are co-resident with an OSM/PSM (including when the only OSM in the household is a child who is not yet eligible for personal interview). TSMs who are not co-resident with an OSM/PSM are not followed and become ineligible for interview.

OSM: All members of *Understanding Society* General Population Sample households enumerated at Wave 1, including absent household members and those living in institutions who would otherwise be resident, are OSMs. All **ethnic minority members** of an enumerated household eligible for inclusion in the Ethnic Minority Boost sample are OSMs. In all of these samples, any child born to an OSM mother after Wave 1 and observed to be co-resident with the mother at the survey wave following the child's birth is an OSM. In Wave 6, all non-UK born TSMs became OSMs.

TSM: Anyone who becomes co-resident with a OSM from Wave 2 onwards is considered to be a TSM. This would include any child born to an OSM father but not an OSM mother after Wave 1 and observed to be co-resident with the father (or any other OSM) at the survey wave following the child's birth. Members of an enumerated household included in the Ethnic Minority Boost sample at Wave 1 and the Immigrant and Ethnic Minority Boost sample at Wave 6 who are not from a qualifying **ethnic minority group** or born outside the UK in case of the IEMB are also considered to be TSMs. These were the only cases where members of sampled households at the time of sampling (Wave 1 for EMB and Wave 6 for IEMB) were considered to be TSMs.

PSM: Any TSM father of an OSM child born after Wave 1 and observed to be co-resident with the child at the survey wave following the child's birth is considered to be a PSM.

4. Specific content

4.1 The extra five minutes sample

The Extra Five Minute Sample originally comprised three components: the Ethnic Minority Boost (EMB) sample, the General Population Comparison (GPC) sample, the Ethnic minority in Low Density Area (EM-LDA) sample. In addition, many of the questions, in particular those where responses might be expected to change most quickly among more recent immigrants, are asked of all recent immigrants, that is those who immigrated within the three years prior to the beginning of Wave 1 of *Understanding Society*, even if they were not in the EMB.

From Wave 7, the IEMB was incorporated in the Extra Five Minutes sample alongside all other immigrants not already covered in any of three other extra five minutes samples (sometimes referred to as the FBORN sample).

The Ethnic Minority Boost (EMB) sample

The EMB sample includes sufficient sample sizes of different ethnic groups to allow analysis separately by ethnic group. For a detailed discussion see section 3.2.

The General Population Comparison (GPC) sample

To allow comparative analysis with the White majority (defined as those who choose White – British/English/Scottish/Welsh/Northern Irish as their ethnic group) group the Extra Five Minutes questions were also asked of a random sub-sample of the GP Sample, referred to as the General Population Comparison (GPC) sample. The GPC includes one sampled address from 40 per cent of the selected postal sectors in the Great Britain component GPS. In other words, of the 2,640 general population sectors, 60 per cent of them (1,584) contain 18 GPS addresses and the other 40 per cent contain 17 GPS addresses and one GPC address. The persons in these households were designated as members of the GPC sample, *regardless of ethnic group membership*.

The Ethnic Minority in Low Ethnic Minority Density Area (EM-LDA) sample

Existing studies have shown impact of the composition and quality of residential neighbourhoods on various life outcomes. The EMB sample was drawn from high ethnic minority concentration areas and, as a result, any analysis based on the EMB and GPC samples would not be representative of the experience of all ethnic minorities. It was therefore decided to include ethnic minorities from the GP sample living in areas that would not have been eligible for inclusion in the EMB screening areas on the basis that they did not have sufficient concentrations of the target ethnic minority groups. These areas are referred to as 'low density areas'; and hence this sample is referred to as EM-LDA sample. This status became fixed at the individual level after Wave 1.

All OSM members of EMB and GPC samples and anyone co-resident with them are eligible for the extra five minute questions at each wave. At Wave 1, only the EM-LDA sample members were eligible for the extra five minute questions but from wave 2 onwards anyone co-resident with them were also eligible. From Wave 7 all OSM members of the IEMB and all OSMs who are not UK-born and not already covered in one of the other samples are eligible for the extra five minutes.

The Foreign born (FBORN) sample

From Wave 7 onwards, all non-UK born individuals who were enumerated by Wave 6 (and their status was changed to OSM, see Section 3.5) were also eligible for the extra five minutes questions.

4.2 Extra five minutes questions

Given the importance of ethnicity to *Understanding Society* as a whole, and the potential for addressing specific research interests in a way that had not been possible since the Fourth National Survey of Ethnic Minorities in 1993/4, it was determined from the outset to collect an extra five minutes' worth of questions for specific topics of interest for ethnicity research. Of particular priority were those areas which would enable longitudinal research on pertinent topics. However, extra questions add to respondent burden and overall survey costs and specialist questions could not be fielded across the whole sample, hence it was decided that these would be asked of a sub-sample which would allow comparative analysis. We refer to this as the "Extra Five Minute" sample, which is described further below.

The topics to be carried in to the "Extra Five Minutes" sample, and their rotation over waves of the study was subject to extensive consultation with the ethnicity researchers, co-funders, third sector and the Ethnicity Strand Advisory Committee. Further consultation was held in advance of the IEMB refreshment sample. As a result of this consultation, the broad areas for inclusion and their rotation were determined as illustrated in Table 4.1a. There are also some questions asked on a different (more frequent) rotation for the Extra Five Minutes sample than for the general population sample or questions that commence at a different starting point due to particular interests or topics (See Table 4.1b).

In Wave 6, the IEMB introduced some new questions for non-UK born respondents that were specifically relevant to immigration research (See Table 4.1a). Most of this content was asked of all other immigrants at Waves 7 & 9. See Section 4.3 for further details on these questions.

For a complete list of extra five minutes question variables, see [Appendix II](#). Please also see the current [long term content plan](#) which includes a list of all the questionnaire modules and the frequencies in which they are planned to be asked (extra five minutes question modules are also indicated).

Table 4.1a: Extra Five Minute Questions asked of the Extra Five Minute sample only

Module	Rotation	Wave
Migration history	Initial conditions	1, 6, 7 ^a
Harassment	2 year rotation	1, 3, 5, 7 ^b , 9
Discrimination	2 year rotation	1, 3, 5, 7 ^b
Remittances	3 year rotation	1, 4, 7 ^b
Ethnic identity	6 year rotation	2, 8, 14 ^b
Ethnic identity (only for new immigrants and 16-19 year olds)	3 year rotation	2, 5, 8, 11
Best Friends (part of the module)	3 year rotation	3, 6
Service Use	4 year rotation	4, 6, 10, 14 ^c
Religious practice	6 year rotation	4, 10
Cultural participation	-	5
Migration Reasons	Initial conditions	6, 7 ^d
Migration Intentions	-	6, 7, 9 ^e
First job in the UK	Initial conditions	6, 9 ^f
Additional family background information ^g	Initial conditions	6, 9 ^f
Educational qualification gained abroad	Initial conditions	6, 9 ^f
British citizenship intentions	Initial conditions	6, 9 ^f

^a These questions were asked of the EMB, LDA, GPC in Wave 1, these were asked of the IEMB in Wave 6 and were asked of the FBORN sample in Wave 7.

^b These questions were asked of the EMB, LDA, GPC until Wave 7, these were asked of the IEMB & FBORN samples from onwards Wave 7.

^c These questions were on a two year rotation (Waves 4 and 6) and then on a 4-year rotation (Waves 6, 10, 14)

^d These questions were asked of the IEMB in Wave 6, were asked of the EMB, GPC, LDA, FBORN samples in Wave 7.

^e These questions were asked of the IEMB in Wave 6, were asked of the EMB, GPC, LDA, FBORN samples in Wave 7 and as the routing did not work correctly also in Wave 9.

^f These questions were asked of the IEMB in Wave 6, were asked of the EMB, GPC, LDA, FBORN samples in Wave 9.

^g These include questions on parents' current country of residence and first and current UK job related questions

Table 4.1b: Modules asked of the whole sample and the Extra Five Minutes sample but on a different rotation

Module	Sample asked	Rotation	Wave
Religious belonging (oprlg)	Whole sample		1, 4, 8, 12
	New entrants of only the Extra Five Minutes sample & Northern Ireland residents	-	2, 3, 5, 6, 7, 9, 10, 11
Political engagement	Extra Five Minutes sample	-	2
	Non-Extra Five Minutes sample	-	3
	All	4-year rotation	6 (not IEMB)
Britishness	Extra Five Minutes sample	-	1,3
	All (self-completion)	-	6 (not IEMB)

Notes: Other ethnicity related questions which were asked of the whole sample, such as the questions on ethnic group, childhood language, national identity, are not discussed here as they are not part of the extra five minutes questions. For these questions see [Appendix V](#)

How to identify the extra five minutes questions in the questionnaire?

In the [questionnaires](#) up to and including Wave 6, the Extra Five Minutes questions can be identified by searching for the description of the “Universe” (refers to those who are eligible for a particular question), as follows:

If (HHGRID.EMBoost = 1 | HHGRID.GPCompare = 1 | (HHGRID.LDA = 1 & ETHNICITYANDNATIONALIDENTITY.RACEL > 4 & ETHNICITYANDNATIONALIDENTITY.RACEL < 98)) // Ethnic Minority Boost or General population comparison sample or LDA and any non-white background

From Wave 7 onwards, this text also includes “GRIDVARIABLES.FBORN = 1” and “foreign-born continuing respondents and their co-resident sample members” to correspond with the extended definition of extra five minutes sample which also includes non-UK born (by Wave 6) and their household members who are not in EMB or IEMB.

The text identifies the different components of the overall sample who receive the questions as we go on to discuss next.

4.3 New IEMB questions

In Wave 6, some additional new questions were asked of the non-UK born in the IEMB sample. See Table 4.1a, above, and [Appendix III](#). The “reasons to migrate” questions were asked of all other non-UK born respondents in Wave 7 and can therefore be analysed for all immigrants from Wave 7 onwards. The “intentions to migrate” questions were included in Wave 7 as well but due to question routing error this module was not asked of a substantial proportion of those who were eligible for the module. The “intentions to migrate” questions and rest of the IEMB-specific questions were asked of all other non-UK born sample members at Wave 9. Since, as discussed below, the immigrants in the IEMB sample should not be analysed separately and all immigrants across all samples should be analysed together, these questions should not be analysed in Wave 6 or Wave 7. These should be analysed after combining information collected in Wave 9 for the rest of the sample. The syntax to combine this information from Waves 6, 7 & 9 to produce derived variables (mintent_dv mintwhen_dv mreason?_dv) are provided in [Appendix VI](#). If you have any suggestions or questions about these derived variables please let us know (usersupport@understandingsociety.ac.uk).

The IEMB sample was also asked the Wave 1 Migration history module, which was asked of all immigrants at Wave 1. This information can therefore be analysed across all immigrants in the study, whether included via the IEMB or any of the other samples, using the appropriate weights.

Educational qualification questions were asked differently in the IEMB. See Section 6.4 for further details.

5. Response and attrition

General information on non-response and attrition including up-to-date response tables is available in the [Main User Guide section on Response Rates](#).

The Wave 1 survey fieldwork started on 8th January 2009 and ended on the 7th March 2011 (including the re-issue period). In total, interviews were achieved in 30,169 households (26,089 in the General Population Sample, 4,080 in the Ethnic Minority Boost sample), full or proxy interviews with 50,994 individuals (43,674 in the General Population Sample and 7,320 in the Ethnic Minority Boost sample).

Table 5.1 below shows the household response outcomes including the number of households that were eligible, ineligible or of unknown eligibility at the initial wave of sample selection by Understanding Society sample. The EMB and IEMB samples included a screening stage and if there was non-contact or refusal at the screening stage then the eligibility of these households were not known. See Lynn et al. 2012 and Lynn et al. 2017 for details.

Table 5.1: Household interview outcomes at initial interview (Wave 1 for GPS, EMB and Wave 6 for IEMB)

	General Population Sample			Ethnic Minority Boost Sample	Immigrant and Ethnic Minority Boost Sample
	Great Britain	Northern Ireland	Total		
<i>Productive</i>	24,775	1,291	26,066	4,067	2,922
<i>Refusal</i>	15,400	663	16,063	2,966	1,792
<i>Non-contact</i>	2,061	117	2,178	1,117	705
<i>Other unproductive</i>	774	9	783	798	181
Screened in and eligible	43,010	2,080	45,090	8,948	5,600
Ineligible	4,473	278	4,751	34,293	10,307
Unknown eligibility and unproductive at screening	661	37	698	1,528	3,552
Total Issued	48,144	2,395	50,539	44,769	19,459

Table 5.2 shows individual response rates (among responding households) for the different samples. Individual response rates (including proxy interviews) for the GP-GB sample is higher than that for the GP-NI and EMB samples which are similar at 80%. Refusal and non-contact rates for these two samples are also similar. Response is lower within the EMB and IEMB than across the sample as a whole, both at Wave 1 and over time (see Lynn et al 2012; Lynn et al. 2017).

Table 5.2: Individual response rate^a at initial interview

	General population sample			Ethnic minority boost sample	Immigrant & ethnic minority boost sample
	Great Britain	Northern Ireland	Total		
Adult individual interview rate	82.0%	77.3%	81.8%	72.4%	71.2%
+ proxy interview rate	87.3%	80.8%	87.0%	79.24%	74.3%
Refusal rate	8.0%	11.0%	8.2%	12.2%	16.4%
Non-contact rate	4.7%	8.2%	4.8%	8.6%	9.2%
All enumerated adults	47,615	2,584	50,199	9,237	6,260
Youth response rate	77.2%	73.1%	77.0%	62.9%	N/A ^b
Youth refusal rate	22.8%	26.9%	23.0%	37.1%	
All enumerated youths	4,900	290	5,190	1,437	753
All enumerated children	8,082	477	8,559	2,687	1474
All enumerated individuals	60,597	3,351	63,948	13,361	8,517

^a Among responding households and includes OSM and TSMs. ^b Youth questionnaire was not included in IEMB

Table 5.3 shows the number of adult interviews achieved in Wave 1 by ethnic group and sample. This excludes proxy respondents.

Table 5.3: Wave 1 adult (16+ year) interviews by ethnic group and sample^a

	Total	GPS	EMB	X5min Sample	Born in UK	Not born In UK
Asian/Asian British: Indian	1,910	825	1085	1,219	572	1337
Asian/Asian British: Pakistani	1,441	498	943	981	546	895
Asian/Asian British: Bangladeshi	1,130	177	953	971	318	812
Black/African/Caribbean: Caribbean	1,149	354	795	832	555	594
Black/African/Caribbean: African	1,459	500	959	1024	193	1264
Five target ethnic groups	7,089	2,354	4,735	5,027	2,184	4,902
Mixed ^a	874	422	452	575	565	309
Asian/Asian British: Chinese	322	130	192	242	49	273
Other ethnic group: Arab	282	145	137	178	27	255
White: Irish	718	696	22	25	458	259
White: Any other white background	1,470	1,336	134	217	142	1328
Other ethnic groups ^b	1,004	517	487	622	194	809
All ethnic minority groups	4,670	3,246	1,424	1,859	1,435	3,233
white: British/ English/ Scottish/Welsh/ Northern Irish	35,920	35,404	516	1,265	35,041	876
Total^c	47,732	41,047	6,685	8,161	38,696	9,024

^a Includes "Mixed: white and black Caribbean", "Mixed: white and black African", "Mixed: white and Asian", "Mixed: any other mixed background"; ^b Includes those who reported their ethnic group as "White: gypsy or Irish Traveller", "Asian/Asian British: Any other Asian Background", "black/African/Caribbean: Any other black background", "Other ethnic group"; ^c The discrepancy in the totals is due to small number of cases with missing information.

Table 5.4 shows the numbers of enumerated adults and adult respondents by ethnic group background and immigrant status in the IEMB sample.

Table 5.4: Adult interviews and enumerated individuals in the IEMB sample at Wave 6 by ethnic group and by whether or not UK born

Enumerated Adults		Adult respondents	
Country of birth		Country of birth	
UK	1,983	UK	1,437
Outside UK	4,197	Outside UK	3,219
Missing information	80	Missing information	
Ethnic group ^a		Ethnic group ^a	
Indian	1,007	Indian	727
Pakistani	914	Pakistani	664
Bangladeshi	297	Bangladeshi	212
Caribbean/West Indian	470	Caribbean/West Indian	374
Black African	853	Black African	653
Other ethnic minority groups	1,753	Other ethnic minority groups	1,564
None of the above	890	None of the above	460
Missing information	76	Missing information	2
Total	6,260		4,656

^aUses the self-reported ethnic group question and if that is missing then uses screening ethnic group question included in the household grid

Finally, Tables 5.5a-5.5b show the response outcomes among Wave 1 adult interview respondents across the first four waves of the study by ethnic group, and religion. Cell sizes smaller than 50 are not shown to avoid disclosure. These tables inform users of the expected sample sizes for longitudinal analyses on these dimensions and illustrate that patterns of response may be non-monotonic i.e., that those who are observed at Waves 1 and 4 may not always also have been observed at Waves 2 and 3.

Table 5.5a: Response outcomes among Wave 1 adult interview respondents across four waves by ethnic group

1: full interview, 0: non-interview including proxy interview; Position indicates wave number. For example, 1110 indicates full-adult interview in Waves 1, 2 and 3 but not in Wave 4

	1000	1001	1010	1011	1100	1101	1110	1111
white Majority ^a	6471	215	384	985	3909	925	2819	20212
white Irish	139	*	*	*	94	*	78	358
Any other white background	440	*	*	*	198	*	109	592
Mixed	222	*	*	*	94	*	70	395
Indian	523	*	*	80	248	71	150	775
Pakistani	371	*	*	70	171	71	148	550
Bangladeshi	356	*	*	57	151	64	96	341
Chinese	115	*	*	*	52	*	*	111
Caribbean	302	*	*	*	131	*	107	497
African	446	*	*	61	236	*	142	469
Arab	94	*	*	*	*	*	*	87
Other ethnic groups ^b	293	*	*	*	132	*	85	396

^a White British/ English/ Scottish/ Welsh/ Northern Irish; ^bIncludes those who reported their ethnic group as Gypsy or Irish Traveller, Other Asian Background, Other Black Background, Any Other Ethnic group; *Denotes number of observations<50

Table 5.5b: Response outcomes among Wave 1 adult interview respondents across four waves by religion (brought up in or belongs to)

1: full interview, 0: non-interview including proxy interview; Position indicates wave number. For example, 1110 indicates full-adult interview in Waves 1, 2 and 3 but not in in Wave 4

	1000	1001	1010	1011	1100	1101	1110	1111
None	1602	65	115	228	864	208	599	3322
Catholic	1448	51	89	187	771	186	582	3420
Protestant	4744	140	272	754	2924	713	2057	15262
Jewish	50	*	*	*	*	*	*	93
Muslim	1173	71	115	176	537	178	374	1373
Hindu	333	*	*	*	146	*	98	453
Sikh	146	*	*	*	65	*	*	216
Buddhist	87	*	*	*	*	*	*	124
Other	177	*	*	*	92	*	61	514

*Denotes number of observations<50

6. Data files and question coverage

6.1 Data structure, naming conventions, key variables

We recommend looking at the [Main Survey User Guide Data section](#) for discussion of data structure, filename and variable naming convention, missing values etc. Here we give a brief overview of the main elements and some key variables.

As Understanding Society is a household panel survey, providing all the data in a single flat file is not very efficient. Instead different sets of files are provided for each wave. At each wave, data from adult individual interviews are stored in one file, while responses from household questionnaire are stored in another file. Details of all files provided at each wave are shown in Table 6.1.

Within each wave individual and household level files can be linked by the unique wave specific household identifier: **w_hidp**. Within each wave different individual level files can be linked by the unique wave specific individual identifier: **w_hidp w_pno** (both need to be used together) OR the unique cross-wave individual identifier: **pidp**. Across different waves, individual level files can be linked by using the unique cross-wave individual identifier: **pidp**. Households cannot be linked across waves, as there is no universally accepted definition of a longitudinal household. But household information from each wave can be linked to individuals who can be matched across waves.

The naming convention followed in *Understanding Society* is that all wave-specific variables and data files have a wave specific prefix: **a_** signifies wave one, **b_** signifies wave two and so on. While most variables and data files are wave specific there are some which are not. Such variables and data files do not have the wave prefix. For example, the cross-wave unique identifier, **pidp** has no wave prefix, but the main activity status variable is named **a_jbstat** at Wave 1, and **b_jbstat** at Wave 2 and so on. Similarly, the data file that stores information from adult interviews in Wave 1 is named **a_indresp**, while the data file that stores time-invariant variables such as country of birth, ethnic group, sex, collected from across different waves is named **xwavedat**, that is, without any wave prefix. Also note the suffix **_dv** signifies the variable was derived by data managers.

Table 6.1: Some of the key data file names and their description

File name	File description
w_indresp	Data from adult interviews based on individual questionnaire including proxy, telephone and self-completion questionnaires
w_indall	Enumeration and household grid information on all household members including children and non-responding adults
w_hhresp	Data from household interviews based on household questionnaire
xwavedat	Basic time-invariant variables. This is a very useful file. It collects fixed information such as ethnic group, country of birth which are only asked the first time a person is interviewed from the wave in which it was asked for that person and puts the information together in a flat file.
w_youth	Data from youth interviews based on youth self-completion questionnaires

For those carrying out analysis of those answering the Extra Five Minutes questions, the Extra Five Minutes sample can be identified by the flag **w_xtra5min_dv** and those who are OSMs in this sample are identified by **w_xtra5minosm_dv**. We discuss the related issues of weighting for the different samples next.

The EMB, IEMB, BHPS and the GP samples can be distinguished using the flag **w_hhorig** (in wave specific files) or **hhorig** (in cross-wave files such as the **xwavedat**). However, analysis should NOT be carried out separately on the EMB or IEMB, since the weights are not estimated for that sample specifically. See section 7 for a discussion on the weights available and the samples for which they have been constructed.

We provide a [list of key variables](#) such as unique person identifiers, relationship pointers, age, sex, region of residence, marital status, main activity status, and so on.

In Understanding Society data, missing values are given negative values instead of leaving these as system missing (.). These missing values provide a lot of useful information.

Item non-response: Don't know responses are given a value of -1, while refusals are given a value of -2 and when the data is missing for coding error or interviewer error or fieldwork error it is coded as -9.

Valid missing values: Not everyone gets asked all questions. For example, those without a job are not asked about their pay. As these missing values are valid, these cases are given a different value of -8. As the proxy questionnaire is shorter than the adult questionnaire administered to full respondents, the value of questions not included in the proxy questionnaires is -7 for proxy respondents.

Since the IEMB questionnaire different slightly from the questionnaire administered to the main wave 6 sample, questions included in IEMB but not in the questionnaire administered to the non-IEMB samples (e.g. migration motivations), are given a value of -11 for sample members/households belonging to non-IEMB samples. In the opposite case (e.g. questions that could not be accommodated in the IEMB questionnaire such as life satisfaction), IEMB sample members get a value of -10.

In the 2017 data release, 18 waves of BHPS data were harmonised with UKHLS data. There was no change to the UKHLS variable or file names. BHPS files and variables were harmonised to match UKHLS variables and file names. All BHPS files and variable names now have the same naming convention as Understanding Society files and BHPS variables considered to be harmonisable with UKHLS variables have the same name, with certain exceptions: (i) the wave prefix is **bw_** instead of **w_** for **BHPS waves a-r** (ii) BHPS variables that are not harmonisable with UKHLS but have the same original name were renamed with a suffix **_bh**. For further details see [BHPS User Manual](#) and the [BHPS harmonised user guide](#) (Fumagali et al. 2017).

6.2 IEMB vs GPS-EMB-BHPS sample Questionnaires in Wave 6

The questionnaire for the IEMB sample that was fielded during the second half of the 24-month fieldwork period for Wave 6, was very similar to the Wave 6 questionnaire for the rest of the samples with some exceptions:

The Youth questionnaire was not included in the IEMB.

A number of new immigration-related questions were included, as noted above, and some other questions included in the non-IEMBS Wave 6 questionnaire were excluded. In addition, as this was the first wave for the IEMB sample, initial partnership and fertility histories were asked: **w_natchild**, **w_adopt**, **w_cohab**, **w_marriage**. However, employment history was not collected. An accompanying file (**6614_wave6_iemb_question_comparison.xlsx**) to the data provides a full comparison of questions in the IEMB compared to the standard Wave 6 questionnaire.

Using information on job description, occupation is coded for that job based on SOC 1990, SOC 2000 and SOC 2010 coding frames. The resulting variables generally have similar names with suffixes soc90, soc00, soc10. The original open ended text variable is not released. However, for some of the new job variables in IEMB, such as first job in UK, last job in country of birth and so on the resulting three occupation variables were given names that did not have the same stem names. This may make it difficult to find and locate the original job description question which was used to derive these occupation variables. So, in [Appendix IV](#) we provide a table which shows the name of the original job description question (as in the IEMB questionnaire) and the three occupation variables computed from it and released with the data.

6.3 Key variables for ethnicity and immigration research

As noted above, researchers bring to the study of ethnicity different conceptual and operational frameworks for their investigations. Some focus on immigrant status (or country of birth), others are concerned with self-ascribed ethnicity, or in the construction of 'ethno-religious groups', and others are interested in ethnic identity as a subject of research in its own right. In *Understanding Society*, the question content aimed to facilitate maximum flexibility for researchers in implementing their own understandings of ethnicity and ethnic group. These concepts are discussed in Section 2. Here we signpost to the specific variables. Unless specified, all questions refer to questions asked of adult respondents in the adult interview questionnaire.

Specific variables for ethnicity and migration research are included in the extra five minutes questions (See Tables 4.1a, 4.1b and [Appendix II](#)), additional questions asked of immigrants in Waves 6, 7 & 9 (See [Appendix III](#)) or related to measurement of ethnicity (see [Appendix V](#)).

In [Appendix V](#), we provide a list of variables to measure ethnicity and for each variable we provide brief description of the question, the variable names, the (sub-)sample who were asked those questions and the wave (first) asked in. Most questions were asked either in Wave 1 or at the first time the respondent was interviewed. However, additional questions on ethnic identity were introduced in Wave 2 and are repeated every three years for new immigrants and 16-19 year olds and every six years for everyone else. The table also highlights if the question was part of the extra 5 minutes module.

As mentioned earlier, the [long term content plan](#) includes a list of all the questionnaire modules and the frequencies in which they are planned to be asked and extra five minutes question modules are also indicated.

Ethnic group

When adults are interviewed for the first time, they are asked the ONS Census ethnic group question. As this has changed across Censuses, the BHPS sample respondents were asked the version **bw_race** from 1991, then **bw_racel_bh** from 2003 onwards, and UKHLS respondents were asked the version **w_racel** & **w_racelo_code** from the outset. Those who were interviewed by telephone were asked a branching version of these questions (**w_racelt**, **w_racelwt**, **w_racelmt**, **w_racelat**, **w_racelbt**, **w_racelot_code**). These reports across the waves have been combined and this combined version, **racel_dv** is available in the cross-wave file **xwavedat**, and as **w_racel_dv** in **w_indresp** files.

Additionally, in Wave 1, everyone was asked to report the ethnic group of their mother and father (**w_maid**, **w_paid**), and in Wave 6, in the IEMB sample, the person completing the household grid was asked about the ethnic group of everyone in the household & that of their parents or grandparents (**w_indethg***, **relethg***). A question about ethnic group is also included in the youth questionnaire (administered to 10-15 year olds) in every other wave starting with Wave 1 (**w_yprace**).

The ethnic group information as reported via different sources (**racel_dv**, **w_ethnic***, **w_indethg***, **relethg***, **w_yprace**, **w_maid**, **w_paid**) is combined to produce the ethnic group variable, **ethn_dv**. This is available in **xwavedat** and as **w_ethn_dv** in **w_indall** & **w_indresp** files. The variable **ethn_dv_source** records the information source used.

Country of birth

A derived variable, **bornuk_dv** has been constructed to identify whether the sample member was born in the UK or not based on country of birth from different sources (**w_ukborn**, **bw_plborn***, **bw_aplbornd**, **f_bornuk**). It is available in the cross-wave file **xwavedat**. Information is also collected about parents' countries of birth (**w_macob***, **w_pacob***) the first time someone is interviewed and about grandparents countries of birth (**w_mgmrob***, **w_mgprob***, **w_pgmrob***, **w_pgprob***) in Wave 1 for GPS+EMB and Wave 6 for IEMB.

Religion

Everyone is asked whether they belong to a religion (**w_oprlg**) and if they say yes, they are asked about the religion (**w_oprlg1**, **w_nirel*** for NI residents) and if they say no they are asked about the religion they were brought up in (**w_oprlg0**, **w_oprlg0ni** for NI residents). These questions were asked in Wave 1 and then every 3 years of the whole sample, but in the intervening waves these were asked of new entrants in the extra 5 minutes sample.

Language

In Wave 2, all adult respondents were asked about the main language spoken at home during their childhood (**w_kidlang**), and in Wave 6 all IEMB sample were asked the same question, and then in Wave 8 all new entrants or those who were previously interviewed but this information was missing were asked this question.

Citizenship

Respondents who were not born in the UK were also asked questions about their citizenship in Wave 1 (**w_citzn***). They could have chosen to respond that the country of their citizenship was UK, and up to two other countries. All new entrants who were not born in UK were also asked these questions in Waves 2, 3 and 4. Then from Wave 5 onwards, in addition to new entrants not born in the UK, continuing non-UK born members who had not yet reported being a UK citizen were asked this question again. In Wave 6, non-UK citizens in the IEMB sample were also asked about whether they had indefinite leave to remain and if they were intending to take UK citizenship.

Ethnicity of friends and associates

Information is also collected about the ethnicity of friends, networks and workplace. In Waves 3, 6, 9 everyone is asked about the ethnic composition of their friends (**w_simrace**), ethnicity of (up to 3) best friends (**w_netet1-3**) and only in Wave 5, extra 5 minutes sample respondents are asked about the ethnicity of their supervisor or manager (**w_bossethn**) and the ethnic composition of the workplace (**w_etcosamn w_etcowhn w_etcootn w_etcosamp w_etcowhp w_etcootp**).

National identity, ethnic identity, Britishness

The first time someone is interviewed they are asked about their national identity (**w_natid***), and in Waves 1 and 3, extra five minutes sample were asked how British they felt and in Wave 6 everyone excluding the IEMB sample were asked the same question (**w_britid**). In Waves 2 and 8, the extra five minutes sample was also asked a series of questions to measure their ethnic identity (**w_ethid* w_ethclose* w_pride***), and in Waves 5 and 11, young adults and recent immigrants in the extra five minutes sample were also asked these questions. Additionally, in Waves 2, 5, 8 and 11, everyone was asked about how

important their racial and ethnic background was to their sense of who they were (**w_scwhorurac**).

Harassment and Discrimination

In Waves 1, 3, 5 and 8, the extra five minutes sample was asked 4 sets of questions. In each set, they were first asked if they had experienced that phenomenon (physically attacked, verbally attacked or insulted, avoided, felt unsafe) in any of the public places listed (including educational institutions, workplace, home, in and around public transport, commercial, entertainment and recreational venues). If they said yes to any of these places, for each place they were asked what the reason was for that experience or feeling or behaviour and they could choose from these protected characteristics (sex, age, ethnicity, sexual orientation, health or disability, nationality, religion, language or accent, dress or appearance). These variables are **w_attack***, **w_resattack***, **w_insult***, **w_resinsult***, **w_avoid***, **w_resavoid***, **w_unsafe*** **w_resunsafe***. In Wave 11, these were asked of the entire sample.

Questions about employment discrimination (for promotion, training, recruitment) were asked in Waves 1 and 3 but as these questions did not work well, in Waves 5, 7 & 9, only the questions about recruitment discrimination were retained (**w_jobdeny**, **w_resjobdeny**, **w_disaffects**). In Wave 11, these questions were dropped completely. Instead, everyone was asked if they would describe themselves as belonging to a member of a group that is discriminated against in this country and if yes, then what are the grounds for that (**w_discrim**, **w_discrimgrd**).

6.4 Notes on educational qualification variables

Educational qualifications are an important measure for a wide range of analyses. However, it is very difficult to create a measure of equivalent highest qualifications for qualifications gained in different countries. Prior to the IEMB, regardless of where they had obtained their highest qualification, all respondents have been asked about their highest qualification using a UK-specific set of categories (**w_qfhigh**), which have been collated into a derived variable **w_hiquual_dv**. However, for those who obtained their qualifications in a different country, most of these categories will have little meaning or direct equivalence, with the possible exception of tertiary qualifications (higher education). As a result, in the IEMB it was established where the respondent obtained their highest qualification: if it was in the UK they were asked the standard question with UK response categories; and if it was outside the UK they were asked to select from set of options that were derived from [ISCED](#) categories (**f_qfhighoth**). This information can be supplemented and contextualised by reference to the country in which the qualification was obtained (**f_cntryqual**). To obtain the highest qualification for all those in the IEMB, analysts therefore need to combine the information from both **w_qfhighoth** and **w_hiquual_dv**. At Wave 9, these alternative questions were also asked of non-UK born sample members in the other samples (GPS,

BHPS, EMBS). Specifically, continuing sample members who were born outside the UK were asked these questions: **Qhigheduk**, **Qualhighoth**, **Qcntryqual**, **Qualhigh** & their follow-ups. New entrants to the sample who were born outside the UK were asked the same type of education questions as of the IEMB sample: **Higheduk**, **Qfhighoth**, **Cntryqual**, **Qfhigh**. Please refer to the [Questionnaires](#).

Using these educational qualification information, two variables have been derived.

- **w_isced11_dv**: the highest educational qualification, using ISCED 2011 coding frame, achieved by wave w
- **w_nisced11_dv**: the *new* highest educational qualification, using ISCED 2011 coding frame, achieved *in* wave w since the last interview

These variables are available in **indresp** files from Wave 9 onwards.

6.5 Notes on determining country of birth for children in Understanding Society

Country of birth questions are asked of all adults (16+ or higher) when they are interviewed for the first time (**w_ukborn**, **w_plborn**). As a result, those who entered the survey when they were 15 years old or less, were never directly asked a question about their country of birth. To determine whether the child was born in the UK or not, we suggest users match information of their parents' country of birth, and year of entering the UK if the parent was born outside UK and the child's year of birth to impute whether the child was born in the UK. Using this method requires the assumption that the child came to the country along with their parents and not before. In Wave 6, for the IEMB, the person who answered the household or enumeration grid was also asked whether each person in the household (or all ages) was born in the UK (**f_bornuk**) and if not, then whether they entered the country before or after 2009 (**f_uk2009**). For the IEMB sample, this additional information can also be used to impute country of birth for children.

7. Issues for analysis: sample design, non-response, weighting

The complex sample design of Understanding Society and possibly non-random non-response has important consequences for analysis.

7.1 Why use weights?

Firstly, the sample design is such that some sections of the population are selected with higher probability than others. Secondly, not everyone eligible for interview is interviewed. Some cannot be contacted, and some of those contacted refuse to give an interview. Finally, if those selected with higher selection probability are different in terms of some variable of interest from those selected with lower probability, then population estimates based on sample statistics will be biased. This is also the case if respondents are systematically different from non-respondents in terms of the variable of interest. In such cases (appropriately) weighted estimates provide unbiased population estimates.

For example, Table 7.1 provides a notional example where the population consists of two groups: 1000 individuals from group A and 500 individuals from group B. There are two sample designs. The first one is such that the sampling fraction is the same across two sub-populations: 0.10. This results in a sample of size 150, with 100 individuals from group A and 50 from group B. But if in order to analyse group B we require a larger sample, we could have a different sample design where the sample size for both groups is the same, 100. This means that the sampling fractions are different: 0.10 for group A and 0.20 for group B in this second sample.

If the mean earnings of Group A are 30 and that of Group B are 15, then

$$\text{the population mean earnings are} = \frac{(1000 \times 30) + (500 \times 15)}{1500} = \frac{37500}{1500} = 25$$

$$\text{the sample mean for sample 1} = \frac{(100 \times 30) + (50 \times 15)}{150} = \frac{3750}{150} = 25$$

$$\text{but the sample mean for sample 2} = \frac{(100 \times 30) + (100 \times 15)}{200} = \frac{4500}{200} = 22.5$$

For sample 2, unequal selection probability along with systematic difference in earnings across the two groups results in a biased estimate of the mean population earnings.

Weights are computed as inverse of the selection probabilities and so the weighted sample mean for sample 2 would be $= \frac{(100 \times 30 \times 10) + (100 \times 15 \times 5)}{(100 \times 10) + (100 \times 5)} = \frac{37500}{1500} = 25$, where the weights for group A sample members is 10 and for group B it is 5.

These are differences resulting from the design of the sample. There are also differences in response, both at Wave 1 and over time that can affect the representativeness of the sample with respect to the population of interest. These also need to be adjusted for by appropriate weights. The available weights include the combined adjustment for both design and non-response or attrition.

Table 7.1: Illustration of impact of not accounting for differential selection probabilities

	Group A	Group B
Population size	1000	500
Earnings	30	15
<i>Sample 1</i>		
Sample size	100	50
Sampling fraction	0.10	0.10
Weights	10	10
<i>Sample 2</i>		
Sample size	100	100
Sampling fraction	0.10	0.20
Weights	10	5

To illustrate the role of weights in producing population estimates the weighted and unweighted estimates of ethnic group distribution of adults living in England are shown in Table 7.2. The distribution for adults living in England and Wales according the 2011 Census is also shown. As is clear, in most cases the weighted estimates are closer to the Census distribution than the unweighted estimates.

Table 7.2: Ethnic group of 16+ year old residents of England

	2011 Census (England)	Understanding Society Wave 1, 2009-10	
		weighted	unweighted
White			
English/Welsh/Scottish/Northern Irish/British	81.1%	84.6%	72.7%
Irish	1.1%	0.9%	0.8%
Gypsy or Irish Traveller	0.1%	0.0%	0.0%
Other White	4.8%	3.9%	3.1%
Mixed/multiple ethnic group			
White and Black Caribbean	0.5%	0.4%	0.8%
White and Black African	0.2%	0.2%	0.3%
White and Asian	0.4%	0.2%	0.4%
Other Mixed	0.4%	0.2%	0.4%
Asian/Asian British			
Indian	2.6%	2.7%	4.6%
Pakistani	1.7%	1.5%	3.5%
Bangladeshi	0.7%	0.6%	2.8%
Chinese	0.8%	0.4%	0.7%
Other Asian	1.5%	0.7%	1.3%
Black/African/Caribbean/Black British			
African	1.6%	1.0%	2.8%
Caribbean	1.1%	1.4%	3.4%
Other Black	0.4%	0.1%	0.2%
Other ethnic group			
Arab	0.4%	0.2%	0.4%
Any other ethnic group	0.6%	0.9%	1.7%

7.2 Available weights

The weighting strategy is described in Lynn and Kaminska (2010). The underlying logic of weighting is to take account of unequal selection probabilities (e.g., different probabilities of selection in the EMB sample and the NI part of the GP sample) and differential nonresponse by assigning higher weights to those with lower selection probabilities and lower response propensities.

A single weight is not appropriate for all the possible different types of analysis that might be carried out and so different weights are provided with the data for different types of analysis; however, this may make the selection of the appropriate weight seem difficult. Please refer to the Main Survey User Guide [Weighting Guidance Section](#), particularly sub-section on [Selecting the correct weight for your analysis](#). ***It is very important that users consult that section before commencing analysis.*** The section describes the different types of weight variables available, their naming convention and also guidance on which weights to use. Weights vary by

- the population you want to generalise the results to (cross-sectional weights to generalise to the population in that wave period, longitudinal weights for generalising to the original sample)
- the sample being analysed or included in analysis (GPS+EMB, GPS+EMB+BHPS, BHPS, GPS+EMB+IEMB+BHPS)
- the units of analysis (adults, youths, households)
- the survey instruments used (household grid, household questionnaire, adult questionnaire, proxy questionnaire, adult self-completion, extra five minute questions, youth questionnaire)

For example, say you wanted to estimate the average wages for black Caribbean women living in England in 2013-2014. As this information is collected in the adult interviews (not proxy but in-person interviews), you would use the cross-sectional adult full-respondent weight for Wave 5 relating to all samples (GPS+EMB+BHPS): **e_indindub_xw**.

All this information is provided [here](#), which list the different weights and the types of analysis for which each should be used. Ethnicity and migration researchers may be particularly interested in the table that describes the weights to use if you use responses from at least one Extra Five Minutes question.

7.3 Analysis and Sample design variables

One important point to note is that even if the analysis is restricted to ethnic minority respondents, researchers should include ethnic minorities across **all sample components** not just in the EMB and/or IEMB sample to avoid coverage error. As discussed in Section 3.2, the EMB and IEMB samples were drawn from high ethnic minority concentration areas and so do not represent ethnic minorities living in low ethnic minority concentration areas. Just using the EMB and/or IEMB sample would mean that results would not be representative of all ethnic minorities living in the UK. The same argument applies for research or analysis of immigrants. Appropriate weights for the full sample should be used.

Another key point to note is that the design for most components of Understanding Society sample is stratified and clustered. Unless explicitly accounted for most statistical software assume that the sample design is a simple random sample and estimate standard errors of estimators based on that assumption. If these design features are not taken into account it will result in incorrect standard error estimates. It is therefore essential that you take explicit account of the stratification and clustering. Variables which indicate the stratum and the primary sampling unit (PSU) or cluster to which the sample member or sample household belongs, are provided with the data. New entrants are assigned the PSU and stratum of the OSM household they have joined. The PSU variable is **psu** and the variable indicating stratum is **strata**. These are available in **xwavedat**. For ease of use these variables are also available in wave specific files. In those files the variable names include a **w_** wave prefix. But note that although these have a wave prefix they *do not change across waves*.

8. Approach to translation and language

One key issue for the ethnicity strand was the extent to which the questionnaire should be translated into other languages and, if so, which languages. Translation of questionnaires is not simply a case of translating word-by-word from an original questionnaire, but ensuring that the construct being measured is equivalent and that responses provide the same information. Much of the work on translating survey instruments has come from the perspective of translating the same instrument in different countries, in order to harmonise the instrument for cross-national comparison. Nevertheless, many of the issues are the same in a multilingual context.

One of the key insights from this literature is that consideration of the need to translate the instrument may have implications for the initial design of the questionnaire and questions. For example, questions will need to be as simple and clear as possible; a question with ambiguous wording will not be a useful question in the original language and is likely to be an even less useful question once translated. It is important to acknowledge that whilst the language in a question may be quite simple, it is also very specific. The language of the question should be free of jargon or slang, or if these are used then a note for the translators should be written which sets out in more detail the concept the word is being used to convey. At the same time in the translation process there has to be careful attention to the principles behind questions and response categories in the original. Response categories are used as a tool of measurement and so are carefully chosen so as to be neutral, with symmetrical response categories. Any translation procedure will need to ensure that those doing the translation are aware of this.

In terms of the translation process, there are a variety of practices. 'Best practice' is often regarded to be represented by the approach used in the European Social Survey, and their TRAPD, Translation, Review, Adjudication, Pretesting and Documentation approach (Harkness 2003). This is an iterative process and covers the stages that a questionnaire goes through, often more than once. There are restrictions in the extent to which it is possible to engage in concurrent consideration of questions in different languages and in the possibility to change questions or suites of questions that are recognised validated measures. However, while much of the literature now suggests that back translation is not an appropriate approach to translation, there is an emphasis on using more than one translator, with the initial translation checked by a second translator and adjudication taken place in the case of disagreements. It is also recognised that a translation pre-test or pilot is an important element of the quality checking of a translation.

In existing UK surveys at the time that UKHLS Wave 1 was in preparation a range of practice was used, from the exclusion of those who did not speak English to the use of translated advance materials and survey leaflets combined with the use of bilingual interviewers, translators or the use of another (adult) household member to translate 'on the fly'. Pre-

translated instruments were used in some cases and administered by a bilingual interviewer, or translator, but this was less common.

Given that at the time of preparation of UKLHS the Census had not included a question on language spoken and fluency in English, information on relative English language fluency across different language groups had to be extrapolated from other surveys.

On this basis UKLHS developed translated instruments using independent translator and checker and adjudication. The translated instruments were implemented in computer assisted personal interviewing (CAPI) software, obviating the need for separate, paper-based versions of the translation, which are associated with much longer interview times. This meant that bilingual interviewers could simply switch to the language of choice (or between that and English), while non-bilingual interviewers could work alongside a translator, helping them to enter the response directly into the CAPI laptop.

A translation pilot was carried out on selected languages and found that this process largely worked well. The selection of languages was based, as noted, on inferences from other surveys combined with the expected composition of the UKLHS overall sample. Note that the inferences were based not just on language use but on inferences about English language fluency for speakers of those languages. Hence it was expected, for example, that Hindi speakers would tend to be relatively fluent in English, while Somali speakers were expected to have much lower levels of English language fluency. On that basis the chosen languages for translation were Welsh (required under the Welsh Language Act), Urdu, Punjabi (in both Urdu and Gurmukhi scripts), Gujarati, Bengali, Cantonese, Somali and Arabic. For all these languages, interviews were carried out using the translation by either accredited bilingual interviewers or by accredited translators alongside the interviewer. Where respondents spoke languages not available in these translations, practice resorted to the use of bilingual interviewers or translators or household members, translating 'on the fly', for the respondent.

In fact, use of translated interviews was in fact rather low, which is consistent with the findings from other surveys. In Wave 1, out of 50,994 interviews, 456 translated individual (adult) questionnaires were used (i.e., around 0.8%). As a result the selection of languages was revised from Wave 6. Moreover, at that point the introduction of the IEMB suggested that there might be a greater need for other languages, such as those spoken by European immigrants. In addition, by the time the IEMB design was being prepared, the 2011 Census information on languages spoken and English language fluency among those whose first language was not English or Welsh was available. This was therefore used to inform decisions about translation. Hence the languages for translation from Wave 6 onwards were: Welsh, Bengali, Gujarati, Polish, Portuguese, Punjabi (Gurmukhi), Punjabi (Urdu), Somali, Turkish, Urdu. Around 3.3% of the adult interviews were translated into one of these languages, with Urdu being the most common language the interview was conducted in.

9. Help and support

9.1. Understanding Society resources

Find out more about the Study online at: www.understandingsociety.ac.uk

To search for variables:

<https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation>

To look through the questionnaires:

<https://www.understandingsociety.ac.uk/documentation/mainstage/questionnaires>

For the main survey user guide, this user guide and other topic specific user guides visit:

<https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/weighting-guidance>

For syntax in Stata, SPSS, SAS and R for some common data management tasks:

<https://www.understandingsociety.ac.uk/documentation/mainstage/syntax>

For Frequently Asked Questions (FAQ) about the study:

<https://www.understandingsociety.ac.uk/help/faqs>

For further help, make use the [Help pages](#) for Understanding Society. If you have questions [Contact User Support](#).

A series of [training courses](#) (including [online DIY Moodle courses](#) and [webinars](#)) are offered to introduce new users to the data and deal with key issues of data manipulation and analysis.

9.2. Resources for ethnicity and immigration research

Visit the Ethnicity and Immigration Topic page to know more about ethnicity and immigration research using Understanding Society:

<https://www.understandingsociety.ac.uk/topic-page/ethnicity-and-immigration/>

To find research publications on ethnicity and immigration using Understanding Society:

<https://www.understandingsociety.ac.uk/research/publications>

[The UKDS ethnicity theme pages](#) provides a useful introduction to ethnicity research and resource in the UK. The UK Data Archive hosts a large number of qualitative and quantitative survey data that enable research on ethnicity and immigration and can be identified using their [searchable data catalogue on UKDS](#).

10. Citation information and ethical approval statement

10.1 Data citation

If you use Understanding Society data you must acknowledge the data source. All works which use or refer to these materials should acknowledge these sources by means of bibliographic citation. To ensure that such source attributions are captured for bibliographic indexes, citations must appear in footnotes or in the reference section of publications. The citation changes at each release to reflect the addition of the data from the new wave. Please visit <https://www.understandingsociety.ac.uk/documentation/citation> for the citation for the latest version of the data. Please cite each dataset that you use.

10.2 Citation for this Guide:

McFall, S., Nandi, A. and Platt, L. (2024). *Understanding Society: UK Household Longitudinal Study: User Guide to Ethnicity and Immigration Research*. 12th Edition. Colchester: University of Essex.

10.3 Ethics

Collecting, using and sharing data in research with people requires that ethical and legal obligations are respected. The Understanding Society study protocols and research programme are scrutinised by a number of research ethics committees to assure that ethical and legal obligations are respected at all times. Most recent version the ethical approval statement can be found on the [Understanding Society website](#). For further details about the various committees which have provided ethical approval of the Understanding Society study and its components see the [Ethics section](#) in the Main User Guide.

11. References

- The American Association for Public Opinion Research. 2016. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition*. AAPOR
- Berthoud, R., Fumagalli, L., Lynn, P. and Platt, L. (2009) Design of the Understanding Society ethnic minority boost sample. *Understanding Society Working Paper 2009-02*. Colchester: University of Essex.
- Burton, J., Nandi, A. and Platt, L. (2010) Measuring ethnicity: challenges and opportunities for survey research. *Ethnic & Racial Studies* 33 (8): 1332 – 1349.
- Fumagalli, Laura, Knies, Gundi and Buck, Nicolas (2017): Understanding Society: The UK Household Longitudinal Study harmonised British Household Panel Survey (BHPS) User Guide. Colchester: University of Essex.
- Harkness, J. A. 2003. Questionnaire Translation. Pp. 35-56 in *Cross-Cultural Survey Methods*, edited by Harkness, J. A., F. J. R. van de Vijver, and P. Ph. Mohler. Hoboken, NJ: Wiley.
- Institute for Social and Economic Research (2021), Understanding Society: Waves 1-10, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009, User Guide, 29 October 2020, Colchester: University of Essex.
- Lynn, P. (2009) Sample design for Understanding Society. *Understanding Society Working Paper 2009-01*. Colchester: University of Essex.
- Lynn, P., Kaminska, O. (2010) Weighting strategy for Understanding Society. *Understanding Society Working Paper 2010-05*. Colchester: University of Essex
- Lynn, P., Burton, J., Kaminska, O., Knies, G. and Nandi, A. (2012) An initial look at non-response and attrition in Understanding Society. *Understanding Society Working Paper 2012-02*. Colchester: University of Essex.
- Lynn, P., Knies, G. (ed.) (2016) Understanding Society: the UK Household Longitudinal Study: Wave 1-5 Quality Profile. Colchester: University of Essex.
- Lynn, P., Nandi, A., Parutis, V. and Platt, L. (2017). Design and implementation of a high quality probability sample of immigrants and ethnic minorities. *Understanding Society Working Paper 2017-11*. Colchester: University of Essex.
- Taylor, M. (ed). with Brice, J., Buck, N. and Prentice-Lane, E. (2018) British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices. Colchester: University of Essex.

12. FIND OUT MORE

Sign up to our newsletter at: www.understandingsociety.ac.uk/email/signup

Email us at: info@understandingsociety.ac.uk

Follow us on Twitter (@usociety - <https://twitter.com/usociety>)

Please let us know about any research you are doing using *Understanding Society* data.

<https://www.understandingsociety.ac.uk/research/request>

Appendix I: Screening questions

The EMB screening question (Wave 1)

D7. [SHOW SCREENING CARD] Does anyone living at this address come from, or have parents or grandparents from any of the following ethnic groups? CODE ALL THAT APPLY

- 1 Indian
- 2 Mixed Indian – (parents or grandparents from Indian ethnic group AND parents or grandparents from a non-Indian group)
- 3 Pakistani
- 4 Bangladeshi
- 5 Sri Lankan
- 6 Caribbean/West Indian
- 7 Mixed Caribbean/West Indian (parents or grandparents from Caribbean/West Indian ethnic group AND parents or grandparents from a non-Caribbean/West Indian ethnic group)
- 8 North African
- 9 Black African
- 10 African Asian
- 11 Chinese
- 12 Far Eastern (includes Filipino, Thai, Malaysian, Japanese, Vietnamese, Singaporean, Indonesian, Korean, Burmese)
- 13 Turkish
- 14 Middle Eastern/Iranian (includes Israeli, Palestinian, Lebanese, Syrian, Jordanian, Temeni, Saudi, Iraqi, Afghani, other Gulf states)
- 96 None of these
- 95 Unable to complete screening questions

IEMB screening questions (Wave 6)

Q6e: Is there anyone living at this address who was born outside the UK?

- 1 Yes
- 2 No
- 3 Unable to establish

Q6f: [SHOW SCREENING CARD] Does anyone living at this address come from any of the following ethnic groups, or have parents or grandparents from any of these groups? CODE ALL THAT APPLY

- 1 Indian
- 2 Mixed Indian – (parents or grandparents from Indian ethnic group AND parents or grandparents from a non-Indian group)
- 3 Pakistani
- 4 Bangladeshi
- 5 Sri Lankan
- 6 Caribbean/West Indian
- 7 Mixed Caribbean/West Indian (parents or grandparents from Caribbean/West Indian ethnic group AND parents or grandparents from a non-Caribbean/West Indian ethnic group)
- 8 North African
- 9 Black African
- 10 African Asian
- 11 Chinese
- 12 Far Eastern (includes Filipino, Thai, Malaysian, Japanese, Vietnamese, Singaporean, Indonesian, Korean, Burmese)
- 13 Turkish
- 14 Middle Eastern/Iranian (includes Israeli, Palestinian, Lebanese, Syrian, Jordanian, Temeni, Saudi, Iraqi, Afghani, other Gulf states)
- 96 No, None of these
- 95 Unable to complete screening questions

Appendix II: List of variables corresponding to the extra five minutes questions, Waves 1-14

These questions were asked of the Extra Five Minutes sample. But some questions were asked of a sub-sample of the Extra Five Minutes sample or of additional sample members in which case these are indicated in parenthesis.

Wave	Module	Variable stem name
1	Migration History	mabroad mnabroad mindirect mnotherc mreturned mlived mnlived moveage mnmoves mlivedist mlivedist5 mlivedistf mabroadc_all*¹ mindirectc_all*² mlivedc_all*³
	Harassment	attacked* resattacked* avoidance* resavoid* insulted* resinsulted* unsafe* resunsafe*
	Discrimination	joblook jobdeny resjobdeny* eed12 promodeny respromodeny* traindeny restraindeny*
	Remittances	remit1-remit5 remitfreq* remitreg* remitam* remitusual remusamt remcntry_all⁴
	Ethnicity and National identity	Britid <i>(In the first six months these were asked of the entire sample)</i>
2	Ethnic identity	ethid* ethclose* pride* food*
	Religion	oprlg* <i>(Asked of new entrants in the Extra Five Minutes sample)</i>
	Political engagement	perpolinf* colbens* civicduty* polcost votenorm perbfts grpbfts voteintent demorient
3	Harassment	attacked* resattacked* avoidance* resavoid* insulted* resinsulted* unsafe* resunsafe*
	Discrimination	joblook jobdeny resjobdeny eed12 promodeny respromodeny* traindeny restraindeny* disaffects
	Religion	oprlg* <i>(Asked of new entrants in the Extra Five Minutes sample)</i>
	Ethnicity and National identity	britid
	Self-completion Best Friends ³	netmet* netweb* nettalk* netdo* <i>(Rest of this module are asked of the whole sample)</i>
4	Remittances	remit1-remit5 remitfreq* remitreg* remitam* remitusual* remusamt* remhow* remcntry_all*⁴
	Prayer Frequency	prayfreq
	Religious practice	rleat rldrnk rlwear rlmarg rlschl rlcharity rlsave rlfriends rljob
	Service use	servuse opserv srvynot
	Wealth, assets and debt ³	hascurr currynot refbank refbnky svacts nosvy svamt svb* svsj* svpn* svsk* savdocs* infsave* debtinfo* debtwho* debtot* creditor* creditamt* unusexp* expmanage* <i>(Rest of this module are asked of the whole sample)</i>

Wave	Module	Variable stem name
5	Ethnic identity module	ethid* ethclose* pride* food* (<i>Asked of 16–19-year-olds in the Extra Five Minutes sample and all recent immigrants⁵</i>)
	Language	englang engspk spkdif engtel teldif engread readdif engform formdif (<i>Asked of those in the Extra Five Minutes sample for whom “englang” info was missing from Wave 1</i>)
	Religion	oprlg* (<i>Asked of new entrants in the Extra Five Minutes sample</i>)
	Harassment	attacked* resattacked* avoidance* resavoid* insulted* resinsulted* unsafe* resunsafe*
	Cultural participation	culturea cultureb culturec cultured culturee culturef cultureg
	Work composition	bossethn bossex etcosamn etcowhn etcootn cowosexn etcosamp etcowhp etcootp cowosexp
	Discrimination	joblook jobdeny* resjobdeny* disaffects
6	Self-completion Best Friends ³	netmet* netweb* nettalk* netdo* (<i>Rest of this module are asked of the whole sample</i>)
	Service use	servuse opserv srvynot
	Religion	oprlg* (<i>Asked of new entrants in the Extra Five Minutes sample</i>)
	Migration history (asked of IEMBS)	mabroad mnabroad mindirect mnotherc mreturned mlived mnlived moveage mnmoves mlivedist mlivedist5 mlivedistf mabroadc_all*¹ mindirectc_all*² mlivedc_all*³
7	Remittances	remit1-remit5 remitfreq* remitreg* remitamt* remitusual* remusamt* remhow* remcntry_all*⁴
	Harassment	attacked* resattacked* avoidance* resavoid* insulted* resinsulted* unsafe* resunsafe*
	Discrimination	joblook jobdeny* resjobdeny* disaffects
	Religion	oprlg* (<i>Asked of new entrants in the Extra Five Minutes sample</i>)
	Migration history (asked of all non-UK born non-IEMBS)	mabroad mnabroad mindirect mnotherc mreturned mlived mnlived mnmoves mlivedist mlivedist5 mlivedistf mabroadc_all*¹ mindirectc_all*² mlivedc_all*³
8	Ethnic identity	ethid* ethclose* pride* food*

Wave	Module	Variable stem name
9	Religion	oprlg* <i>(Asked of new entrants in the Extra Five Minutes sample)</i>
	Harassment	attacked* resattacked* avoidance* resavoid* insulted* resinsulted* unsafe* resunsafe*
	Discrimination	joblook jobdeny* resjobdeny* disaffects
10	Language	englang engspk spkdif engtel teldif engread readdif engform formdif <i>(Asked of those in the Extra Five Minutes sample for whom “englang” info was missing from Wave 1)</i>
	Religion	oprlg* <i>(Asked of new entrants in the Extra Five Minutes sample)</i>
	Religious practice	rleat rldrnk rlwear rlmarg rlschl rlcharity rlsave rlfriends rljob rljoby
	Service use	servuse opserv srvynot
	Remittances	remit1-remit5 remitfreq* remitreg* remitam* remitusual* remusamt* remhow* remcntry_all ⁴
11	Ethnic identity	ethid* ethclose* pride* food* <i>(Asked of 16-19 year olds in the Extra Five Minutes sample and all recent immigrants⁵)</i>
	Religion	oprlg* <i>(Asked of new entrants in the Extra Five Minutes sample)</i>
12	Religion service attendance	oprlg2 <i>(Asked of the whole sample in Waves 1, 4 & 8)</i>
	Prayer Frequency	prayfreq
13	Remittances	remit1-remit5 remitfreq* remitreg* remitam* remitusual* remusamt* remhow* remcntry_all ⁴
	Religion	oprlg* <i>(Asked of new entrants in the Extra Five Minutes sample)</i>
14	Ethnic identity	ethid* ethclose* pride* food*
	Service use	servuse opserv srvynot
Notes: Process-generated and derived variables not included in the listing; Special Licence release variables in bold; Refer to the Long-Term Content Plan for current and future variable frequency of occurrence. ¹ Created from non-released responses to questions mabroadc and mabroadco. ² Created from non-released responses to questions mindirectc and mindirectco. ³ Created from non-released responses to questions mlivedc mlivedco. ⁴ Created from non-released responses to questions remcntry and remcntryo. ⁵ Recent immigrants defined as Wave 1 sample members who came to the UK in the last three years.		

Appendix III: List of variables corresponding to the new questions asked of the IEMB sample in Wave 6

Wave	Module	Variable stem name
6 (asked of IEMBS)	Migration related	mreason1-7, mreason97, mreasono_code
		dvage2uk age2ukc dv2uk16 lkmvcntry xpmvcntry
		mintent mintwhen
	Citizenship, residency	indeflv intendukc
	Non-UK educational qualifications	qfvocnonuk qfvocuk cntryqual qfhhighoth higheduc qfctry qfctry_all ¹
	Parents migration information	payruk1 pacurr mayruk1 macurr
	Parents’ job information before and after moving to UK	majuk majukevr majuke190 majuke100 majuke110 majcb majcbe90 majcbe00 majcbe10 pajuk pajukevr pajuke190 pajuke100 pajuke110 pajcb pajcbe90 pajcbe00 pajcbe10 ⁷
7 (asked of all non-UK born non-IEMBS)	Migration related	mreason1-7, mreason97, mreasono_code
		dvage2uk age2ukc dv2uk16
		mintent
9 (asked of all non-UK born non-IEMBS)		mintent ²
Notes: Process-generated and derived variables not included in the listing; Special Licence release variables in bold ; ¹ Created from non-released responses to questions qfctry. ² The filtering for this question didn’t work correctly in Wave 7 and so this question was asked again in Wave 9		

Appendix IV: Names of occupation (SOC) variables

The names of IEMB new questions related to occupation were changed in the data. As with all occupation questions these were coded to three variables according to SOC90, SOC200 and SOC2010 coding frames.

Name in questionnaire	Question	Corresponding occupation variables in data files	New in IEMB	Special Vs End User Licence
j1soc00	First job since leaving full-time education	j1soc90 j1soc00 j1soc10	In both	These variables are available under Special Licence. Under the End User Licence these variables are not available. Instead a condensed version of these variables are available. These variables have the same name with an additional suffix “_cc”
jbsoc00	Occupation of current job	jbsoc90 jbsoc00 jbsoc10	In both	
j2soc00	Second job currently held	j2soc90 j2soc00 j2soc10	In both	
jsoc00	Occupation of last job	jsoc90 jsoc00 jsoc10	In both	
j1soc00uk	First job since coming to UK	j1uksoc90 j1uksoc00 j1uksoc10	Only in IEMB	
j1joboth	Last job before moving to the UK	jlcsoc90 jlcsoc00 jlcsoc10	Only in IEMB	
masoc00	Mother’s occupation (SOC 2000), resp. aged 14	masoc90 masoc00 masoc10	In both	
majuke1	Mother’s first UK job	maj1uksoc90 maj1uksoc00 maj1uksoc10	Only in IEMB	
majcbe	Mother’s occupation in Country of Birth	majlcsoc90 majlcsoc00 majlcsoc10	Only in IEMB	
pasoc00	Father's occupation (SOC 2000), resp. aged 14	pasoc90 pasoc00 pasoc10	In both	
pakuke1	Father’s first UK job	paj1uksoc90 paj1uksoc00 paj1uksoc10	Only in IEMB	
pajcbe	Father’s occupation in Country of Birth	pajlcsoc90 pajlcsoc00 pajlcsoc10	Only in IEMB	

Appendix V: List of questions to measure ethnicity

Questions asked in the adult questionnaire about the respondent

Question	Variable Name	Description	Sample	Wave asked in
Ethnic group	w_racel, w_racelo_code	modified ONS 2011 Census ethnic group question ^a	All	When first interviewed
	w_racelt, w_racelwt, w_racelmt, w_racelat, w_racelbt, w_racelot_code	modified ONS 2011 Census ethnic group question ^a , Telephone Version	All	When first interviewed, Wave 2 onwards – only Telephone mode
	bw_racel_bh	2001 ONS harmonised UK ethnic group	BHPS	First interviewed, BHPS Waves 11-18
	bw_race_bh	1991 ONS harmonised UK ethnic group	BHPS	First interviewed, BHPS Waves 1-10
Ethnic identity	w_ethid*	Whether own country, own language, parents cob if different from own etc. are “important to your sense of who you are”	Extra 5 minutes	Repeated. First asked Wave 2. See Tables 4.1a & 4.1b for rotation.
	W_ethclose*	Whether feel happy to meet someone from same country as own etc.	Extra 5 minutes	
	w_pride*	Pride in own country of birth etc.	Extra 5 minutes	
Religion	w_oprlg1, w_nirel	Religion belong to	All	Repeated. First asked Wave 1. See Table 4.1b for rotation.
	W_oprlg0ni, w_oprlg0	Religion brought up in (asked if does not belong to a religion)	All	

Questions asked in the adult questionnaire about the respondent (continued)

Question	Variable Name	Description	Sample	Wave asked in
Own country of birth ^b	w_ukborn	Whether born in England, Scotland, Wales, Northern Ireland or non-UK	All	When first Interviewed
	w_plbornc, w_plbornc_all	Which specific country, if born outside UK	All	
	bw_plbornc	Country of birth if born outside UK	BHPS	
	bw_plbornd	District of birth if born in UK	BHPS	
	w_macob, w_macob_all	Mother's country or birth	All	
	w_pacob, w_pacob_all	Father's country of birth	All	
Language/ Mother tongue	w_kidlang	The main language that was spoken at home during childhood	All (not BHPS)	Wave 2
Britishness	w_britid	How important is being British to the respondent	Extra 5 minutes	Repeated. First asked Wave 1. See Tables 4.1a & 41.b for rotation.
National identity	w_natid1-w_natid6, w_natid97	Whether national identity is English, Scottish, Welsh, Northern Irish, British, Irish, Other (code all that apply)	All	When first Interviewed

Questions asked in the adult questionnaire about respondent's parents and grandparents

Question	Variable Name	Description	Sample	Wave asked in
Grandparents' countries of birth ^b	w_mgmrob, w_mgmrob_all	Maternal grandmother's country of birth	All	Wave 1 for GPS+ EMB, Wave 6 for IEMB, BHPS Wave 18
	w_mgprob, w_mgprob_all	Maternal grandfather's country of birth	All	
	w_pgmrob, w_pgmrob_all	Paternal grandmother's country of birth	All	
	w_pgprob, w_pgprob_all	Paternal grandfather's country of birth	All	
Parents' Ethnic group	w_maid	Mother's ethnic group	GPS+EMB	Wave 1
	w_paid	Father's ethnic group	GPS+EMB	
Strength of identification with parents' ethnic group	w_smaid	Strength of identification with mother's ethnic group, if different from father's ethnic group	GPS+EMB and if w_racel is not White majority ^c	
	w_spaid	Strength of identification with father's ethnic group		

^aNote that the values corresponding to the categories "black or black British: Caribbean" and "black or black British: African" have been switched here as compared to the ONS Census 2011 question. Also, the BHPS sample members were asked this question in Wave 18 and the harmonized version is available in xwavedat file

^bFor all country of birth variables, those who report other country are asked to name those countries and these are coded later and renamed with a suffix "_all". Some of these are available in the Special Licence version of the data

^cDefined as those who choose white: British/English/Scottish/Welsh/Northern Irish

Questions asked in the household grid

Question	Variable Name	Description	Sample	Wave asked in
Ethnic minority boost sample screening questions ^a	w_ethnic*	Whether the person or their parents/grandparents come from one of the ethnic groups (as reported in the household grid)	10+ year old GPS+EMB enumerated individuals	Wave 1
Immigrant and ethnic minority boost sample screening questions ^a	w_uk2009	Whether born in England, Scotland, Wales, Northern Ireland or non-UK country	All IEMB enumerated individuals	Wave 6
	w_indethg*	Ethnic group	All IEMB enumerated individuals	
	w_relethg*	Ethnic group of parents or grandparents	All IEMB enumerated individuals	

^a The ethnic group question used for screening has a different coding frame from the ONS harmonised ethnic group question (w_race1). The exact question and the response options are available in the online documentation.

Questions asked in the screeners

Question	Variable Name	Description	Sample	Wave asked in
Ethnic minority boost sample screening questions ^a	w_d7*	Whether there is a person in the household who or their parents/grandparents come from one of the ethnic groups	EMB screened households	Wave 1
Immigrant and ethnic minority boost sample screening questions ^a	f_q6e	Whether there is a person in the household who was not born in the UK	IEMB screened households	Wave 6
	f_q6f*	Whether there is a person in the household who or their parents or grandparents were from that ethnic group		
	f_bornuk	Whether born in the UK		
	f_uk2009	Whether came to the UK before or after 2009		

^a The ethnic group question used for screening has a different coding frame from the ONS harmonised ethnic group question (w_race1). The exact question and the response options are available in the online documentation.

Appendix VI: Syntax for derived variables mintent_dv mintwhen_dv mreason?_dv

```
// Syntax for creating mintent_dv mintwhen_dv mreason?_dv

foreach w in f g i {
    local i=strpos("abcdefghi","`w'")
    use "$ms/ukhls_w`i'/'`w'_indresp", clear
    isvar pidp *mint* *mreas*
    keep `r(varlist)'
    tempfile `w'
    save ``w'', replace
}
use `f', clear
merge 1:1 pidp using `g', nogen
merge 1:1 pidp using `i', nogen

generat mintent_dv=-9
generat mintwhen_dv=-9
generat wmintqasked_dv=-9

replace wmintqasked_dv=6      if f_mintent>=-2 & f_mintent<. & mintent_dv===-9
replace mintwhen_dv=f_mintwhen if f_mintent>=-2 & f_mintent<. & mintent_dv===-9
replace mintent_dv =f_mintent if f_mintent>=-2 & f_mintent<. & mintent_dv===-9

replace wmintqasked_dv=7      if g_mintent>=-2 & g_mintent<. & mintent_dv===-9
replace mintwhen_dv=g_mintwhen if g_mintent>=-2 & g_mintent<. & mintent_dv===-9
replace mintent_dv =g_mintent if g_mintent>=-2 & g_mintent<. & mintent_dv===-9

replace wmintqasked_dv=9      if i_mintent>=-2 & i_mintent<. & mintent_dv===-9
replace mintwhen_dv=i_mintwhen if i_mintent>=-2 & i_mintent<. & mintent_dv===-9
replace mintent_dv =i_mintent if i_mintent>=-2 & i_mintent<. & mintent_dv===-9

lab copy i_mintent  mintent_dv
lab copy i_mintwhen mintwhen_dv

lab val mintent_dv mintent_dv
```

```

lab val mintwhen_dv mintwhen_dv

lab def wmintqasked_dv -9 "missing" 6 "Wave 6" 7 "Wave 7" 9 "Wave 9"
lab val wmintqasked_dv wmintqasked_dv

lab var mintent_dv "current migration intentions"
lab var mintwhen_dv "When expects to migrate"

lab var wmintqasked_dv "Wave mintent questions asked"

forvalues i=1/7 {
    generat mreason`i'_dv=-9
    replace mreason`i'_dv=f_mreason`i' if f_mreason97>=-2 & f_mreason97<.
    replace mreason`i'_dv=g_mreason`i' if g_mreason97>=-2 & g_mreason97<.
    lab copy g_mreason`i' mreason`i'_dv
    lab val mreason`i'_dv mreason`i'_dv
}
foreach i in 97 o_code {
    generat mreason`i'_dv=-9
    replace mreason`i'_dv=f_mreason`i' if f_mreason97>=-2 & f_mreason97<.
    replace mreason`i'_dv=g_mreason`i' if g_mreason97>=-2 & g_mreason97<.
    lab copy g_mreason`i' mreason`i'_dv
    lab val mreason`i'_dv mreason`i'_dv
}
lab var mreason1_dv "for work (e.g. wanted to work and earn money in the uk, to do an internship"
lab var mreason2_dv "joined partner or spouse already living in the uk"
lab var mreason3_dv "joined other family members already living in the uk"
lab var mreason4_dv "moved together with family members"
lab var mreason5_dv "education (e.g. wanted to study here or take language classes etc.)"
lab var mreason6_dv "political (did not feel safe in the country of birth)"
lab var mreason7_dv "simply wanted to live in another country / in the uk"
lab var mreason97_dv "other"
lab var mreasoно_code_dv "mreason - reason for moving to the UK"

keep pidp mintent_dv mintwhen_dv wmintqasked_dv mreason*

```

Appendix VII: Fieldwork periods for different samples

	2009		2010	2011	2012	2013	2014	2015	2016	2017		2022	2023
General Population Sample – Great Britain part		Wave 1											
			Wave 2										
				Wave 3									
					Wave 4								
						Wave 5							
							Wave 6						
								Wave 7					
									Wave 8				
General Population Sample –Northern Ireland part	Wave 1												
			Wave 2										
				Wave 3									
					Wave 4								
						Wave 5							
							Wave 6						
								Wave 7					
									Wave 8				
Ethnic Minority Boost Sample		Wave 1											
			Wave 2										
				Wave 3									
					Wave 4								
						Wave 5							
							Wave 6						
								Wave 7					
									Wave 8				
British Household Panel Survey sample													
			Wave 2										
				Wave 3									
					Wave 4								
						Wave 5							
							Wave 6						
								Wave 7					
									Wave 8				
Immigrant and Ethnic Minority Boost Sample								Wave 6					
									Wave 7				
										Wave 8			
General Population Sample Boost GPS2 – Great Britain part												Wave 14	
General Population Sample Boost GPS2 – Northern Ireland part												Wave 14	

