

---

# Chapter 1

## Sampling

---

The English Housing Survey (EHS) consists of two main elements: an interview survey of approximately 13,300 households and a follow up physical inspection of the dwelling of 6,000 of the participating households together with an inspection of around 200 vacant dwellings. This chapter provides information on how the household and dwelling samples were selected for inclusion in the interview and physical inspection elements of the 2015-16 EHS.

### Overview

- 1.1 As in previous years, addresses for the initial 2015-16 EHS sample were selected using a systematic random sample design. Interviews were attempted at all of these addresses. This is referred to throughout this report as the 'interview survey sample'.
- 1.2 The design delivers a representative sample of households in England for 2015-16, with the sample being unclustered when combined with the 2014-15 sample.
- 1.3 In addition to the interview survey sample, a sub-sample of addresses was selected for physical inspection. This sub-sample included both occupied and vacant dwellings and is referred to as the 'physical survey sample'. To ensure that a sufficient number of rented properties were included in the dwelling sample, all rented properties were selected for physical inspection while around half of owner occupied dwellings were randomly selected for a physical inspection (the percentage selected varied each quarter, ranging from 51% to 55%).
- 1.4 In 2015-16, 26,260 addresses were issued to interviewers. A small proportion of these (898 addresses) were found to be ineligible because they fall outside the scope of the EHS, they include addresses that were found to be commercial premises, second and holiday homes, demolished properties or instances where the address was not found.<sup>1</sup> Productive interviews were

---

<sup>1</sup> These categories of ineligible cases are those that do not fall within the scope of the English Housing Survey - second homes, demolished addresses, commercial addresses or instances where the address is not found. When calculating ineligible cases for the purpose of response rates we include cases that fall within the scope of the EHS, i.e. cases with an identifiable primary residential address, but are ineligible for an interview or survey, such as vacant properties and derelict properties (both ineligible for household interview but eligible for a physical survey), institutions and addresses under construction. See Annex Table 4.1. for details of cases ineligible for interview/survey.

---

achieved at 13,468 of the eligible addresses. Just under three-quarters of the productive addresses (10,469) were passed to surveyors, and physical surveys were achieved at 6,248 of these.

## Interview survey sample

- 1.5 The requirement for the 2015-16 EHS was to achieve 13,300 interviews across the four housing tenure types (i.e. owner occupied, private rented, local authority and housing association).

### Initial sample

- 1.6 To achieve these interviews, an initial sample of 42,000 addresses was drawn from the Postcode Address File (PAF). These addresses were drawn as a systematic two-stage random sample from the Royal Mail's Small User PAF.
- 1.7 The EHS sample is designed so that fieldwork takes place in one half of England in one year and the other half in the following year. (The design is referred to as the 'Half-England Model'.) For the first stage of sampling, the whole of England was grouped into 1,808 geographic 'merged areas' (clusters). Those areas were generated by combining neighbouring (18 on average) Lower Layer Super Output Areas (LSOAs) within region, so that each cluster contained about 12,500 addresses (with a range from about 11,200 to 13,800).
- 1.8 The 1,808 clusters were stratified by region, Census estimates of the percentage of dwellings in the cluster that are owner occupied (in tertiles within region) and the percentage of owner occupied dwellings (for the tertile with the highest percentage of owner occupied dwellings) and Census estimates of the proportion of households with a Household Reference Person (HRP) who works in non-manual occupations (for the other two tertiles).
- 1.9 The clusters were then randomly allocated to years, by assigning them in pairs going down the list, and then randomly (and independently) allocating one of each pair to "Year 1" and the other to "Year 2" (i.e. 904 clusters per year). The allocation of clusters to years by the 'Half-England Model' was fixed since 2012-13: "Year 1" clusters were used for 2012-13 and 2014-15 and "Year 2" clusters for 2013-14 and 2015-16.
- 1.10 At the second stage of sampling, all addresses in the PAF in the 904 clusters selected for the 2015-16 sample were stratified by cluster, LSOA ordered by Census estimates of the percentage of owner occupied dwelling, Census output area and postcode. Then a systematic random sample of 42,000 initial addresses was selected.

- 1.11 The advantage of using a two-stage approach involving the clusters is that it reduces the fieldwork area to half the country so interviewer and surveyor travel time and costs are reduced. The disadvantage is that for any single survey year the survey is partially clustered, which results in a small loss<sup>2</sup> in statistical efficiency. However, when analysing any two years of survey data, the combined sample is entirely unclustered.

### Sample issued to interviewers

- 1.12 The social rented sector (local authority and housing association) is less prevalent compared with the other sectors. To ensure that there is a sufficiently large sample of social renters for analysis, the 42,000 addresses were sub-sampled to sift out about half of the owner occupiers and private renters. The sub-sampling was carried out based on the 'predicted tenure' of the sampled addresses derived from the predominant tenure within the postcode that contained that address. Predominant tenure was identified using Experian's Residata<sup>3</sup> classifications and attached to the address records. Sub-sampling was carried out by grouping the addresses into the four housing tenure types and sub-sampling at the rates of 55% for owner occupied and private rented and 97% for social rented. That process produced 26,260 addresses for issuing to interviewers, Table 1.1.

**Table 1.1: Sub-sampling of PAF addresses, 2015-16**

	<b>PAF sample</b>	<b>Sub-sampling rate</b>	<b>Issued EHS sample</b>
	<i>addresses</i>	<i>percentages</i>	<i>addresses</i>
<b>predominant tenure</b>			
owner occupied	26,642	54.6%	14,553
private rented	7,496	54.7%	4,104
social rented	7,712	96.7%	7,457
unknown tenure	150	97.3%	146
<b>Total</b>	<b>42,000</b>	<b>62.5%</b>	<b>26,260</b>

<sup>2</sup> NatCen estimates that the maximum design factor due to clustering would be about 1.08.

<sup>3</sup> Experian possess a database that contains information obtained from a number of sources including insurance companies, Census, etc. referred to as Residata. It is from this that information was taken on predominant tenure within a postcode as well as other information. The matching of the EHS sample to Residata was carried out by BRE.

1.13 Interviews were achieved at 13,468 households, Table 1.2.

**Table 1.2: Number of interviews achieved, 2015-16**

<b>tenure</b>	
owner occupiers	7,748
private rented	2,061
local authority	1,536
housing association	2,123
<b>Total</b>	<b>13,468</b>

## Physical survey sample

1.14 The requirement for the 2015-16 physical survey sample was 6,200 physical surveys across the four housing tenures.

1.15 To ensure that the EHS delivers findings on renters that have the same level of precision as its predecessor, the English House Condition Survey, a disproportionate number of renters is included by under-sampling owner-occupied dwellings, Table 1.3.

**Table 1.3: Tenure distribution of achieved physical survey sample compared with the national stock**

<b>tenure</b>	<b>Achieved sample</b>		<b>National stock</b>
	<i>number</i>	<i>percentages</i>	<i>percentages</i>
owner occupiers	2,474	39.6%	62.5%
private rented	1,227	19.6%	20.4%
local authority	1,079	17.3%	6.9%
housing association	1,468	23.5%	10.2%
<b>Total</b>	<b>6,248</b>	<b>100.0%</b>	<b>100.0%</b>

1.16 The issued sample for the physical survey was drawn as a stratified sub-sample of the dwellings of those households who responded to the interview survey, together with a stratified sub-sample of dwellings found to be vacant during fieldwork. Calculation of the size of the sample to be issued took account of the expected physical survey response rates by tenure.

1.17 The sub-sampling of interview survey cases for the physical survey was carried out during the interview by the Computer-Assisted Personal Interviewing (CAPI) program. The program used the tenure of the dwelling established at the interview to select the subsample and advised the interviewers if the household they were interviewing was eligible for the

physical survey. If so, they would attempt to gain agreement from respondents to take part in the physical survey and pass on the address details to CADS Housing Surveys who managed the fieldwork of the physical surveys.

- 1.18 Different sub-sampling rates were applied to each tenure group to identify cases eligible for the physical survey. As the owner occupied sector is larger than the rented sector, it was under-sampled to ensure sufficient numbers of renters for analysis. Sub-sampling rates were reviewed at the start of the year and were kept under review throughout the year in order to ensure a sufficiently large sample was achieved. The sub-sampling rates for 2015-16 are shown in Table 1.4.

**Table 1.4: Sub-sampling rates, 2015-16 physical survey**

	Sub-sample rates			
	Quarter 1 <sup>4</sup>	Quarter 2	Quarter 3	Quarter 4
	<i>percentages</i>			
<b>tenure</b>				
owner occupiers	-	51.0%	51.0%	55.0%
private rented	-	100.0%	100.0%	100.0%
local authority	-	100.0%	100.0%	100.0%
housing association	-	100.0%	100.0%	100.0%

- 1.19 To ensure that there were roughly equal numbers of surveys across the quarters to reduce the impact of seasonality (for example, damp problems are more likely to be identified in the winter), the sub-sampling rate was kept fairly constant. However, the sub-sample rate for owner occupiers was increased from 51% to 55% in quarter 4 to ensure the overall target for 6,200 physical surveys across the four housing tenure groups was achieved given the response rates in previous quarters.
- 1.20 Vacant properties were sub-sampled at the same rates as occupied cases based on information about their last known tenure. This information was gathered by interviewers as part of their initial visit (from talking to the landlord or neighbours) or based on the interviewer's best estimate of tenure derived from available evidence. Permission and access for the survey was then sought by the surveyors managed by CADS Housing Surveys.
- 1.21 Not surprisingly, a lower proportion of full physical surveys were obtained in unoccupied dwellings, compared with occupied dwellings, because of the difficulty in gaining access to a property that was unoccupied. In 2015-16, surveyors managed to gain access and obtain full physical surveys in 28% of dwellings that were unoccupied at the time fieldwork took place.

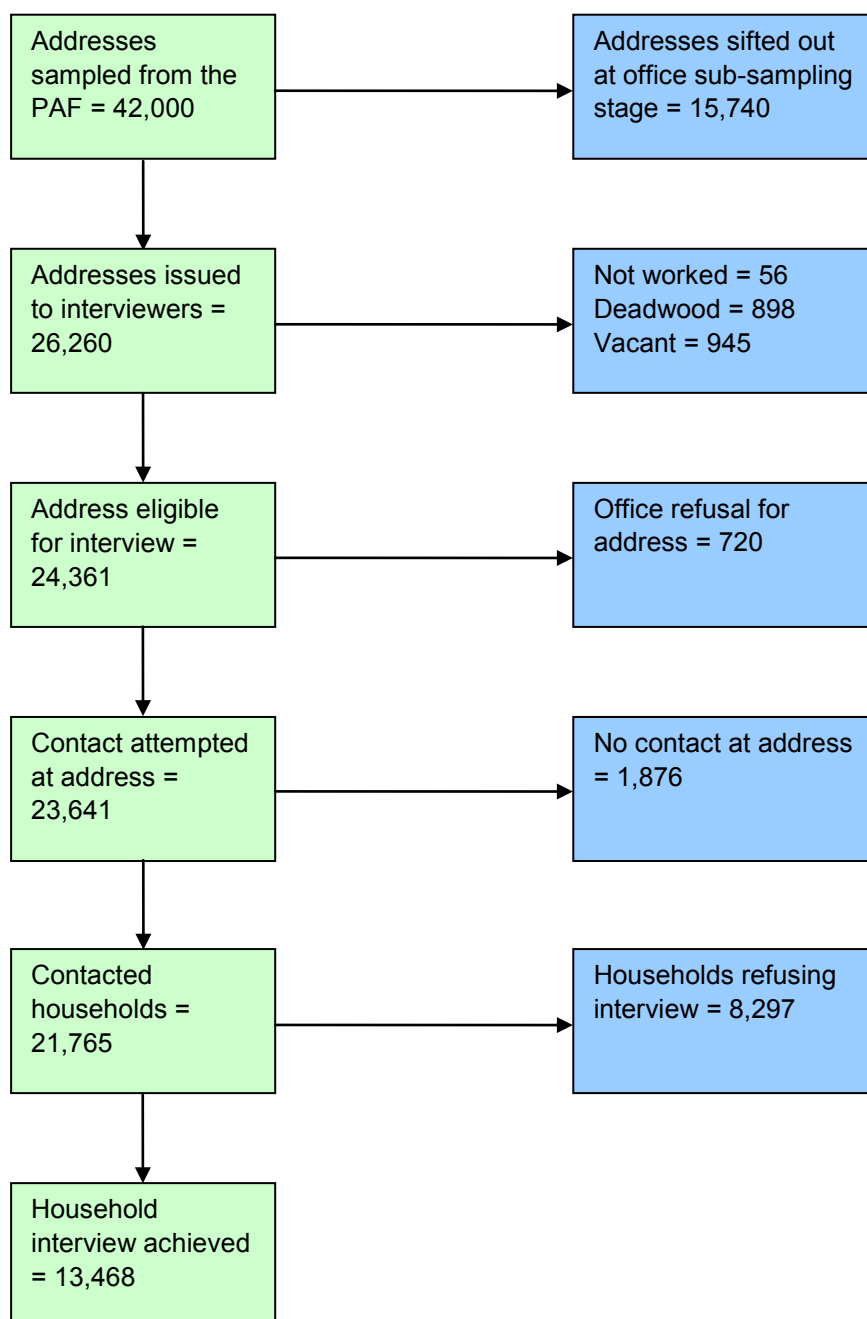
<sup>4</sup> Fieldwork for EHS 2015-16 started at Quarter 2 and was constrained to 3 quarters.

---

1.22 The 2015-16 sampling and response process is summarised in Figure 1.1.

**Figure 1.1: Sample structure of the EHS, 2015-16<sup>5</sup>**

**Interview survey sample**



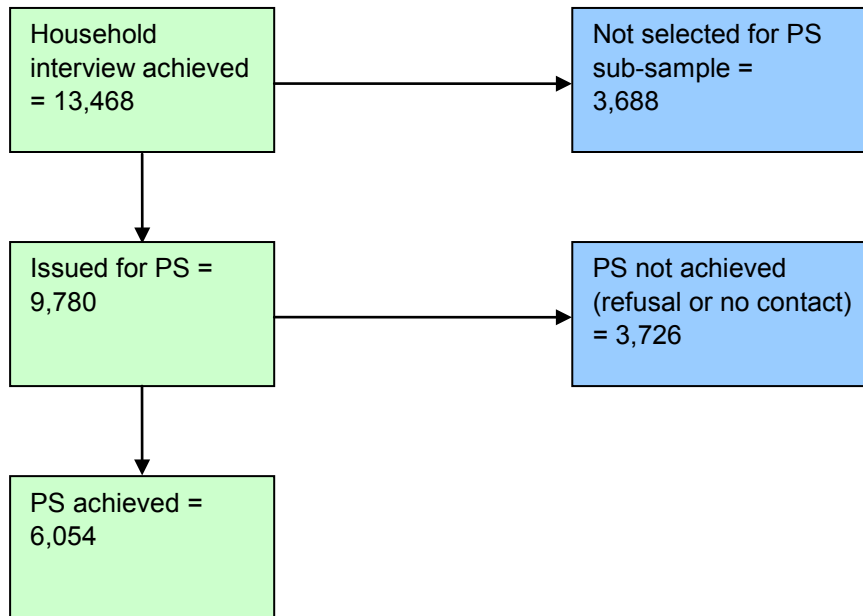
<sup>5</sup> Different concepts are required for sampling and response calculation so the way case outcomes are grouped for sampling purposes differs from the approach used for calculating response rates in three:

- 1) Categories 'unable to locate address' and 'not worked' are included under in-scope addresses for the purposes of response rate calculations, but grouped within ineligible in the above.
- 2) In the diagram above 'no contact' includes categories 'unknown whether residential: no contact', 'residential but unknown eligibility: no contact', and 'other unknown eligibility'. These are treated as a separate category 'unknown eligibility' for the purposes of response rate calculations.
- 3) 'Households refusing interview' above includes other unproductive categories including 'language difficulties', 'lost productive' and 'away in hospital during fieldwork period'. These are treated as a separate category 'other unproductive' for the purposes of calculating response rates.

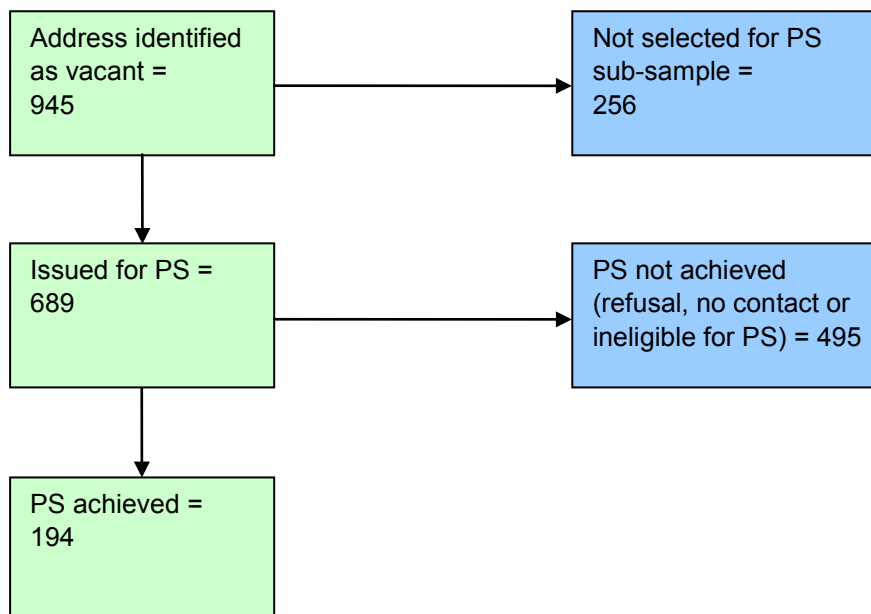
Details of household interview outcomes are presented in Annex Table 4.1.

---

### Physical survey sample: occupied dwellings



### Physical survey sample: vacant dwellings





---

## Sampling at addresses

- 1.23 Most addresses contained a single dwelling and a single household. However, at a small proportion of addresses (less than 1%) this was not the case. Interviewers on the 2015-16 EHS used standard procedures to select one dwelling and/or one household at random when more than one was identified.
- 1.24 The interviewer listed the dwellings identified at multi-dwelling address and then randomly selected one from the list, using a pre-selected random number. The random number was obtained from a sheet (called a Kish grid) which had a column for the number of dwellings identified, and a column for which numbered dwelling to select.
- 1.25 The same procedure was used to select the household to interview when more than one was identified at a dwelling.