Chapter 6
Weighting

The English Housing Survey (EHS) is weighted to take account of the over-sampling of the less prevalent tenure groups and differential non-response, in order to provide unbiased national estimates. This chapter provides details of the weighting methodology and gives advice on which sets of weights to use when conducting analysis using EHS data.

Overview

6.1 The following weights have been calculated for the 2013-14 EHS data:

- Household weights for the full sample interview survey 2013-14 – aagfh13
- Average dwelling weights for cases in 2012-13 and 2013-14 that had both the interview and physical survey conducted (paired cases) [1] – aagpd1213
- Average household weights for cases in 2012-13 and 2013-14 that had both the interview and physical survey conducted (paired cases) [1] – aagph1213

6.2 These weights adjust the sample to correct for the over-sampling of the less prevalent tenure groups and reduce the bias from differential non-response. The resulting weights sum to estimated population totals, enabling the survey to provide estimates of the total population of dwellings and households in England.

6.3 The weighting of the 2013-14 EHS data was undertaken by NatCen who managed the survey on behalf of the Department for Communities and Local Government (DCLG).

[1] Cases which were eligible for PS (the paired sample) included addresses which the interviewers had determined to be vacant and where a physical survey had also been achieved.

Application of weights during analysis

6.4 The EHS comes with its own weights covering the household and dwelling sample of cases.

6.5 The weight for the household sample can be found in the file generalfs13.sav and is called agfh13. This should be used for any analysis for which the aim is to provide estimates of households, based on the interview survey data.

6.6 The weights covering the dwelling sample of cases can be found in the file general 12+13.sav. Aagpd1213 should be used for any analysis in which the aim is to provide estimates of dwellings and that includes physical survey data (e.g. percentage of non-decent dwellings) while aagph1213 should be used for any analysis in which the aim is to provide estimates of households and that includes interview survey data (e.g. percentage of households in non-decent dwellings). The weights can only be used on the full 2-year dataset. They cannot be used on the data split into separate years.

6.7 The recommended application of weights is summarised in Table 6.1.

Table 6.1: Application of weights during analysis

Weight	Description	Base
aagfh13	Household weights for the full sample interview survey 2013-14	13,276
aagpd1213	Average dwelling weights for cases in 2012-13 and 2013-14 that had both the interview and physical survey conducted (paired cases)	12,498
aagph1213	Average household weights for cases in 2012-13 and 2013-14 that had both the interview and physical survey conducted (paired cases)	12,008

Summary of weighting methodology

6.8 The weighting methodology[2] used a sequence of stages[3] described below. Some stages correct for the disproportionate sampling by tenure; others for the differential non-response in the Interview Survey (IS) and Physical Survey (PS) response process.

[2] The current weighting methodology was introduced in 2013-14 as a result of a weighting review which was carried out after the 2012-13 survey. The results of the review are published here: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/406771/EHS_2013-14_weighting_methodology_paper_FINAL.pdf.

[3] To avoid confusion with the numbering of stages under the previous methodology and for consistency with the weighting review report, the prefix "N" is used to describe the stages.

6.9 The following stages were applied to both the full household sample and the dwelling sample:

- Stage N1: Calculating the probability of selection for addresses
- Stage N2a: Adjusting for not-worked cases
- Stage N2: Filtering out office refusals
- Stage N3: Contact at IS
- Stage N4: Cooperation at IS
- Stage N5: Address-to-dwelling ratio
- Stage N6: Dwelling-to-household ratio

6.10 The following stage was applied to the full household sample:

- Stage N7: Calibration weighting for IS (full household weights)

6.11 The following stages were applied to the dwelling sample:

- Stage N8: Sub-sampling by tenure for PS
- Stage N9: Cooperation at PS
- Stage N10: Calibration weighting for PS
- Stage N11: Adjustment for new build (final paired dwelling weights)
- Stage N12: Creation of final paired household weights

6.12 Stages N1 to N6, N8 to N9 and N12 were implemented using a set of SPSS syntax scripts in combination with the specialist SPSS module AnswerTree. Stages N7, N10 and N11 used a combination of SPSS syntax and STATA's calibration command.

6.13 The weighting stages are described in more detail below.

Weighting the full household sample

6.14 Stage N1 generated the selection weights for the issued sample of addresses. This had two components: a grossing constant equal to the total number of delivery points on the Postcode Address File (PAF) divided by the number of delivery points sampled; and, to allow for the under-sampling of properties that were likely to be owner-occupiers, a correction factor to account for the unequal selection of addresses within each predicted tenure type (within each

quarter). That correction factor was equal to the number predicted to be tenure X in the PAF divided by the number selected in predicted tenure X.

6.15 A very small number of the addresses sampled were not issued to interviewers for operational reasons. Not-worked addresses were filtered out at stage N2a by applying an adjustment to the Stage N1 weights so that the distribution of the weighted worked addresses by region matched that based on all the issued addresses.

6.16 Non-response at the key fieldwork stages (refusal to co-operate prior to the interview, non-contact at the interview and refusal to co-operate at the interview) do not happen completely at random and the factors associated with each of these three processes may differ. Stages N2, N3, and N4 made a series of adjustments for non-participation. The adjustments were were made using weighting classes (groups of cases) generated by the CHAID[4] algorithm of the SPSS AnswerTree software. The models generated from the algorithm were developed to identify the factors (or predictor variables) that were significantly associated with each of the three causes of non-response. These classes were derived at each stage using an unweighted CHAID model at address level which partitioned the sample of occupied dwellings based on predictor variables significantly associated with the propensity to respond at each of the key fieldwork stages above. Typical predictor variables for stage N2 included geographical area, predominant tenure, dwelling age and dwelling type in the area, and urban/rural classification; for subsequent stages, information collected by the interviewer was also used.

6.17 Once the classes had been generated, the adjustment within each class was made based on the weighted estimates (using the combined weights up to that stage). The main reason for doing the correction in stages was to use any additional information available at each stage.

6.18 The EHS analyses are concerned with dwellings and households rather than addresses, and there is not always a one-to-one relationship between an address, a dwelling, and a household[5]. Usually there is only one dwelling at each address sampled from the PAF, but addresses are occasionally found to cover more than one dwelling (for example if a house has been converted into self-contained flats) or only part of a dwelling (for example a bedsit which shares facilities with a household at a separate postal address). As only one dwelling was selected at these addresses and one household was selected at dwellings that contain more than one household, weights were required.

6.19 Where an address refers to more/less than one dwelling, each dwelling at that address would have a lower/higher chance of selection. Similarly, when a dwelling contains more than one household, each household at that dwelling

[4] Chi-squared Automatic Interaction Detector
[5] For the purposes of the survey, a dwelling is defined as ‘a self-contained unit of accommodation where the occupants of that accommodation have sole use of all the rooms and facilities’.

would have a lower chance of selection. Stages N5 and N6 corrected for the selection of one dwelling unit at the sampled address and one household at the selected dwelling by deriving address-to-dwelling and dwelling-to-household ratios designed to be applied with the weights developed further downstream to derive the final weight. The address-to-dwelling[6] and dwelling-to-household ratios were averaged (smoothed) within weighting (smoothing) classes defined by tenure and region to remove large peaks and troughs within the classes.

6.20 The previous stages account for the sampling and response probabilities. Applying the weights derived from Stages N1 to N5 to the household-level data would provide a preliminary survey estimate of the total number of households in England. However, this will differ from the true value because of sampling error, under-coverage of the frame and inability of the model-based process to all allow for all the factors associated with non-response. Estimates for subgroups such as tenures will differ from their true values for the same reason. These differences in the survey estimates can be reduced by calibrating the weights.

6.21 Calibration takes an initial set of weights (e.g. the composite weights from stages N1 to N5) and then adjusts (or calibrates) it to given control totals. The control totals are usually a population count of a specific attribute or set of characteristic derived from a source external to the survey. The process generates weights which produces survey estimates that exactly match the population for the specific characteristics (the control totals) used in the adjustment.

6.22 The calibration of the interview sample was carried out on the part of the sample that contained occupied dwellings only. That section of the sample was first weighted by the composite weights from stages N1 to N5 and then calibrated to the control totals below.

- population proportions for age/sex based on ONS population projections (for dwellings with more than one household, the household counts were adjusted using the dwelling-to-household ratio from Stage N6);
- counts of occupied dwellings by tenure for each region (these control totals were estimated by adjusting the DCLG dwelling counts, which include both occupied and vacant dwellings, using estimates of the proportion of dwellings that are occupied from the current and four previous years of the EHS).
- The control totals were measured as at 1 October 2013.

[6] Apart from occupied dwellings, the address-to-dwelling ratio was also computed for vacant dwellings as this was needed for the sub-sampling stage (stage N8).

6.23 The calibration weights from Stage N7 was combined with the dwelling-to-household ratio from stage N6 to give the weights for the full household sample.

Weighting the dwelling sample

6.24 A lot of analyses using the EHS data are carried out using the dwelling as the unit of analysis. These analyses usually make use of cases containing both the interview and physical survey data (or cases in the paired sample). Cases which were eligible for PS (the paired sample) included addresses which the interviewers had determined to be vacant and where a physical survey had also been achieved. The inclusion of vacant addresses was thus necessary for the production of the final paired dwelling weights.

6.25 The weighting process started by using stages N1 to N5 as described above.

6.26 The PS sample included a disproportionately larger number of dwellings from renting tenure groups to enable detailed analysis of these. This was achieved by under-sampling of properties likely to be owner occupied (sub-sampling rates varied by quarter). Stage N8 calculated selection weights to take account of the under-sampling of owner occupied properties.

6.27 Stage N9 adjusted for non-response to PS using weighting classes generated from a CHAID model designed to identify the factors significantly associated with non-response to the physical survey. Data collected during the interview survey (for occupied dwellings only) were also used to help determine the weighting classes. Vacant cases were treated using a separate CHAID model because the process of participation for vacant dwellings is generally different from that for occupied dwellings and this needed to be reflected in the weights.

6.28 The occupied paired sample (i.e. cases with both IS and PS) was weighted by the composite weights from stages N1 to N5 and N8 to N9. The vacant sample (i.e. vacant addresses at PS) was weighted by the composite weights from stages N1, N2a, N5, N8, and N9. This gives the initial weights for dwellings.

6.29 At stage N10, both the occupied and vacant samples were then combined and calibrated to:

- population proportions for age/sex (for dwellings with more than one household, the household counts were adjusted using the dwelling-to-household ratio from Stage N6; for vacant dwellings the household counts were all zero);

- DCLG counts of all dwellings (occupied + vacant) by tenure for each region;

- counts of vacant dwellings by tenure (RSL and LA collapsed) for each region (these control totals were estimated by adjusting the DCLG dwelling counts using estimates of vacancy rates from the current and four previous years of the EHS).

6.30 The achieved sample of dwellings did not include any dwellings built since the sample was drawn; stage N11 adjusted the weights to allow for those new dwellings. Because of the small number of recently-built dwellings in the survey, the weights of all cases with a construction date of 1990 onwards were weighted-up to cover for new addresses on the PAF by re-calibrating to the same overall control totals, but with re-estimated counts of post-1990 build. This was carried out separately for areas with a high/low rate of new build, and for private/social sector housing (excluding local authority housing because the rate of new building in this sector is negligible).This process generated the final paired sample dwelling weights.

6.31 At Stage N12, the final paired dwelling weights from Stage N11 (after removing the vacant dwellings) were adjusted using the dwelling-to-household ratio from stage N6 to generate the final paired sample household weights.

Calculating two year weights

6.32 Because of the smaller annual sample sizes involved, analysis of the dwelling sample is normally carried out using 2 years' weighted data. This section sets out how the combined weights were calculated.

6.33 The individual year (2012-13 and 2013-14) datasets with the dwelling weights after the PS cooperation stage (stage N9 in 2013-14 and stage 10 in 2012-13) were merged together and two-year dwelling weights were created by dividing each year's weights by 2, so that each dataset has equal influence on the weighting process.

6.34 The two-year dwelling weights were calibrated by repeating stages N10 and N11 of the annual weighting[7]. This process generated the final two-year paired sample dwelling weights.

6.35 Similarly to stage N12 of the annual weighting, the final two-year paired dwelling weights (after removing the vacant dwellings) were adjusted using the dwelling-to-household ratio to generate the final two-year paired sample household weights.

[7] The reference date for the 2-year weights was 1st April 2013, therefore the estimated DCLG dwelling counts used as control totals were adjusted accordingly.