national
STaTiSTiCS

General Household Survey 2005

Weighting and grossing

Appendix D

© Crown copyright

Office for National Statistics

Government Buildings

Cardiff Rd

Newport

NP10 8XG

Tel: 01633 655740

Email: ghs@ons.gsi.gov.uk

All surveys accept that there will be some degree of nonresponse, although great efforts are made to keep it to a minimum[1]. During the review of the General Household Survey (GHS) in 1999, two methods of compensating for nonresponse were examined with the aim of improving the quality of data. The method adopted to compensate for **unit nonresponse** (where all survey information for a sampled household is missing) will be described here. The method adopted to reduce **item nonresponse** (where information for particular questions is missing as the result of conducting proxy interviews) is discussed in *Appendix B Sample design and response*.

The 2005 GHS is weighted using a two-step approach. In the first step, the data are weighted to compensate for nonresponse (sample-based weighting). The second step weights the sample distribution so that it matches the population distribution in terms of region, age group and sex (population-based weighting)[2].

Weighting for nonresponse

Weighting for unit nonresponse involves giving each respondent a weight so that they represent the non-respondents who are similar to them in terms of survey characteristics. To be able to use this method, information about non-respondents is needed. However, by their very nature, non-responding households yield little information. Although some surveys collect information about the characteristics of these households, information about non-respondents is not routinely collected on the GHS. As such an alternative approach to gaining information about the GHS's non-responding households was needed to carry out such a weighting procedure.

Sample-based weighting using the Census

The decennial Census was found to be the most appropriate source of information about non-responding addresses on the GHS. Unlike the GHS, which relies upon voluntary co-operation from respondents, the Census is mandatory therefore nonresponse is kept to an absolute minimum.

After the 1991 Census, methodological work was conducted to match Census addresses with the sampled addresses of some of the large continuous surveys, of which the GHS was one. In this way it was possible to match the address details of the GHS respondents as well as the non-respondents with corresponding information gathered from the Census for the same address. It was then possible to identify any types of household that were being under-represented in the survey. Similar work is currently being conducted using the 2001 census.

The information collected during the 1991 Census/GHS matching work was used to weight the 2005 GHS data by identifying types of households that differed in terms of response rates. A combination of household variables such as household type, social class, region and car ownership were analysed using the software package Answer Tree (using the chi-squared statistic CHAID)[3] to identify which characteristics were most significant in distinguishing between responding and non-responding households. These characteristics were sorted by the program to produce the weighting classes shown in Figure D.A. The variables used to identify the weighting classes were restricted to those that appear annually on the GHS.

Figure DA

Figure DA **Weighting classes formed in the CHAID analysis**

Level 1 Split	Level 2 Split	Level 3 split	Level 4 split	Weight class
Region North East Merseyside Yorks & Humbs W Midlands South East Scotland	**No. of cars** 0 or 1	**No. of dependent children** 0 or 1	**Household type 1** 1 adult 16-59 Youngest 5-15 3+ adults no child	1
			Household type 1 2 adults 16-59 Youngest 0-4 2 adults, 1 or 2 60+ 1 adult only 60+	2
		No. of dependent children 2 or more		3
	No. of cars 2 or more			4
Region North West E Midlands Eastern South West	**Pensioner HH** Pensioner in HH	**SEG (grouped)** Skilled manual Partly-skilled manual Unskilled manual & others Not employed in last 10 years		5
		SEG (grouped) Professional Manager/employer Intermediate/junior		6
	Pensioner HH No pensioner in HH	**No. of adults** 1		7
		No. of adults 2 or more	**Social class** I, II, IV Not employed in last 10 years	8
			Social class IIInm, IIInm, V Other	9
Region London	**Type of building** Detached Semi-detached Terraced Converted flat/other			10
	Type of building Purpose built flat			11
Region Wales				12

Population-based weighting (grossing)

Population-based weighting schemes address deficiencies in the data due to sample-non coverage. They can also further reduce nonresponse bias and reduce the variance (sampling error) of survey estimates.

The population-based (grossing) method

The GHS sample is based on private households, which means that the population totals used in the weighting need to relate to people in private households. These totals are taken from the Labour Force Survey (LFS). The LFS derives household population estimates by excluding residents of institutions initially from population projections based on mid-year estimates.

The population information and GHS data were grouped into twelve age by sex categories within six region categories to form weighting classes as shown in Figure D.B. The population-based weighting consists of adjusting the existing weights (including factors for design and nonresponse) so that the final weights ensure that weighted totals for the above demographic categories match the population totals.

This procedure, also known as calibration, was carried out using the GES SAS macro. This was implemented in such a way as to ensure that all individuals within a household were given the same final weights[4].

Figure D.B

Figure DB **Weighting classes used for GES analysis**

Age/sex	Region
0-4	London
5-15	Scotland
16-24 male	Wales
16-24 female	Other Metropolitan
25-44 male	Other Non-metropolitan
25-44 female	South East
45-64 male	
45-64 female	
65-74 male	
65-74 female	
75+ male	
75+ female	

Presentation and interpretation of weighted data

Weighted data cannot be meaningfully compared to unweighted data from previous years without knowledge of how the weighting changes the estimates. In the GHS trend tables, weighted and unweighted data are presented for 1998 and weighted data are only shown from 2000 to 2005. Care should be taken when interpreting trend data or individual tables compared with other years as part of a time series.

It should be noted that the weighted bases used in this report are not recommended as a source for population estimates. They should primarily be used as bases for the percentages shown in tables rather than estimates of population size[5].

Weighting the 2005 Dataset

Since April 1994, the GHS has been conducted on a financial year basis, with fieldwork spread evenly across the year April-March. However, in 2005 the survey period reverted to a calendar year. Since the first quarter of 2005 was also the last quarter of the previous year's survey, the whole of the usual annual set sample of approximately 13,000 addresses, selected from the Postcode Address File, was dealt with in the nine months April to December 2005. Future surveys will run from January to December each year.

Since the 2005 survey does not cover the January-March quarter, this affects annual estimates for topics which are subject to seasonal variation. To rectify this, where the questions were the same in 2005 as in 2004-05, the final quarter of the 2004-05 survey has been added, weighted in the correct proportion, to the nine months of the 2005 survey.

It should also be noted that comparison of 2005 with 2004-05 is affected by having some individuals in both samples (those interviewed from January-March 2005). This reduces the sampling errors attached to estimates of change between the two surveys, so use of the usual recommended design effect of 1.2 gives a conservative estimate of errors – i.e. more, rather than less, likely to reject a difference as not statistically significant than is really the case.

Effects of weighting on data

Tables D.1 and D.2 identify the effects of weighting by comparing unweighted and weighted data for 2005. They also show the differences between the weighted and unweighted estimates for 1998 and 2000 to 2005, on a selection of household and individual level variables.

Tables D.1 and D.2

A comparison of the characteristics recorded on the 1991 Census forms of respondents and non-respondents in the 1991 GHS sample showed that households comprising one adult

aged 16 to 59 or a couple with non-dependent children were under-represented. Households containing dependent children were over-represented in the responding sample. As would be expected, weighting has changed the value of the estimate for some variables, but the overall changes have been relatively small.

For the 2005 estimates, the most marked effect of weighting was seen in the following variables. None of the effects are large.

Increase in value of estimate.

- 1 person households from 29% to 31%.

- 1 adult households from 34% to 37%.

- Households with no access to car or light van from 23% to 25%.

Decrease in value of estimate.

- 2 adult households from 52% to 48%.

- Households with access to two cars or vans from 27% to 22%.

- Households containing a married couple and no children from 25% to 23%.

The differences between the weighted and unweighted data for 1998 and 2000 to 2005 are also shown in Tables D.1 and D.2. It can be seen that the differences produced by weighting in 2005 were similar to those in previous years for the same variables.

Notes and references

1. Appendix B describes the variation in response for the GHS since it began in 1971.

2. Barton, J. Developing a Weighting and Grossing System for the General Household Survey: *Social Survey Methodology Bulletin* (Issue 49 July 2001).

3. CHAID is an acronym that stands for Chi-squared Automatic Interaction Detection. As is suggested by its name, CHAID uses chi-squared statistics to identify optimal splits or groupings of independent variables in terms of predicting the outcome of a dependent variable, in this case response.

4. GES, or *Generalized Estimation System*, is a SAS macro produced by Statistics Canada. The weights are formed using a form of calibration called Generalized Regression, or GREG estimation. The macro allows bounds to be set on the adjustment factors in the calibration.

5. Missing answers are excluded from the tables and in some cases this is reflected in the weighted bases, i.e. these numbers vary between tables. For this reason, the bases themselves are not recommended as a source for population estimates. Recommended data sources for population estimates for most socio-demographic groups are: ONS mid-year estimates, the Labour Force Survey, or Housing Statistics from the Department for Communities and Local Government.

Table D1 Weighted versus unweighted data for years 1998 to 2005 - household level

Household level variables

% of households		2005		Effect of weighting						
		Unweighted (a)	Weighted (b)	Weighted 1998 - Unweighted 1998	Weighted 2000 - Unweighted 2000	Weighted[1] 2001 - Unweighted 2001	Weighted 2002 - Unweighted 2002	Weighted 2003 - Unweighted 2003	Weighted 2004 - Unweighted 2004	Weighted 2005 - Unweighted 2005 (b-a)
Household size										
1 person		28.9	31.3	1.9	2.4	1.7	2.0	2.3	2.3	2.4
2 persons		36.3	35.0	-1.3	-1.6	-1.4	-1.7	-1.0	-1.4	-1.3
3 persons		14.8	14.8	0.2	0.2	0.2	0.1	0.0	0.0	0.0
4 persons		13.6	12.9	-0.4	-0.5	-0.3	-0.3	-0.8	-0.5	-0.7
5 persons		4.6	4.2	-0.1	-0.3	-0.1	-0.1	-0.4	-0.2	-0.4
6 or more persons		1.8	1.7	-0.2	-0.1	-0.1	0.0	-0.1	-0.1	-0.1
	Base	12802	24,829,368							
Number of adults										
1 adult		34.2	36.5	1.4	1.9	1.2	2.1	2.2	2.2	2.3
2 adults		51.6	48.4	-2.7	-2.6	-2.0	-2.5	-3.2	-3.0	-3.2
3 adults		9.9	10.3	0.5	0.3	0.4	0.1	0.4	0.3	0.4
4 or more adults		4.3	4.8	0.7	0.5	0.3	0.5	0.6	0.5	0.5
	Base	12802	24,829,368							
Number of children										
No children		72.0	73.8	1.5	1.5	1.1	0.6	2.1	1.3	1.8
1 child		12.4	12.4	0.3	0.0	0.1	0.3	0.1	0.1	0.0
2 children		11.1	9.9	-1.0	-0.8	-0.7	-0.6	-1.4	-0.9	-1.2
3 or more children		4.5	3.9	-0.8	-0.6	-0.4	-0.4	-0.8	-0.5	-0.6
	Base	12802	24,829,368							
Household type										
1 person only		29.0	31.4	1.9	2.4	1.8	2.0	2.3	2.3	2.4
2 or more unrelated adults		2.5	2.9	0.4	0.3	0.2	0.3	0.3	0.4	0.4
Married couple, dependent children		18.6	17.2	-0.8	-0.8	-0.6	-0.6	-1.4	-1.0	-1.4
Married couple, independent children		5.2	5.4	0.3	0.4	0.3	0.1	0.3	0.1	0.2
Married couple, no children		25.4	23.4	-2.0	-2.3	-1.7	-2.0	-1.7	-1.8	-2.0
Lone parent, dependent children		7.0	7.0	-0.4	-0.5	-0.5	0.0	-0.1	0.0	0.0
Lone parent, independent children		2.6	2.7	0.1	0.2	0.2	0.1	0.2	0.2	0.1
2 or more families (inc. same sex cohab)		1.1	1.0	-0.1	0.0	0.0	0.0	0.0	0.0	-0.1
Cohabiting couple, with children		3.9	3.6	0.0	0.0	0.0	0.1	-0.3	-0.1	-0.3
Cohabiting couple, no children		5.0	5.3	0.4	0.4	0.3	0.1	0.2	0.0	0.3
	Base	12802	24,829,368							
Tenure - harmonised										
Owns outright		31.9	30.3	-1.7	-1.9	-1.7	-1.6	-1.5	-1.1	-1.6
Buying on mortgage		38.8	38.4	0.3	0.1	0.5	0.0	-0.1	-0.7	-0.4
Rents from Council/Local Authority		11.5	12.3	0.1	0.5	0.2	0.7	0.6	0.7	0.8
Rents from HA/Reg. Social Landlord		7.3	7.4	0.1	0.1	0.1	0.2	0.1	0.3	0.1
Rents privately - unfurnished/nk		7.7	8.2	0.4	0.5	0.4	0.4	0.6	0.4	0.5
Rents privately - furnished		2.8	3.3	0.6	0.6	0.6	0.5	0.4	0.5	0.5
	Base	12802	24,829,368							
Ownership of consumer durables										
Washing machine		95.0	94.6	-0.8	-0.7	-0.6	-0.5	-0.6	-0.5	-0.4
Telephone		99.1	99.0	-0.3	-0.2	-0.1	0.0	-0.1	0.0	-0.1
Home computer		64.0	63.4	0.4	-0.1	0.4	-0.4	-0.3	-0.7	-0.6
	Base	12802	24,829,368							
Central heating		95.0	94.7	-0.3	-0.3	-0.2	-0.3	-0.1	-0.2	-0.3
	Base	12802	24,829,368							
Car or van ownership										
No car or van		23.4	25.1	0.6	1.1	0.6	1.4	1.4	1.7	1.7
One car or van		44.2	44.8	0.6	0.4	0.4	0.5	0.3	0.5	0.6
Two cars or vans		26.5	24.5	-1.1	-1.3	-1.0	-1.7	-1.7	-2.0	-2.0
Three or more cars or vans		5.8	5.6	-0.1	-0.2	0.0	-0.2	-0.1	-0.2	-0.2
	Base	12802	24,829,368							

1 Original 2001 weighting (based on LFS 2000 population estimates).

Table D2 Weighted versus unweighted data for years 1998 to 2005 - individual level

Table D2   Weighted versus unweighted data for years 1998 to 2005 - individual level

Individual level variables

% of individuals	2005 Unweighted (a)	Weighted (b)	Effect of weighting Weighted 1998 - Unweighted 1998	Weighted 2000 - Unweighted 2000	Weighted[1] 2001 - Unweighted 2001	Weighted 2002 - Unweighted 2002	Weighted 2003 - Unweighted 2003	Weighted 2004 - Unweighted 2004	Weighted 2005 - Unweighted 2005
									(b-a)
Limiting longstanding illness									
Male	32.7	32.2	-0.2	-0.3	-0.3	-0.4	-0.1	-0.1	-0.5
Female	33.0	33.3	-0.1	0.0	-0.1	0.1	0.3	0.3	0.3
Total	32.9	32.8	-0.1	-0.1	-0.2	-0.2	0.1	0.2	-0.1
Non-limiting longstanding illness									
Male	18.1	18.1	-0.4	-0.2	-0.2	-0.4	-0.2	-0.3	0.0
Female	19.6	20.0	0.0	-0.1	-0.1	-0.2	-0.1	0.0	0.4
Total	18.9	19.1	-0.2	-0.1	-0.2	-0.3	-0.2	-0.2	0.2
No longstanding illness									
Male	67.3	67.8	0.6	0.5	0.6	0.8	0.4	0.3	0.5
Female	67.0	66.7	0.1	-0.1	0.2	0.1	-0.1	-0.2	-0.3
Total	67.1	67.2	0.4	0.2	0.4	0.4	0.1	0.0	0.1
General health **Good**									
Male	65.7	65.7	0.6	0.1	0.0	0.1	0.1	0.1	0.0
Female	63.0	62.3	0.2	-0.3	-0.1	-0.4	-0.4	-0.4	-0.7
Total	64.3	63.9	0.5	-0.1	0.0	-0.1	-0.2	-0.2	-0.4
Fairly good									
Male	23.4	23.3	-0.2	0.0	0.0	0.0	-0.1	-0.2	-0.1
Female	24.6	25.0	0.0	0.2	0.0	0.2	0.2	0.1	0.4
Total	24.0	24.2	-0.2	0.0	0.1	0.1	0.1	0.0	0.2
Not good									
Male	10.9	11.0	-0.4	0.0	-0.1	0.0	0.0	0.1	0.1
Female	12.4	12.7	-0.2	0.1	0.1	0.1	0.2	0.2	0.3
Total	11.7	11.9	-0.3	0.0	0.0	0.0	0.1	0.2	0.2
Restricted activity in the last 14 days									
Male	11.3	11.4	0.0	-0.1	0.1	-0.1	-0.3	0.0	0.1
Female	14.3	14.6	0.0	0.1	0.0	0.1	-0.1	0.1	0.3
Total	12.8	13.1	-0.1	-0.1	0.1	0.0	-0.2	0.0	0.3
Cigarette smoking by sex **Men**									
Current cigarette smokers	24.0	25.3	1.4	1.2	1.2	1.0	1.2	0.9	1.3
Ex-regular cigarette smokers	28.8	27.3	-2.0	-2.0	-1.4	-1.5	-1.2	-1.1	-1.5
Never or (only occasionally) smoked	47.2	47.5	0.6	0.8	0.1	0.5	0.0	0.1	0.3
Women									
Current cigarette smokers	22.2	22.7	0.4	0.2	0.3	0.4	0.3	0.2	0.5
Ex-regular cigarette smokers	20.9	20.5	-0.3	-0.3	0.0	-0.2	-0.1	-0.4	-0.4
Never or (only occasionally) smoked	56.9	56.8	-0.1	0.1	-0.3	-0.2	-0.2	0.1	-0.1
Total									
Current cigarette smokers	23.0	23.9	0.8	0.8	0.8	0.7	0.8	0.6	0.9
Ex-regular cigarette smokers	24.6	23.7	-0.9	-1.0	-0.6	-0.7	-0.7	-0.6	-0.9
Never or (only occasionally) smoked	52.4	52.5	0.1	0.3	-0.2	0.1	-0.1	0.1	0.1
Maximum daily amount of alcohol drank last week by sex **Men**									
Drank nothing last week	27.4	27.8	0.0	0.3	0.0	0.4	0.1	0.6	0.4
Up to 4 units	39.1	37.6	-1.6	-1.2	-1.1	-1.2	-1.1	-1.3	-1.5
More than 4 units and up to 8	16.1	16.0	0.0	-0.1	0.1	-0.1	0.2	-0.1	-0.1
More than 8 units	17.4	18.6	1.6	1.1	1.1	1.0	0.8	0.8	1.2
Women									
Drank nothing last week	42.7	43.1	0.2	0.5	0.2	0.5	0.4	0.7	0.4
Up to 3 units	37.5	36.8	-0.7	-0.7	-0.5	-0.8	-0.7	-0.8	-0.7
More than 3 units and up to 6	12.3	12.3	0.1	0.0	0.0	-0.1	-0.1	-0.1	0.0
More than 6 units	7.5	7.8	0.4	0.2	0.3	0.3	0.4	0.1	0.3

1   Original 2001 weighting (based on the LFS 2000 population estimates).