



General Household Survey 2005

Sample Design and Response

Appendix B

© Crown copyright

Office for National Statistics
Government Buildings
Cardiff Rd
Newport
NP10 8XG
Tel: 01633 655740
Email: ghs@ons.gsi.gov.uk

Sample design and response

In 2005, 16,560 addresses were sampled. The GHS aims to interview all adults aged 16 or over at every household at the sampled address¹. It uses a probability, stratified two-stage sample design. The Primary Sampling Units (PSUs) are postcode sectors, which are similar in size to wards and the secondary sampling units are addresses within those sectors.

Sample design

The revised 2000 survey design introduced new stratifiers². Stratification involves the division of the population into sub-groups, or strata, from which independent samples are taken. This ensures that a representative sample will be drawn with respect to the stratifiers (i.e. the proportion of units sampled from any particular stratum will equal the proportion in the population with that characteristic). Stratification of a sample can lead to substantial improvements in the precision of survey estimates. Optimal precision is achieved where the factors used as strata are those that correlate most highly with the survey variables. From 2000, the stratification factors were based on an area classifier and selected indicators from the 1991 census. Details of how these were selected were reported in the January 2000 edition of the ONS Survey Methodology bulletin³.

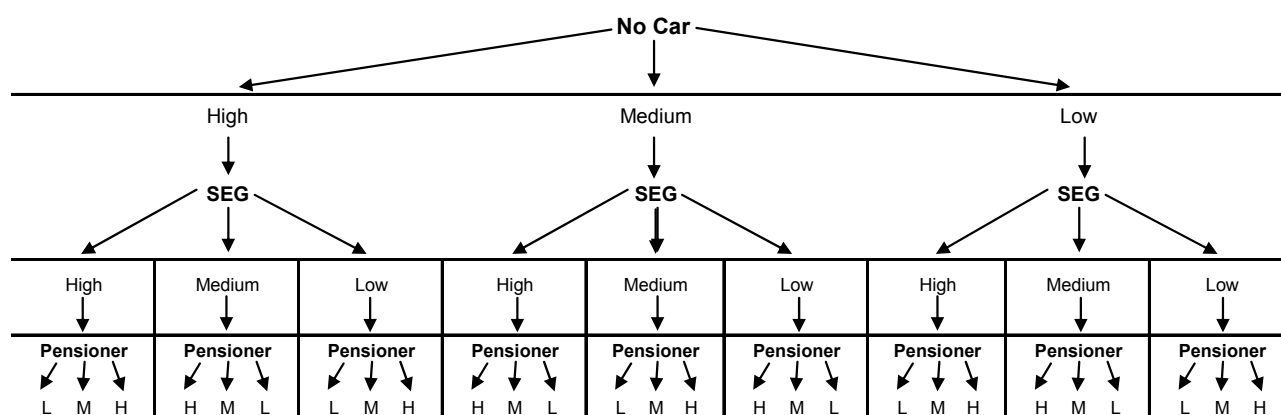
Initially, postcode sectors were allocated to 30 major strata. These were based on the 10 Government Office Regions in England, 5 subdivisions in Scotland and 2 in Wales. The English regions were divided between the former Metropolitan and non-Metropolitan counties. In addition London was subdivided into quadrants (Northwest, Northeast, Southwest and Southeast) with each quadrant being divided into inner and outer areas⁴. Using a finer division of London in the regional stratifier had a large effect on the increase in precision.

Within each major stratum, postcode sectors were then stratified according to the selected indicators taken from the 1991 Census. Sectors were initially ranked according to the proportion of households with no car, then divided into three bands containing approximately the same number of households. Within each band, sectors were re-ranked according to the proportion of households with household reference person in socio-economic groups 1 to 5 and 13, and these bands were then sub-divided into three further bands of approximately equal size. Finally, within each of these bands, sectors were re-ranked according to the proportion of people who were pensioners. In order to minimise the difference between one band and the next, the ranking by the pensioners and socio-economic group criteria were in the reverse order in consecutive bands, as shown in Figure B.A.

Figure B.A

Major strata were then divided into minor strata with equal numbers of addresses, the number of minor strata per major strata being proportionate to the size of the major stratum. Since 1984, the frame has been divided into 576 minor strata and one PSU has been selected from each per year. Of the 576 PSUs selected, 48 are randomly allocated to each month of the year. Each PSU forms a quota of work for an interviewer. Within each PSU, 23 addresses are randomly selected.

Figure B.A



Sample design on the longitudinal General Household Survey

In 2005, the GHS adopted a new sample design in line with European requirements, changing from a cross-sectional to a longitudinal design. This will help monitor European social policy by comparing poverty indicators and changes over time across the European Community. The GHS was identified as the best vehicle for this work over other ONS social surveys because there were many overlaps in the topics covered. The GHS design has changed to a four-yearly rotation, with increased sample size, and additional core questions.

The new sample design will follow a four-year sample rotation in which households remain in the sample for four years (waves) and one quarter of the sample is replaced each year. Each quarter of the sample is known as a replication, and each replication is representative of the target population. Figure B.B illustrates how the design operates. Once the system is fully established (from year 4 onwards) the sample for any one year consists of four replications which have been in the survey for 1, 2, 3 or 4 years. Each year one of the four replications is dropped and a new one added, giving an overlap of 75 per cent between successive years. This has been implemented to avoid high attrition rates due to repeated interviewing. Because 2005 is the first year of this longitudinal design, this sample does not contain any follow-up interviews.

Figure B.B

Figure **B.B**

Sample replication	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	
1	1st						
2	1st	2nd					
3	1st	2nd	3rd				
4	1st	2nd	3rd	4th			
5		1st	2nd	3rd	4th		
6			1st	2nd	3rd	4th	
7				1st	2nd	3rd	etc.
8					1st	2nd	etc.
9						1st	etc.

	New sample
	Follow-up sample

Conversion of multi-occupancy addresses to households

Most addresses contain just one private household, a few - such as institutions and purely business addresses⁵ - contain no private households, while others contain more than one private household. For addresses containing more than one household, set procedures are laid down in order to give each household one and only one chance of selection.

As the Postcode Address File (PAF) does not give names of occupants of addresses, it is not possible to use the number of different surnames at an address as an indicator of the number of households living there. A rough guide to the number of households at an address is provided on the PAF by the multi-occupancy (MO) count. The MO count is a fairly accurate indicator in Scotland but is less accurate in England and Wales, so it is used only when sampling at addresses in Scotland.

All addresses in England and Wales, and those in Scotland with an MO count of two or less, are given only one chance of selection for the sample. At such addresses, interviewers interview all the households they find up to a maximum of three. If there are more than three households at the address, the interviewer selects the households for interview by listing all households at the address systematically then making a random choice by referring to a household selection table.

Addresses in Scotland with an MO count of three or more, where the probability that there is more than one household is fairly high, are given as many chances of selection as the value of the MO count. When the interviewer arrives at such an address, he or she checks the actual number of households and interviews a proportion of them according to instructions. The proportion is set originally by the MO count and adjusted according to the number of households actually found, with a maximum of three households being interviewed at any address. The interviewer selects the households for interview by listing

all households at the address systematically and making a random choice, as above, by means of a table.

No addresses are deleted from the sample to compensate for the extra interviews that may result from these multi-household addresses, but a maximum of four extra interviews per quota of addresses is allowed. Once four extra interviews have been carried out in an interviewer's quota, only the first household selected at each multi-occupancy address is included. As a result of the limits on additional interviews, households in concealed multi-occupied addresses may be slightly under-represented in the GHS sample.

Data collection

Information for the GHS is collected week by week throughout the year⁶ by personal interview. In 2005, the field period changed from financial year to calendar year. As the 2004/5 survey finished in March 2005, interviewing for the 2005 survey started in April 2005 and ended in December 2005. The 2005 survey data contains interviews from January 2005 to December 2005. Since 1994 the survey has been carried out using Computer Assisted Personal Interviewing (CAPI) on laptop computers and Blaise software by face-to-face interviewers. Since 2000, telephone interviewers have converted GHS proxy interviews to full interviews using Computer Assisted Telephone Interviewing (CATI) from a central unit. Interviews are sought with all adult members (aged 16 or over) of the sample of private households and some information about children in the household is also collected.

A letter is sent in advance of an interviewer calling at an address. The letter briefly describes the purpose and nature of the survey and prepares the recipient for a visit by an interviewer. Since 2001, postage stamps have been included in the advance letter (see 'Improving response').

Data quality

The face-to-face and telephone interviewers who work on the GHS are recruited only after careful selection procedures after which they take part in an initial training course. Before working on the GHS they attend a briefing and new recruits are always supervised either by being accompanied in the field by a Field Manager or monitored by a Telephone Interviewing Unit (TIU) supervisor. All interviewers who continue to work on the GHS are observed regularly in their work.

Proxy interviews and the proxy conversion exercise

On occasion it may prove impossible, despite repeated calls, to contact a particular member of a household in person and, in strictly controlled circumstances, interviewers are permitted to conduct a proxy interview with a close household member. In these cases

opinion-type questions and questions on smoking and drinking behaviour, qualifications, health, family information and income are omitted.

During the review of the GHS⁷ the conversion of proxy interviews to full interviews was examined in order to improve the quality of data (a full interview is one in which the respondent has answered all sections of the questionnaire in person, either face-to-face or by telephone). This was achieved by re-contacting the household member, who was unavailable during the initial face-to-face interview, to answer the questions that were not asked of the proxy respondent on his/her behalf. The most efficient way of re-contacting these respondents was by employing Telephone Interviewing Unit (TIU) interviewers who could contact a widely dispersed population more efficiently than would be possible by conducting face-to-face interviews. Table B.1 shows the percentages of the types of interview taken for all persons in the co-operating households since 2000 before and after the proxy conversion exercise. In 2005 the TIU increased the percentage of full interviews conducted on the GHS from 70 per cent before proxy conversion to 72 per cent. The process of proxy conversion allows the GHS to provide more information on the topics not asked in detail of proxy respondents.

Table B.1

Table B1 Type of interview taken by proxy conversion status

All persons		Great Britain									
		Proxy conversion status									
		2001		2002		2003		2004		2005	
		Before	After	Before	After	Before	After	Before	After	Before	After
		%	%	%	%	%	%	%	%	%	%
Full interview		71	73	71	74	70	71	71	73	70	72
Proxy		7	5	7	5	7	6	7	5	8	6
Child/Other		22	22	22	22	23	23	21	21	22	22
Unweighted sample		21180	21180	20149	20149	24489	24489	20421	20421	30069	30069

2005 data includes last quarter of 2004/5 data due to survey change from financial year to calendar year.

Response

The GHS is conducted with people who volunteer their time to answer questions about themselves. The voluntary nature of the survey means that people who do not wish to take part in the survey can refuse to do so. Reasons for not participating in the survey vary from a dislike of surveys to poor health that prevents them from taking part. The sample is designed to ensure that the results of the survey represent the population of Great Britain. The risk of the survey not being representative is likely to increase with every refusal or non-contact with a sampled household (survey non-response). One measure of the quality of survey results is therefore the response rate.

Harmonised outcome codes and survey response rates

Harmonised outcome codes and definitions of response rates⁸ were introduced for the first time on the 2002/3 GHS and other large household surveys, following recommendations from National Statistics and the National Centre for Social Research joint working group on standard survey outcomes. The harmonised outcome codes are categorised as complete and partial interviews, non-contact, refusals and other non-responders, and unknown and known ineligibility.

The joint working group also recommended that surveys include an estimate of the proportion of cases of unknown eligibility that are eligible from 2002 onwards. It is assumed that the proportion of eligible cases amongst those cases where eligibility is unknown is the same as that amongst cases where eligibility has been established.

Four response rates are now calculated for the GHS on the basis of these outcome codes (full interviews also include cases where the complete interview has been partly completed by the selected person, and partly by proxy):

Overall response rate

This indicates how many full and partial interviews were achieved as a proportion of those eligible for the survey. In order to obtain the most conservative response rate measures, the denominator includes an estimate of the proportion of cases of unknown eligibility that would in fact be eligible for interview. In 2005 the overall response rate for the GHS was 74 per cent.

Full response rate

This is similar to the overall response rate calculated above, but only full interviews are included in the numerator. The full response rate for the GHS in 2005 was 64 per cent.

Co-operation rate

This indicates the number of achieved interviews as a proportion of those contacted during the fieldwork period. The co-operation rate for the GHS in 2005 was 75 per cent.

Contact rate

The contact rate measures the proportion of cases in which some household members were reached by the interviewer even though they might then have refused or been unable to give further information about the household or to participate in the survey. In 2005 the contact rate for the GHS was 96 per cent.

Table B.2 shows the outcome of visits to the addresses selected for the 2005 sample and the resultant number of households interviewed. Out of the 18,695 addresses that were selected, 17,184 were eligible and this yielded a sample of 17,310 eligible households. In 11,980 households, interviews (including proxy interviews) were achieved with every member of the household. In a further 291 households interviews were achieved with some but not all members of the household. This produced a total of 12,271 full or partial interviews.

Sample Design and Response

In total, 25 per cent of households selected for interview in 2005 were lost to the sample altogether, because they did not wish to take part (21 per cent) or because they could not be contacted (4 per cent) (table not shown). Table B.3 shows annual response by interview outcome category.

Tables B.2-B.3**Table B2** The sample of addresses and households*Great Britain: 2005*

Selected addresses		18695
Ineligible addresses:		
Demolished or derelict		
Used wholly for business purposes		
Empty		
Institutions		
Other ineligible		1511
No sample selected at address		
Address not traced		
Eligible addresses		17184
Number of households at eligible addresses		17310
Number of households where all individual interviews achieved (including proxies)	11980	
Number of households where some but not all individual interviews achieved	291	12271

2005 data includes last quarter of 2004/5 data due to survey change from financial year to calendar year.

Table B3 Annual Response

<i>Households</i>		<i>Great Britain: 2005</i>	
Outcome category		No.	%
1	Complete household co-operation	10650	63.5
2	Non-interview of one of more household members, proxy taken	1330	7.9
3	Non-contact with one or more household members	162	1.0
4	Refusal by at least one household member	129	0.8
5	Non-contact with household/resident	734	4.4
6	HQ refusal	627	3.7
7	Other refusal	3139	18.7
8	Other non-response	7	0.0
<i>Unweighted sample base = 100%</i>		<i>16,778</i>	

2005 data includes last quarter of 2004/5 data due to survey change from financial year to calendar year.

Trends in response

In order to continue to measure the response to the survey over time, a 'middle' response rate has been calculated since 1971. The middle response rate includes full interviews and accepts some of the partial household interviews as response – that is, it includes households where information has been collected by proxy and is therefore missing certain sections (category 2 in Table B.3), but does not include those where information is missing altogether for one or more household members (categories 3 and 4 in Table B.3). In other words, this middle rate can be thought of as the proportion of the sample of households known to be eligible from whom all or nearly all the information was obtained.

For the purposes of comparison, the middle response rate has been calculated, although it is not in itself a classification of the harmonised response rates. In 2005, full interviews also included complete interviews completed partly by the selected person and partly by proxy.

In 2005 the middle response rate was 72 per cent. Table B.4 shows middle response rates by Government Office Region. Trends in the middle response rate are shown in Table B.5. Since 1971, the middle response rate has shown some fluctuation. The decline in response rate since the early 1990s is due to an increase in the proportion of households refusing to participate (12 per cent in 1991 rising to 21 per cent in 2005) rather than failure to contact people. This decline reflects a general trend in decreasing response experienced by all survey organisations.

Tables B.4-B.5

Figure B.4 Middle response rates by Government Office Region

<i>Households</i>	<i>Great Britain: 2005</i>	
Government Office Region	%	Rank
North East	72.4	8
North West	71.6	9
Yorkshire and the Humber	74.1	5
East Midlands	75.7	1
West Midlands	71.3	10
East of England	72.5	7
London	63.1	11
South East (excluding Greater London)	74.2	4
South West	74.4	3
Wales	74.7	2
Scotland	73.2	6
Great Britain	72.1	

2005 data includes last quarter of 2004/5 data due to survey change from financial year to calendar year.

Full interviews include complete interviews partly completed by selected person and partly by proxy.

Table B5 Trends in the middle response rate: 1971 to 2005

<i>Households</i>	<i>Great Britain</i>
Year	Response rate
	%
1971	83
1972	81
1973	81
1974	83
1975	84
1976	84
1977	83
1978	82
1979	83
1980	82
1981	84
1982	84
1983	82
1984	81
1985	82
1986	84
1987	85
1988	85
1989	84
1990	81
1991	84
1992	83
1993	82
1994	80
1995	80
1996	76
1998	72
2000	67
2001	72
2002	69
2003	70
2004	69
2005	72

2005 data includes last quarter of 2004/5 data due to survey change from financial year to calendar year.

2005: Full interviews include complete interviews partly completed by selected person and partly by proxy.

Improving response

The GHS uses a number of methods to try to improve response. One ongoing method of improving response has been to reissue addresses to interviewers where there is a

Sample Design and Response

possibility of obtaining a better outcome, for example if there was initially a non-contact or a circumstantial refusal.

During the 2001/2 survey the advance letter was changed to mention that the survey had been running for 30 years. This change shows respondents how important the survey is and how many people take part. For subsequent years a similar advance letter has been produced informing respondents that the survey has been running for over 30 years.

Another method was introduced in 2002 following the findings of the Response Working Group of the Office for National Statistics. The Group found that response was higher among households who received postage stamps with their advance letters. Since August 2002, stamps have been included in all advance letters.

Sample sizes

Tables B.6 and B.7 show the numbers of households and individuals interviewed on the 2005 GHS by age, sex, region and country.

Tables B.6-B.7

Table B6 Unweighted bases: number of household reference persons in GHS 2005 by age, sex, region and country

Household reference persons								Great Britain: 2005
	Age							All
	16-24	25-34	35-44	45-54	55-64	65-74	75+	
Sex								
Male	199	1118	1676	1514	1535	1140	842	8024
Female	266	762	891	750	705	641	763	4778
Government Office Region								
North East	26	86	118	101	99	85	78	593
North West	56	223	306	248	277	219	204	1533
Yorkshire and the Humber	73	158	238	219	196	147	137	1168
East Midlands	38	160	188	166	180	129	128	989
West Midlands	37	170	216	193	202	180	135	1133
East of England	49	175	237	230	200	182	136	1209
London	41	265	305	263	203	148	127	1352
South East	47	257	373	317	346	262	234	1836
South West	34	158	232	199	196	161	175	1155
Country								
England	401	1652	2213	1936	1899	1513	1354	10968
Wales	15	77	119	125	129	95	103	663
Scotland	49	151	235	203	212	173	148	1171
Total	465	1880	2567	2264	2240	1781	1605	12802

Shaded figures also show the number of households in each region and country.

2005 data includes last quarter of 2004/5 data due to survey change from financial year to calendar year.

Sample Design and Response

Table **B7** Unweighted bases: number of people in GHS by age, sex, region and country

<i>All persons</i>		<i>Great Britain: 2005</i>								
	Age									All
	0-4	5-15	16-24	25-34	35-44	45-54	55-64	65-74	75+	
Sex										
Male	966	2247	1455	1809	2141	1845	1887	1293	937	14580
Female	935	2171	1535	1962	2320	1976	1940	1461	1189	15489
Government Office Region										
North East	79	190	142	156	195	171	161	127	99	1320
North West	214	543	361	430	522	430	457	323	264	3544
Yorkshire and the Humber	165	378	318	319	418	361	326	222	180	2687
East Midlands	155	337	270	317	329	294	324	208	178	2412
West Midlands	183	403	229	351	368	321	360	274	175	2664
East of England	184	417	303	346	439	388	351	291	185	2904
London	249	579	344	551	511	423	321	217	182	3377
South East	250	636	384	509	674	551	595	416	311	4326
South West	175	385	231	329	411	337	347	250	237	2702
Country										
England	1654	3868	2582	3308	3867	3276	3242	2328	1811	25936
Wales	87	200	152	156	192	205	236	157	128	1513
Scotland	160	350	256	307	402	340	349	269	187	2620
Total	1901	4418	2990	3771	4461	3821	3827	2754	2126	30069

2005 data includes last quarter of 2004/5 data due to survey change from financial year to calendar year.

Notes and references

¹ A limit is put on the number of households that are contacted per address. This is explained in detail at the 'Conversion of multi-occupancy addresses to households' section of Appendix B.

² From 1984 to 1998, the stratifiers used were a regional variable (based on the standard statistical region until 1996 and on the Government Office Region in 1998) and variables that measured the prevalence of privately rented accommodation, local authority accommodation and people in professional and managerial socio-economic groups.

³ Insalaco F Choosing stratifiers for the General Household Survey *ONS Social Survey Division, Survey Methodology Bulletin*, No. 46, January 2000.

⁴ The GOR regional stratifier:

1. North East Met
2. North East Non Met
3. North West Met
4. North West Non Met

Sample Design and Response

5. Merseyside
6. Yorks and Humberside Met
7. Yorks and Humberside Non Met
8. East Midlands
9. West Midlands Met
10. West Midlands Non Met
11. Eastern Outer Met
12. Eastern Other
13. Inner London North-East
14. Inner London North-West
15. Inner London South-East
16. Inner London South-West
17. Outer London North-East
18. Outer London North-West
19. Outer London South-East
20. Outer London South-West
21. South East Outer Met
22. South East Other
23. South West
24. Wales 1 – Glamorgan, Gwent
25. Wales 2 – Clwydd, Gwynedd, Dyfed, Powys
26. Highlands, Grampian, Tayside
27. Fife, Central, Lothian
28. Glasgow Met
29. Strathclyde (excl. Glasgow)
30. Borders, Dumfries, Galloway

⁵ Most institutions and business addresses are not listed on the small-user PAF. If an address was found in the field to be non-private (e.g. boarding house containing four or more boarders at the time the interviewer calls), the interviewer was instructed not to take an interview. However, a household member in hospital at the time of interview was included in the sample provided that he or she had not been away from home for more than six months and was expected to return. In this case a proxy interview was taken.

⁶ In 1988, the GHS interviewing year was changed from a calendar year to a financial year basis. From 2005, the GHS interviewing year was changed back from financial year to calendar year.

⁷ Walker A et al *Living in Britain Results from the 2000 General Household Survey: Appendix E*. TSO London 2002. Also available on the web: www.statistics.gov.uk/lib

⁸ Lynn P, Beerten R, Laiho J and Martin J. Recommended Standard Final Outcome Categories and Standard Definition of Response Rate for Social Surveys. ISER Working Papers. Number 2001 – 23. <http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2001-23.pdf>