

SAMPLE DESIGN

The *Understanding Society* main survey sample consists of a new large General Population Sample (GPS) plus three other components: the Ethnic Minority Boost Sample (EMBS), the former BHPS sample, and the Immigrant and Ethnic Minority Boost Sample (IEMBS). The design of the first three components is described in more detail in an *Understanding Society* working paper, see [Lynn \(2009\)](#). The design of the IEMBS is described in [Lynn, Nandi et al. \(2016\)](#). The GPS is based on two separate samples of residential addresses in England, Scotland and Wales and in Northern Ireland. The England, Scotland and Wales sample is a proportionately stratified (equal probability), clustered sample of addresses selected from the Postcode Address File. Northern Ireland has an unclustered systematic random sample of addresses selected from the Land and Property Services Agency list of domestic addresses.

GENERAL POPULATION SAMPLE

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

In Northern Ireland, 2,395 addresses were selected in a single stage from the list of domestic addresses. In combination, this data release is therefore based upon a total of 49,915 addresses.

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

GENERAL POPULATION COMPARISON SAMPLE

The General Population Comparison sample (GPC) has one sampled address for 40% of the selected postal sectors in General Population Sample (GPS) component for Great Britain. In other words, of the 2,640 general population sectors, 60% of

them (1,584) contain 18 GPS addresses and the other 40% contain 17 GPS addresses and one GPC address. The persons in these households will be designated as members of the GPC sample, regardless of ethnic group membership. Members of the GPC sample are a random subsample of the GPS component and they should be included in analyses of the GPS component.

ETHNIC MINORITY BOOST SAMPLE

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

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

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

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

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

- A “screen” was carried out to identify whether there were any persons from target ethnic groups in the household.
- A random mechanism was applied to certain target groups identified by the screen in order to select only a desired proportion into the sample (non-mixed Indian, Pakistani, non-mixed Caribbean, African, Far Eastern, Middle Eastern). For other target groups, all resident persons were included in the sample (mixed Indian, Bangladeshi, mixed Caribbean, Sri Lankan, Chinese, Turkish).

- In households included in the sample in the previous two steps, all members of target ethnic groups were deemed to be members of the EMBS (including children). All persons of other ethnic groups are not EMBS members. They will be interviewed as temporary sample members for so long as they remain co-resident with at least one EMBS member.

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

FORMER BHPS SAMPLE

The sample issued at Wave 2 consisted of all members from the BHPS sample who were still active at Wave 18 of the BHPS and who had not refused consent to be issued as part of the *Understanding Society* sample. It should be noted that the BHPS sample contains different components, including the original sample (first selected in 1991), boost samples in Scotland and Wales (first selected in 1999), and a Northern Ireland sample (selected in 2001). For further details of the BHPS sample, see Section IV of the BHPS User Guide ([Taylor 2010](#)).

IMMIGRANT AND ETHNIC MINORITY BOOST SAMPLE

This sample was introduced at Wave 6. It includes people who were born outside the United Kingdom (“immigrants”) and members of five ethnic minority groups: Indian, Pakistani, Bangladeshi, Caribbean, and African. Some people, of course, fall into both categories. This sample therefore provides coverage for the first time of people who have entered the UK since Wave 1 of the Study (“new immigrants”), while also boosting the numbers of immigrants who arrived earlier and of ethnic minorities who either arrived earlier or were born in the UK. The IEMBS was designed to provide around 2,000 adult immigrant respondents and around 2,500 from the target ethnic minority groups.

The sample was identified through in-person doorstep screening of a set of addresses that were sampled from the Postcode Address File following a stratified multi-stage design in which the strata were defined by small area level indicators from the 2011 population census of the distribution of ethnic groups and immigrants. Five strata were created. Sampling was restricted to four strata, the fifth consisting of the sectors with the very lowest proportions of immigrants and ethnic minorities. Sampling fractions varied between the four strata, with the highest sampling fraction applied to a stratum with the highest proportions of Africans. In each sampled stratum, a number of postcode sectors were selected with probability proportional to the predicted number of eligible households. In each sampled sector, a number of addresses were selected such that the predicted number of eligible households in the sample did not vary between sectors within a stratum (so the number of selected addresses was larger in sectors with a lower predicted proportion of eligible households). A screened household was eligible for interview if it contained at least one person who was born outside the UK and/or a member of a relevant ethnic minority group, even if that person was a child.

The “boost” samples do not therefore provide complete population coverage of the relevant subgroups but are instead designed to be used in combination with the

other samples, as described above. The sample of “new immigrants” is estimated to provide around 74% population coverage.

SAMPLE STATUS AND FOLLOWING RULES

There are three possible sample statuses: Original Sample Members (OSMs), Temporary Sample Members (TSMs), and Permanent Sample members (PSMs).

Original Sample Members (OSMs)

All members of *Understanding Society* GPS households enumerated at Wave 1 - including absent household members and those living in institutions who would otherwise be resident - are Original Sample Members (OSMs). All ethnic minority members of an enumerated household eligible for inclusion in the EMBS are OSMs. In the IEMBS, each household member who met the eligibility criteria at Wave 6 was deemed an OSM.

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 the former BHPS sample, OSMs are those who were enumerated at the first wave of the sample from which they come (Wave 1 for the original sample, Wave 9 for the Scotland and Wales boost samples, Wave 11 for Northern Ireland) or who were subsequently born to an OSM mother or father (or both). Following the incorporation into *Understanding Society* from Wave 2 onward, in the former BHPS sample, as for all other *Understanding Society* samples, only children born to an OSM mother will themselves become an OSM. OSMs, of all ages, are followed for interview and remain eligible as long as they are resident within the UK. They remain potentially eligible sample members for the life of Study.

The case may arise where the only OSM in the household is a child. Other household members are then TSMs so long as they are co-resident with the child, and therefore eligible for interview, even if the child is not yet old enough to be eligible for interview. If the OSM child moves house, they are followed to their new address and those living with the OSM child are eligible for interview. If the OSM child moves into an institution, where normally just the OSM/PSM would be interviewed and not co-residents, a split-off household is created containing only the OSM child and the household enumeration grid completed. The child OSM is an eligible sample member, even if they are not eligible for interview because of their age.

Temporary Sample Members (TSMs)

Any members of an enumerated household eligible for inclusion in the EMBS at Wave 1 who are not from a qualifying ethnic minority are Temporary Sample Members (TSMs) at Wave 1. This was the only category of TSM at Wave 1. Likewise, any members of an enumerated household eligible for inclusion in the IEMBS at Wave 6 who do not have a qualifying ethnic minority or immigration background (non-ethnic minorities who were born in the UK) were deemed to be TSMs at Wave 6.

In all parts of the sample, any new person found to be co-resident in an OSM or PSM household after Wave 1 is a TSM. This would include any child born to an OSM

father after Wave 1 but not an OSM mother and observed to be co-resident with the father (or any other OSM) at the survey wave following the child's birth. TSMs remain eligible for interview as long as co-resident in an OSM/PSM household. TSMs who are not co-resident in an OSM/PSM household are not followed and become ineligible for interview. TSMs are identified as re-joiners if they are subsequently found in an OSM/PSM household and then become eligible for interview.

Permanent Sample Members (PSMs)

PSMs are TSMs who are followed for interview after they no longer live with an OSM. This is done for substantive research reasons because of the additional contextual information they may provide for the analysis of OSMs. At present, there is only one category of PSM, but others may be defined in the future. Any TSM father of an OSM child born after Wave 1 and observed to be co-resident with the child at the survey wave following the child's birth is a PSM. PSMs remain potentially eligible for interview for the life of survey.

Further readings on Understanding Society sample design

- Lynn (2009) "Sample design for Understanding Society". Understanding Society Working Paper 2009-01, Understanding Society Working Paper Series No. 2009-01
- Berthoud et al. (2009) "Design of the Understanding Society ethnic minority boost sample". Understanding Society Working Paper Series No. 2009-02
- McFall et al. (2012) "Implementing the Biosocial Component of Understanding Society – Nurse Collection of Biomeasures" Understanding Society Working Paper Series No. 2012-04
- 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.