



## General Household Survey 2007

---

# Weighting and grossing

## Appendix D

---

**Riaz Ali**

**Rebekah Binmore**

**Steven Dunstan**

**Julia Greer**

**David Matthews**

**Liam Murray**

**Simon Robinson**

© Crown copyright

Office for National Statistics

Government Buildings

Cardiff Rd

Newport

NP10 8XG

Tel: 01633 455877

Email: [ghs@ons.gsi.gov.uk](mailto: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<sup>1</sup>. The General Household Survey (GHS) compensates for **unit nonresponse** (where all survey information for a sampled household is missing) through weighting, which 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*.

Longitudinal surveys like the GHS experience two forms of nonresponse: non-participation to wave one and non-participation to later waves of the survey (also known as attrition). Here we will use the term nonresponse to refer to the former and attrition to refer to the latter.

The 2007 GHS is weighted using a two-step approach. The first step uses sample-based weighting to compensate for nonresponse and attrition. The second step uses population-based weighting to match the sample distribution to the population distribution in terms of age group, sex and region<sup>2</sup>.

## Weighting for nonresponse and attrition

Weighting for nonresponse and attrition can be seen as giving each respondent a weight so that they represent non-respondents who are similar to them in terms of survey characteristics. To be able to use this method consistent information about both respondents *and* non-respondents is needed to model the likelihood of response for different groups. In the case of nonresponse, little information is available on non-responding households directly from the survey<sup>3</sup>, so an external data source is required. For attrition, we are able to match back to previous years of the survey where longitudinal households have provided information.

The sample-based procedure for determining weighting characteristics is considered separately below for those sampled households in wave one and those in waves other than one.

## Sample-based weighting for nonresponse using the 2001 Census

Although we have no direct data on non-responding households to the GHS we use information from the Census to indirectly estimate the likelihood of response. Unlike the GHS, which relies upon voluntary co-operation from respondents, the Census is mandatory and therefore nonresponse is kept to an absolute minimum.

After the 2001 Census, methodological work was conducted to match Census addresses with the sampled addresses for some of the large household surveys, which included the GHS. Therefore, it was possible to match GHS respondents *and* non-respondents with corresponding information from the Census for the same address. We could then model and calculate response rates for types of household that were being under-represented in the survey. A combination of household variables available on an annual basis, such as household type, social class, region and car ownership were analysed using the software package AnswerTree (which uses chi-squared statistics to group households with similar response patterns)<sup>4</sup>. These chosen characteristics were used to produce the weighting classes shown in Figure D.A.

The weighting classes and weights determined from the work described above are then applied to weight the GHS data (for wave one respondents) on an annual basis.

Figure D.A

Figure D.A 2001 Weighting classes formed in the AnswerTree analysis

LEVEL 1 SPLIT	LEVEL 2 SPLIT	LEVEL 3 SPLIT	LEVEL 4 SPLIT	WEIGHT CLASS
<b>Region</b>	<b>Number of Rooms</b>	<b>Number of Pensioners in the household</b>	<b>Sex of the Household Reference Person</b>	
West Midlands  London	More than three	All pensioners	Female	1
			Male	2
		Two or two or more persons in the household but only 1 pensioner		3
		Two or two or more persons in the household where more than one person is a pensioner No pensioners in the household		4
	Three or fewer		5	
Yorkshire & Humberside East Midlands  South West Wales	<b>Household size</b>			
	More than two			6
	Two			7
	One	<b>Adults not employed</b>		
		One		8
		Zero	9	
North East  North West & Merseyside East of England South East Scotland	<b>Number of Adults</b>			
	More than two			10
	Two	<b>Accommodation Type</b>		
		Purpose built flat Part of a converted or shared house Other		11
		Detached Semi-detached Terraced		12
One			13	

## Sample-based weighting for attrition using the 2006 GHS

As mentioned earlier, attrition is a form of nonresponse found on longitudinal surveys between waves. The GHS is currently in year three of its longitudinal implementation and in 2007; this meant that approximately three-quarters of sampled households had been surveyed in 2006. As these sampled households had previously participated in the survey, details of respondents and non-respondents were linked back to their corresponding information at the previous wave. Logistic regression was used to model the likelihood of response in the current wave against the characteristics of households at their interview in the previous wave. A variety of household variables such as household type, socio-economic class, region and car ownership were tested for inclusion. Characteristics determined as significant by the logistic regression model (at the five per cent significance level) were used to weight for this attrition<sup>5</sup>. The variables reaching significance are listed in Figure D.B below.

**Figure D.B**

**Figure D.B Significant weighting variables formed in the logistic regression analysis**

Variable
Current wave
Government Office Region
Accommodation type
Household composition
Age of household reference person
Ethnicity of household reference person
Year of arrival into the United Kingdom of the household reference person
ILO employment status of household reference person
Employment contract of household reference person
Total number of hours worked by household reference person
Personal income benefits received by the household reference person
Total number of residents in the household who smoke
Total number of residents in the household who consume alcohol

## 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 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 population projections based on mid-year estimates and are adjusted to exclude residents of certain institutions.

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.C. 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<sup>6</sup>.

**Figure D.C**

**Figure D.C** 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 shown only from 2000 to 2007. 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 are the denominator for the percentages shown in tables and should not be regarded as estimates of population size<sup>7</sup>.

## Effects of weighting on data

Tables D.1 and D.2 identify the effects of weighting by comparing unweighted and weighted data for 2007. They also show the differences between the weighted and unweighted estimates for 1998 and 2000 to 2007, on a selection of household and individual level variables.

### **Tables D.1 and D.2**

A comparison of the characteristics recorded on the 2001 Census forms of respondents and non-respondents in the 2001 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 2007 estimates, the most marked effect of weighting was seen in the following variables. None of the effects are large.

Increase in value of estimate.

- Household type of 1 person only from 27.9% to 30.5%.
- 1 adult households from 32.3% to 34.7%.
- Households with no access to car or light van from 20.7% to 22.7%.

Decrease in value of estimate.

- 2 adult households from 52.4% to 47.6%.
- Households containing a married couple and no children from 26.8% to 22.4%.
- Households that are owned outright from 35.2% to 31.3%

The differences between the weighted and unweighted data for 1998 and 2000 to 2007 are also shown in Tables D.1 and D.2. It can be seen that the differences produced by weighting in 2007 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. Some surveys collect information about the characteristics of non-responding households although this is not routinely the case on the GHS.
4. AnswerTree uses the CHAID (Chi-squared Automatic Interaction Detection) algorithm, which 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.

5. The attrition weights build on the weights calculated for when these longitudinal cases were in their first wave. This means both nonresponse and attrition are covered within this weight.

6. 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.

7. 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.

**Table D.1** Weighted versus unweighted data for years 1998 to 2007 - household level

**Table D1** Weighted versus unweighted data for years 1998 to 2007 - household level

*Household level variables*

% of households	2007		Effect of weighting							
	Unweighted ( a )	Weighted ( b )	Weighted 2000 - Unweighted 2000	Weighted <sup>1</sup> 2001 - Unweighted 2001	Weighted 2002 - Unweighted 2002	Weighted 2003 - Unweighted 2003	Weighted 2004 - Unweighted 2004	Weighted 2005 - Unweighted 2005	Weighted 2006 - Unweighted 2006	Weighted 2007 - Unweighted 2007 (b-a)
<b>Household size</b>										
1 person	11.8	12.9	2.4	1.7	2.0	2.3	2.3	2.4	1.1	1.1
2 persons	31.2	28.4	-1.6	-1.4	-1.7	-1.0	-1.4	-1.3	-2.6	-2.8
3 persons	19.2	20.5	0.2	0.2	0.1	0.0	0.0	0.0	1.2	1.3
4 persons	23.5	23.6	-0.5	-0.3	-0.3	-0.8	-0.5	-0.7	0.3	0.1
5 persons	10.1	10.0	-0.3	-0.1	-0.1	-0.4	-0.2	-0.4	-0.2	-0.1
6 or more persons	4.2	4.6	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.2	0.4
<i>Base</i>	<i>21,472</i>	<i>58,080,776</i>								
<b>Number of adults</b>										
1 adult	32.3	34.7	1.9	1.2	2.1	2.2	2.2	2.3	2.2	2.4
2 adults	52.4	47.6	-2.6	-2.0	-2.5	-3.2	-3.0	-3.2	-4.9	-4.8
3 adults	10.9	12.0	0.3	0.4	0.1	0.4	0.3	0.4	1.5	1.1
4 or more adults	4.4	5.7	0.5	0.3	0.5	0.6	0.5	0.5	1.1	1.3
<i>Base</i>	<i>9,091</i>	<i>24,723,932</i>								
<b>Number of children</b>										
No children	72.5	73.3	1.5	1.1	0.6	2.1	1.3	1.8	1.0	0.8
1 child	12.4	13.2	0.0	0.1	0.3	0.1	0.1	0.0	0.8	0.8
2 children	11.1	10.1	-0.8	-0.7	-0.6	-1.4	-0.9	-1.2	-1.1	-1.0
3 or more children	4.1	3.5	-0.6	-0.4	-0.4	-0.8	-0.5	-0.6	-0.6	-0.6
<i>Base</i>	<i>9,091</i>	<i>24,723,932</i>								
<b>Household type</b>										
1 person only	27.9	30.5	2.4	1.8	2.0	2.3	2.3	2.4	2.4	2.6
2 or more unrelated adults	1.6	1.6	0.3	0.2	0.3	0.3	0.4	0.4	0.1	0.0
Married couple, dependent children	18.4	18.1	-0.8	-0.6	-0.6	-1.4	-1.0	-1.4	-0.5	-0.3
Married couple, independent children	5.7	6.7	0.4	0.3	0.1	0.3	0.1	0.2	0.9	1.0
Married couple, no children	26.8	22.4	-2.3	-1.7	-2.0	-1.7	-1.8	-2.0	-3.8	-4.4
Lone parent, dependent children	6.4	6.4	-0.5	-0.5	0.0	-0.1	0.0	0.0	0.0	0.0
Lone parent, independent children	3.0	3.3	0.2	0.2	0.1	0.2	0.2	0.1	0.4	0.3
2 or more families (inc. same sex cohab)	1.1	1.4	0.0	0.0	0.0	0.0	0.0	-0.1	0.2	0.3
Cohabiting couple, with children	4.4	4.5	0.0	0.0	0.1	-0.3	-0.1	-0.3	0.0	0.1
Cohabiting couple, no children	4.7	5.0	0.4	0.3	0.1	0.2	0.0	0.3	0.1	0.3
<i>Base</i>	<i>9,053</i>	<i>24,621,024</i>								
<b>Tenure - harmonised</b>										
Owns outright	35.2	31.3	-1.9	-1.7	-1.6	-1.5	-1.1	-1.6	-3.4	-3.9
Buying on mortgage	39.2	40.4	0.1	0.5	0.0	-0.1	-0.7	-0.4	0.4	1.2
Rents from Council/Local Authority	10.0	10.9	0.5	0.2	0.7	0.6	0.7	0.8	0.9	0.9
Rents from HA/Reg. Social Landlord	7.5	8.1	0.1	0.1	0.2	0.1	0.3	0.1	0.5	0.6
Rents privately - unfurnished/nk	6.3	7.0	0.5	0.4	0.4	0.6	0.4	0.5	0.6	0.7
Rents privately - furnished	1.8	2.2	0.6	0.6	0.5	0.4	0.5	0.5	0.7	0.4
<i>Base</i>	<i>9,088</i>	<i>24,718,306</i>								
<b>Ownership of consumer durables</b>										
Washing machine	96.7	96.3	-0.7	-0.6	-0.5	-0.6	-0.5	-0.4	-0.5	-0.4
Telephone	99.5	99.5	-0.2	-0.1	0.0	-0.1	0.0	-0.1	0.0	0.0
Home computer	70.6	71.1	-0.1	0.4	-0.4	-0.3	-0.7	-0.6	0.2	0.5
<i>Base</i>	<i>9,091</i>	<i>24,723,932</i>								
<b>Central heating</b>										
<i>Base</i>	<i>96.1</i>	<i>95.8</i>	<i>-0.3</i>	<i>-0.2</i>	<i>-0.3</i>	<i>-0.1</i>	<i>-0.2</i>	<i>-0.3</i>	<i>-0.3</i>	<i>-0.3</i>
<i>Base</i>	<i>9,091</i>	<i>24,723,932</i>								
<b>Car or van ownership</b>										
No car or van	20.7	22.7	1.1	0.6	1.4	1.4	1.7	1.7	2.0	2.0
One car or van	44.0	43.6	0.4	0.4	0.5	0.3	0.5	0.6	-0.2	-0.4
Two cars or vans	29.2	27.2	-1.3	-1.0	-1.7	-1.7	-2.0	-2.0	-2.1	-2.0
Three or more cars or vans	6.1	6.5	-0.2	0.0	-0.2	-0.1	-0.2	-0.2	0.3	0.4
<i>Base</i>	<i>9,091</i>	<i>24,723,932</i>								

<sup>1</sup> Original 2001 weighting (based on LFS 2000 population estimates).

**Table D.2** Weighted versus unweighted data for years 1998 to 2007 - individual level

**Table D2** Weighted versus unweighted data for years 1998 to 2007 - individual level

*Individual level variables*

% of individuals	2007		Effect of weighting							
	Unweighted (a)	Weighted (b)	Weighted 2000 - Unweighted 2000	Weighted <sup>1</sup> 2001 - Unweighted 2001	Weighted 2002 - Unweighted 2002	Weighted 2003 - Unweighted 2003	Weighted 2004 - Unweighted 2004	Weighted 2005 - Unweighted 2005	Weighted 2006 - Unweighted 2006	Weighted 2007 - Unweighted 2007
<b>(b-a)</b>										
<b>Limiting longstanding illness</b>										
Male	18.2	17.5	-0.3	-0.3	-0.4	-0.1	-0.1	-0.5	-0.6	-0.7
Female	19.3	19.2	0.0	-0.1	0.1	0.3	0.3	0.3	-0.2	-0.1
Total	18.7	18.4	-0.1	-0.2	-0.2	0.1	0.2	-0.1	-0.3	-0.3
<b>Non-limiting longstanding illness</b>										
Male	14.4	13.5	-0.2	-0.2	-0.4	-0.2	-0.3	0	-0.7	-0.9
Female	13.2	12.7	-0.1	-0.1	-0.2	-0.1	0.0	0.4	-0.4	-0.5
Total	13.8	13.1	-0.1	-0.2	-0.3	-0.2	-0.2	0.2	-0.5	-0.7
<b>No longstanding illness</b>										
Male	67.5	69.0	0.5	0.6	0.8	0.4	0.3	0.5	1.3	1.5
Female	67.5	68.1	-0.1	0.2	0.1	-0.1	-0.2	-0.3	0.5	0.6
Total	67.5	68.5	0.2	0.4	0.4	0.1	0.0	0.1	0.9	1.0
<b>General health</b>										
<b>Good</b>										
Male	68.5	69.6	0.1	0.0	0.1	0.1	0.1	0	0.6	1.1
Female	66.3	66.5	-0.3	-0.1	-0.4	-0.4	-0.4	-0.7	-0.2	0.2
Total	67.3	68.0	-0.1	0.0	-0.1	-0.2	-0.2	-0.4	0.1	0.7
<b>Fairly good</b>										
Male	21.8	21.2	0.0	0.0	0.0	-0.1	-0.2	-0.1	-0.2	-0.6
Female	22.8	22.5	0.2	0.0	0.2	0.2	0.1	0.4	0.2	-0.3
Total	22.3	21.9	0.0	0.1	0.1	0.1	0.0	0.2	-0.1	-0.4
<b>Not good</b>										
Male	9.6	9.3	0.0	-0.1	0.0	0.0	0.1	0.1	-0.4	-0.3
Female	11.0	11.0	0.1	0.1	0.1	0.2	0.2	0.3	0.0	0.0
Total	10.3	10.2	0.0	0.0	0.0	0.1	0.2	0.2	-0.2	-0.1
<b>Restricted activity in the last 14 days</b>										
Male	10.9	10.9	-0.1	0.1	-0.1	-0.3	0.0	0.1	-0.2	0.0
Female	12.7	12.8	0.1	0.0	0.1	-0.1	0.1	0.3	0.1	0.1
Total	11.9	11.9	-0.1	0.1	0.0	-0.2	0.0	0.3	0.0	0.0
<b>Cigarette smoking by sex</b>										
<b>Men</b>										
Current cigarette smokers	20.3	22.1	1.2	1.2	1.0	1.2	0.9	1.3	1.6	1.8
Ex-regular cigarette smokers	31.0	28.0	-2.0	-1.4	-1.5	-1.2	-1.1	-1.5	-2.5	-3.0
Never or (only occasionally) smoked	48.7	49.9	0.8	0.1	0.5	0.0	0.1	0.3	1.0	1.2
<b>Women</b>										
Current cigarette smokers	19.1	19.7	0.2	0.3	0.4	0.3	0.2	0.5	0.7	0.6
Ex-regular cigarette smokers	22.3	21.2	-0.3	0.0	-0.2	-0.1	-0.4	-0.4	-0.8	-1.1
Never or (only occasionally) smoked	58.6	59.1	0.1	-0.3	-0.2	-0.2	0.1	-0.1	0.1	0.5
<b>Total</b>										
Current cigarette smokers	19.6	20.9	0.8	0.8	0.7	0.8	0.6	0.9	1.1	1.3
Ex-regular cigarette smokers	26.4	24.4	-1.0	-0.6	-0.7	-0.7	-0.6	-0.9	-1.5	-2.0
Never or (only occasionally) smoked	54.0	54.8	0.3	-0.2	0.1	-0.1	0.1	0.1	0.4	0.8
<b>Maximum daily amount of alcohol drank last week by sex</b>										
<b>Men</b>										
Drank nothing last week	26.4	28.0	0.3	0.0	0.4	0.1	0.6	0.4	1.5	1.6
Up to 4 units	32.8	30.8	-1.2	-1.1	-1.2	-1.1	-1.3	-1.5	4.7	-2.0
More than 4 units and up to 8	17.5	16.8	-0.1	0.1	-0.1	0.2	-0.1	-0.1	-2.6	-0.7
More than 8 units	23.3	24.4	1.1	1.1	1.0	0.8	0.8	1.2	-3.7	1.1
<b>Women</b>										
Drank nothing last week	41.0	43.2	0.5	0.2	0.5	0.4	0.7	0.4	1.5	2.2
Up to 3 units	24.4	22.9	-0.7	-0.5	-0.8	-0.7	-0.8	-0.7	12.0	-1.5
More than 3 units and up to 6	19.6	18.6	0.0	0.0	-0.1	-0.1	-0.1	0	-7.0	-1.0
More than 6 units	15.0	15.3	0.2	0.3	0.3	0.4	0.1	0.3	-6.5	0.3

<sup>1</sup> Original 2001 weighting (based on the LFS 2000 population estimates).