



Sampling and weighting

English Housing Survey technical advice note



Sampling and weighting

English Housing Survey technical advice note

Department for Communities and Local Government
Eland House
Bressenden Place
London
SW1E 5DU
Telephone: 030 3444 0000
Website: www.communities.gov.uk

© Crown Copyright, 2010

Copyright in the typographical arrangement rests with the Crown.

This publication, excluding logos, may be reproduced free of charge in any format or medium for research, private study or for internal circulation within an organisation. This is subject to it being reproduced accurately and not used in a misleading context. The material must be acknowledged as Crown copyright and the title of the publication specified.

You may re-use this information (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence. To view this licence, visit <http://www.nationalarchives.gov.uk/doc/open-government-licence/> or write to the Information Policy Team, The National Archives, Kew, London TW9 4DU, or e-mail: psi@nationalarchives.gsi.gov.uk.

If you require this publication in an alternative format please email alternativeformats@communities.gsi.gov.uk

DCLG Publications
Tel: 030 0123 1124
Fax: 030 0123 1125

Email: product@communities.gsi.gov.uk
Online via the website: www.communities.gov.uk

December 2010

ISBN: 978 1 4098 2713 9

Contents

Introduction

Overview of sampling and weighting

Survey samples

- Interview survey sample

- Physical survey sample

Weighting methodology

- Overview

- Adjusting for the relationship between addresses and dwellings

- Weighting the full household sample

- Weighting the dwelling sample

- Calculating the two-year dwelling sample weights

Differences from the former SEH and EHCS

Introduction

1. This document is one in a series of Technical Advice Notes providing background information about the survey methodology to assist users in their analysis and interpretation of the survey findings.
2. This note provides information on how the samples of households and dwellings were selected for inclusion in the **interview** and **physical survey** elements of the English Housing Survey (EHS) in 2008-09. It also provides information on the weighting methodology used to enable national-level estimates to be obtained from the survey results for both the full household sample and the dwelling sample. This information will be updated periodically as necessary to cover any major changes occurring in subsequent years.

Overview of sampling and weighting

3. As explained in Technical Advice Note: Survey Overview and Methodology, two key elements of the EHS fieldwork are:
 - An achieved interview survey of about 17,000 households per year
 - An achieved physical survey of about 8,000 occupied and unoccupied dwellings per year.
4. Sampling and weighting for both these survey components are carried out by the Office for National Statistics (ONS) who manage the survey on behalf of the Department for Communities and Local Government (DCLG). ONS also undertake the interviews with householders as well as the interview data validation processes. Physical surveys are carried out by a large field force of professional surveyors employed by Miller Mitchell Burley Lane, who work in close co-operation with the interviewers from ONS to maximise response to the survey and deliver high quality data.
5. DCLG require the sample issued for interview to be a random sample of households in England. The cases for which an interview is obtained are known as the **full household sample**. For the physical survey, they require that a disproportionate number of the achieved sample of dwellings should come from properties rented from local authorities and housing associations, in order to meet set tenure targets and provide sufficient information on rarer tenures. All occupied dwellings in the physical survey are properties where an interview survey has been conducted previously. Vacant properties are also drawn from those addresses found to be vacant at the interview stage. The subset of cases for which a physical survey is achieved form the **dwelling sample**, and are also sometimes referred to as 'paired cases' because data are available on them from both the interview and physical surveys.

6. In 2008-09, 32,065 addresses were issued to interviewers. A small proportion of these were found to be ineligible, including addresses that were found to be commercial premises, second and holiday homes or demolished properties. Of the remainder, interviews were achieved at 17,691 addresses (the full household sample). A further 1,390 addresses were found to relate to vacant dwellings. After subsampling, 12,129 dwelling addresses were passed to surveyors, and physical surveys were achieved at 7,972 of these (the 2008-09 component of the dwelling sample).
7. For analysis purposes, weights are calculated for the full household survey cases achieved in one year of the survey, but due to the smaller sample for which physical surveys are achieved, two years' dwelling sample cases are weighted together.

Survey samples

Interview survey sample

8. DCLG require sufficient household interviews to achieve the same level of precision as delivered through the SEH (between 19,000 and 20,000 household interviews, through a clustered sample design). The EHS sample is unclustered, which allows for a reduction in the size of the achieved household interview sample to 17,000 while maintaining the previous level of precision and providing sufficient candidate cases for the physical survey. Calculation of the size of the sample to issue takes account of the expected interview survey response rates by tenure as well as the proportion of ineligible addresses on the Postal Address File.
9. In order to obtain the required achieved sample of about 17,000 households, an initial sample of some 32,000 addresses is drawn from the Royal Mail's Small User Postal Address File (PAF), sorted by postcode. The sample is drawn as a systematic random sample to ensure that the distribution of the sample across local authorities (LAs) is close to that occurring in the PAF.
10. Interviewers working on the EHS also work on other surveys conducted by ONS, of which the largest is the Labour Force Survey (LFS). Addresses sampled for the EHS are allocated to fieldwork quarters in such a way as to create groups of addresses within LFS interviewer areas. Interviewing then takes place during the fieldwork period of each quarter, as described in Technical Advice Note: Survey Overview and Methodology.

Physical survey sample

11. DCLG also require the delivery of sufficient paired cases with physical surveys to deliver the same precision as achieved by the EHCS (8,000

cases a year through a clustered sample design). This sample is stratified to ensure disproportionate numbers of the rarer tenures (social renters). The alternative to this approach would be to issue a very much larger random sample and carry out unnecessary owner occupier and private renter surveys in order to accumulate sufficient social renting cases.

12. The table below compares the target paired case annual tenure distribution with the national stock:

Tenure distribution of target achieved physical survey sample compared with the national stock

Tenure		Target sample (%)	National stock (%) ^{1,3}
Owner-occupied	4,000	50	68
Private rented	1,000	12	14
Local authority	2,000	25	
Registered social landlord	1,000	12	18
Total²	8,000	100	100

Notes:

1. Taken from Table 1.01 Trends in tenure, Survey of English Housing 2007

2. Percentages may not add up to 100 per cent due to rounding

3. The split between local authority and registered social landlord is not available

Size of physical survey sample to issue

13. The issued sample for the physical survey is drawn as a stratified sub-sample of the dwellings of those households who respond to the interview survey, together with a stratified subsample of dwellings found to be vacant at that stage. Calculation of the size of the sample to be issued takes account of the expected physical survey response rates by tenure.

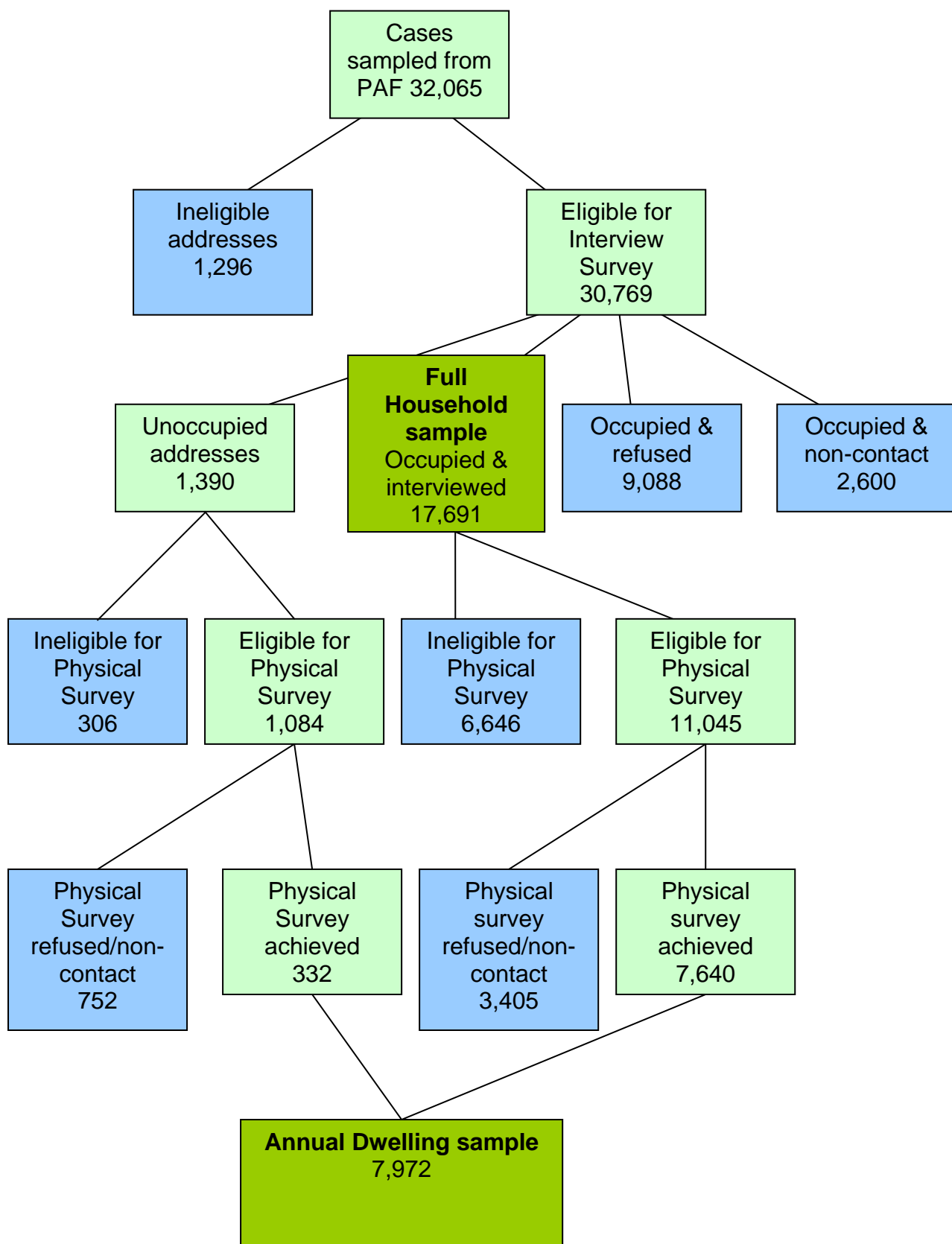
Subsampling cases for physical survey

14. The sub-sampling of interview survey cases for the physical survey is carried out in the field during the interview, using the tenure of the dwelling established during the interview. Different sub-sampling rates are applied to each tenure group to identify cases eligible for the physical survey. These rates can be varied quarterly if required in order to achieve the required annual total sample.
15. Interviewers are advised via the CAPI (computer-aided personal interviewing) instrument if the household they are interviewing is eligible for the physical survey, and they then attempt to gain agreement from respondents to take part before passing on the

address details to MMBL.

16. For vacant properties, cases are selected for physical survey based on the interviewer's best estimate of tenure using local enquiries and using the same sub-sampling rates as for occupied cases. Permission and access for the survey is then sought by the MMBL surveyors.
17. The figure below shows details of the numbers of cases going forward to each successive stage of the survey process in 2008-09.

Figure: Sample structure of 2008-09 EHS



Weighting methodology

Overview

18. The weighting methodology involves a series of steps, each of which is designed to take account of the selection and response processes involved in successive steps in the sampling and interview/ physical survey response process. These steps in the weighting should give us unbiased estimates for the populations of households and dwellings in England.
19. This method is very similar to that of the former Survey of English Housing (SEH), the main difference being that much more detailed bias adjustment is carried out in the EHS. The weighting process for a single year's data is broken down into fifteen stages:

Stages 1 and 2 apply to both the full household sample and the dwelling sample and adjust for:

Stage 1: Address selection probabilities

Stage 2: Address to dwelling relationship

Stages 3 – 6 apply to the full household sample and adjust for:

Stage 3: Dwelling to household relationship

Stage 4a: Accounting for office refusals (refusal to cooperate prior to the interview)

Stage 4: Probability of contact at the interview survey

Stage 5: Probability of response at the interview survey

Stage 6: Calibration weighting for the interview survey – matching to known population totals as at 1 October of the survey year

Stages 7a – 12 apply to the dwelling sample and adjust for:

Stage 7a: Office refusals – refusal to cooperate prior to contact at the dwelling

Stage 7: Probability of contact at the dwelling

Stage 8: Probability of cooperation at the dwelling

Stage 9: Sub-sampling by tenure for the physical survey

Stage 10: Cooperation at the physical survey

Stage 11: Calibration weighting to dwelling totals, including an adjustment for new builds

Stage 12: Calibrating the household weights for the paired sample to those from the full household sample

Stage 12a: Recalibrating the dwelling weights from Stage 11 to be consistent with the household weights from Stage 12

Stages 1 to 5 and 7a to 10 are implemented using a set of SPSS syntax scripts in combination with the specialist SPSS module AnswerTree.

Stages 6, 11, 12 and 12a are referred to as calibration stages because they scale the weighted data to predetermined totals. These stages use the Generalised Estimation System (GES) macro implemented in SAS.

Once the results obtained at Stage 12a for a single year's dwelling sample have been checked and signed off, a further weighting phase repeats the methodology of stages 11 – 12a on the combined data from two years' dwelling sample data, to create the final dataset for analysis.

These processes are described in more detail below.

Adjusting for the relationship between addresses and dwellings

20. Stage 1 and 2 are common to the weighting of both the full household sample and the paired cases sample. Stage 1 takes account of the sampling fractions involved in drawing the equal-probability sample of addresses from the PAF. The weight for this stage is simply:

$$w_1 = \frac{\text{delivery points on PAF}}{\text{delivery points sampled}}$$

Subsequent stages involve multiplying these initial weights by successive adjustment factors.

21. 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. (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'). 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). The weighting methodology must take account of this.

22. Where an address refers to more/less than one dwelling, each dwelling at that address would have a lower/higher chance of selection. Stage 2 adjusts the weight calculated at Stage 1 to take account of the address to dwelling relationship for such cases.

Weighting the full household sample

23. Stage 3 makes a similar adjustment where the dwelling contains more than one household, and each household therefore had a lower chance of selection.
24. Refusal to co-operate prior to interview, non-contact at interview and refusal to co-operate at interview do not happen completely at random and the factors associated with each of these three processes may differ. Stages 4a, 4, and 5 each use available information about each case as predictor variables in a model to partition the sample into groups in order to describe as much variation in the response variable as possible. These models are constructed using the CHAID algorithm in the SPSS AnswerTree software. Typical predictor variables for Stage 4a are region; predominant tenure, dwelling age and dwelling type in the area; urban/rural classification. For subsequent stages, information collected by the interviewer is also used.
25. Response weights are calculated for each of the groups produced by AnswerTree at each of Stage 4a, 4 and 5.
26. The previous five stages account for the sampling and response probabilities. Applying the weight from Stage 5 to the household-level data would provide a 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 inadequacies in capturing the non-response mechanisms. 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 adjusting the weights so that the total final weights match certain control totals.
27. The control totals used for Stage 6 were based on ONS population projections by sex and age group and by region, derived from the 2001 Census, and tenure (owner-occupied, social sector and private rented sector) from the Labour Force Survey (LFS).

Weighting the dwelling sample

28. The approach to weighting the paired cases which make up the dwelling sample is dwelling-based rather than household-based so, in contrast to the weighting of the full household sample described above, all dwellings identified as vacant at which a physical survey was achieved are included in the sample to be weighted.
29. This process is more complex than that for the full household sample, partly because of the need to create internally consistent household and dwelling weights, and also because of the need to combine two years' dwelling sample data to obtain a large enough sample for analysis.
30. In outline, the approach is to first weight up the dwellings data for the current year to estimated dwelling controls by tenure, then to adjust this weighting so that the number of weighted households that result from it is consistent, within tenure and region, with the weighted full household sample. The data for two consecutive years are combined and reweighted to the reference date of 1 April midway between the two survey periods: this process is described in the next section.
31. The weighting process starts by using Stages 1 and 2, described above, to take account of the initial sampling fractions and the address to dwelling relationship.
32. Stage 7a then compensates for office refusals at either interview or physical survey stage. This process uses AnswerTree as in Stage 4a but with the inclusion of dwellings identified as vacant.
33. Stages 7 and 8 adjust for interviewer contact at the dwelling and for co-operation with the interview survey. All vacant dwellings are deemed to have been 'contacted' provided they have been located by the interviewer, and to have 'cooperated' with the interview survey phase with the exception of those in respect of which a refusal was sent to the Field Office.
34. As explained above, the dwelling sample is required to contain a disproportionate number of social rented properties. This is achieved by subsampling the issued sample during the interview survey once information about tenure, including tenure of vacant properties, has been collected. Only dwellings selected at this stage are eligible for the physical survey. Sub-sampling rates for the physical survey can also be varied between quarters: in 2008/09 they were changed between quarter 2 and quarter 3. Stage 9 calculates adjustments to the weights to take account of this subsampling.
35. Stage 10 adjusts for response to the physical survey, within weighting groups created by AnswerTree. For this stage, data collected during the interview survey are also used to help determine the weighting groups, and vacant cases are treated separately if the data show this

to give a better result.

36. As with the interview survey weighting, the initial weighting stages of the physical survey (Stages 1, 2, and 7a - 10) attempt to account for sampling and response probabilities and so the total weight of the data gives a survey estimate of the total number of dwellings in the population. However, this will differ somewhat from the true value because of factors such as sampling error, under-coverage of the frame and inadequacies in capturing the non-response mechanisms. These differences in the survey estimate can be reduced by adjusting the weights to match chosen control totals.
37. The calibration of the single-year weights to control totals was carried out in Stages 11 – 12a. At Stage 11, the weights from Stage 10 were first calibrated to DCLG dwelling estimates split by tenure and region, projected forward to the required 1 October reference date.
38. The achieved sample of dwellings do not include any built since the creation dates of the two PAF files from which the sample was taken, so the weights are adjusted using the number of new dwellings built between the PAF date and the reference date for the weighting. This process involves separate adjustments, for each PAF subsample, for:
- regions with a high/low rate of new build, and
 - private/social sector housing (excluding local authority housing)

Because of the small number of recently-built dwellings in the survey, the weights of all cases with a construction date of 1990 onwards are adjusted in this process. The calibration of the revised weights to tenure by region totals was then rerun.

39. At Stage 12, paired cases household weights were derived from the dwelling weights produced at the end of Stage 11, using the dwelling to household relationships found during the interview and physical surveys.
40. These weights were then recalibrated to give the same distribution of population by age and sex; region; and tenure, as the weights from the full household sample (at Stage 6).
41. At Stage 12a, the averaged adjustments made to each tenure by type by region group of household weights were then applied to the corresponding dwelling weights, including those for vacant dwellings. This ensured that the dwelling and household weights resulting from this stage were consistent for each dwelling case while ensuring that the household weights produced the same weighted totals as those for the full household sample.

Calculating the two year dwelling sample weights

42. As explained above, 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 combined weights were calculated for the '2008' dwelling sample.
43. This two year dataset combined the final EHCS dataset, for 2007-08, with the first EHS dataset, for 2008-09. There were differences in the weighting processes of the EHCS and EHS, not least because the EHCS had a longitudinal component. It was decided to combine the two datasets after the sampling and response adjustments and calibrate the combined dataset.
44. The new and longitudinal components of the EHCS dataset were re-calibrated separately to the control totals at the reference point of April 1 2008, with appropriate adjustments for newbuild dwellings. The new and longitudinal cases were then recombined in proportion to the sample sizes within tenure and GOR groups.
45. The EHS 2008/09 data were also reweighted to the 1 April 2008 reference point. The EHCS 2007-08 and EHS 2008-09 datasets were then merged together and the weights halved so that each dataset had equal influence.
46. Finally, the dwelling and household weights were scaled in the same manner as described above for the one-year paired cases dataset.

Differences from the former SEH and EHCS

47. The principal differences in the sampling and weighting methodology adopted for the EHS and that of its predecessors, the Survey of English Housing and the English House Condition Survey, are described in Appendix A to the EHS Household Report 2008/09 and the EHS 2008 Housing Stock Report respectively. See links below. Information on the impact of these changes is also provided.

<http://www.communities.gov.uk/publications/corporate/statistics/ehs200809householdreport>

<http://www.communities.gov.uk/documents/statistics/pdf/1750754.pdf>

Department for Communities and Local Government
© Crown Copyright, December 2010

ISBN: 978 1 4098 2713 9