

# Methodology

## Population

The Family Resources Survey sample aims to cover private households in Great Britain. Like some other household surveys, the Scottish Islands and the area to the North of the Caledonian Canal are excluded due to disproportionate fieldwork costs in this area. This coverage excludes approximately three per cent of the delivery points (letterboxes) in Scotland, which equates to 0.25 per cent of delivery points in Great Britain.

## Framework for sample selection

The FRS uses a stratified clustered probability sample drawn from the Royal Mail's small users Postcode Address File (PAF). The PAF is a list of all addresses where less than 50 items of mail are received a day, and is updated twice a year.

The survey selects 1,680 postcode sectors with a probability of selection that is proportional to size. Each sector is known as a Primary Sampling Unit (PSU).

The PSUs are stratified by 24 regions and also by three other variables derived from the 1991 Census of Population. Stratifying ensures that proportions of the sample falling into each group reflect those of the population.

Within each region the postcode sectors are ranked and grouped into six equal bands using the proportion of heads of household in socio-economic groups one to five and 13. Within each of these six bands, the PSUs are ranked by the total unemployment rate and formed into three further bands, resulting in 18 bands for each region. These are then ranked according to the proportion of households that are owner occupied. This set of stratifiers is chosen to have a maximum effectiveness on the accuracy of two key variables: household income and housing costs. The table which follows summarises the stratifiers.

Regions	17	in England (Metropolitan/non-Metropolitan/4 in London)
	2	in Wales
	5	in Scotland
Socio-economic groups	1	Employers/Managers in large establishments
	2	Employers/Managers in small establishments
	3	Professional workers (self employed)
	4	Professional workers (employees)
	5	Non-manual ancillary workers, foremen and supervisors
	13	Farmers (employers and managers)

---

Total unemployment rate

---

Owner occupancy

---

Within each PSU a sample of addresses is selected. In 1998-99, 23 addresses were selected per PSU.

Each year, one half of the PSUs are retained from the previous year's sample, but with new addresses chosen; while for the other half of the sample, a fresh selection of PSUs is made (which in turn will be retained for the following year). This is to improve comparability between years.

## Data collection methods

The Consortium of Social Survey Division (SSD) of the Office for National Statistics and the National Centre for Social Research (formerly SCPR) have been conducting fieldwork for the FRS since 1992. In September 1996 the FRS was tendered as part of the good practice in government programme. The consortium were again successful in winning the contract for a three year period (1997-98 - 1999-2000) with a one year extension to 2000-01.

Interviews are carried out jointly on behalf of the DSS by interviewers from the Office for National Statistics and the National Centre for Social Research. Each month the PSUs are systematically divided between the two organisations and then assigned to the field staff.

Before interviewers make contact with the selected addresses, a letter is sent to the address, explaining that it has been chosen for the survey, and that an interviewer will call. Participation in the FRS is voluntary. In October 1997 the FRS advance letter was revised following methodological work carried out by the Office for National Statistics and also a split sample test conducted jointly by the Office for National Statistics and SCPR on the FRS. The letter was simplified and its length reduced.

The interviewers are asked to call at the address. A lower limit of four calls is set and these calls have to be made at different times of the day and on different days of the week. In 1998-99, FRS interviewers averaged 7.7 calls per address before returning it as a non-contact.

The FRS was one of the first Government surveys to use Computer Assisted Personal Interviewing (CAPI). There are advantages to this over the traditional paper interviews, primarily:

- in-built checks for consistency can be made at the time of the interview,
- respondents are automatically routed only to those questions relevant to them,
- there is no need for a data input stage as the data are already available,
- questions with alternate wordings (eg is/was, his/her) can be automatically tailored to the situation,
- interviewers receive and transmit work via a modem in their own homes.

The average interview lasts around one hour and 20 minutes, but the time will vary according to the size of household and its circumstances.

The questionnaire itself is divided into three parts. First, the household schedule which is addressed to one person in the household (usually the head, although other members are encouraged to be present) and which mainly asks household level information, such as relationship of individuals to each other, tenure and housing costs. Second, the individual schedule which is addressed to each adult in turn and asks questions about employment, benefits, pensions, investments and other income. A final section goes on to ask the value of investments for relevant respondents.

Interviewers new to the FRS are briefed on the questionnaire and an annual re-briefing is given to all interviewers on changes to the questionnaire. Those who have been working on the survey for some time also complete a written field report each year, describing their experiences with particular parts of the questionnaire, and commenting on how changes are received in the field.

## Consultation of documentation

Interviewers are encouraged to consult documentation from respondents at all stages of the interview to ensure that the data are as accurate as possible. For some items, whether or not certain documents are consulted is recorded on the questionnaire, helping users of the data to judge the accuracy.

Of all employees interviewed in 1998-99, 60 per cent consulted a payslip in order to answer questions on the income from their employment. However, it should be noted that nine per cent of employees simply did not have a payslip to consult.

In recording data on benefit receipt, some form of documentation (an order book, a letter from the DSS or Benefits Agency or a bank statement) was consulted for 75 per cent of all benefits received.

The questionnaire records consultation of documentation for questions relating to Council Tax. 49 per cent of households consulted a Council Tax bill or statement in answering questions on their Council Tax payments.

In addition, from 1998-99 self-employed respondents are asked if they have documentation when they provide information about the profit or loss of their business. Of the 76 per cent of self employed respondents who had prepared business accounts, 32 per cent were able to refer to such documentation.

## Response

The FRS aims to interview all adults in a household. A household is defined as fully co-operating when it meets this requirement. In addition, to count as fully co-operating, there must be less than 13 'don't know' or 'refusal' answers to monetary amount questions in the benefit unit schedule (ie excluding the assets section of the questionnaire). Proxy interviews are accepted only under restricted circumstances. In 1998-99, for those households classed as fully co-operating, proxy responses were obtained for 14 per cent of adults.

**Table 1: Response in the 1998-99 Family Resources Survey**

	Number of households	Percentage of effective sample
Sampled addresses	39,127	
Ineligible addresses	4,491	
Effective sample (eligible households)	34,636	100
Fully co-operating households	22,913	66
Partially co-operating households	382	1
Refusals	10,014	29
Households with no contact	1,327	4

If a household is partially co-operating, the minimum requirement is that a full interview has been obtained from the head of household's benefit unit.

It should be noted that all data shown in the main body of this publication refer to fully co-operating households.

Table 1 summarises the household response. The original sample chosen for 1998-99 consisted of 39,127 addresses. However, 4,491 were then found to be ineligible because they were not defined as private households or were empty households. This left an effective sample of 34,636 households. Of these, 22,913 fully co-operated (66 per cent), 382 only partially co-operated (one per cent) and 10,014 refused to proceed with the interview (29 per cent). The interviewer was unable to make contact with 1,327 households (four per cent), the same proportion as in 1997-98.

The reasons for refusal and non-contact are recorded. The most common reason for refusal given was the feeling that answering questions from the FRS would be an 'invasion of privacy' (18 per cent); followed by 15 per cent who said they 'couldn't be bothered'. Concerns about confidentiality were only raised by four per cent of households. Seven per cent said they 'disliked a survey of income'.

The main reason given for non-contact was that there was rarely anybody at the address (28 per cent). A further 12 per cent of households could not be contacted because of the working shifts or odd hours of people in the household.

The achieved sample size was 22,913 households.

Table 2 shows response rates broken down by Government Office Region. Response rates are calculated as follows:

$$\frac{\text{Number of fully co-operating households} \times 100}{\text{Number of eligible households}}$$

The overall response rate for the FRS for the year 1998-99 was 66 per cent.

The region with the highest response rate was Wales, where 73 per cent of all households selected responded fully, followed closely by the North East (72 per cent). The region with the lowest response rate was London where only 62 per cent of the chosen households fully co-operated. The variation in response rates reflects those of other major surveys including the Census of Population. That is that response rates are generally lower in large city areas.

**Table 2: Regional response rates**

Government Office Region	Percentage of households
North East	72
North West and Merseyside	66
Yorkshire and the Humber	66
East Midlands	66
West Midlands	64
Eastern	65
London	62
South East	66
South West	69
England	66
Wales	73
Scotland	65
<b>Great Britain</b>	<b>66</b>

## Non-response

The lower the response rate to a survey, the greater the likelihood that those who responded are significantly unlike those who did not, and so the greater the risk of systematic bias in the survey results. Unless information is available about the nature and extent of such bias there are likely to be problems in generalising the sample results to the population. For a British survey of the size and complexity of the FRS the total non-response rate in 1998-99 of 34 per cent is not considered unreasonable. However, any information that can be obtained about the non-responders is useful both in terms of future attempts to improve the overall response rate and also potentially in improving the weighting of the sample results. It is considered a priority issue for the FRS to obtain as much information as possible about non-responders. The following sections outline some of the analysis that has been carried out in this direction.

### FRS non-response and ACORN classifications

A number of household characteristics are often associated with a higher or lower response rate on surveys<sup>1</sup>. Lower response rates than the average are often seen in:

- households in inner-city areas (especially London);
- single person households;
- head of household born outside the UK.

In contrast, higher response rates occur, for example, in households with dependent children. A greater understanding of how these varied characteristics interact and influence survey response can be useful for weighting of results and for maximising response rates in the future.

The DSS commissioned a study of non-response to the Family Resources Survey in relation to the ACORN<sup>2</sup> code of a household. ACORN is a geo-demographic classification system developed by CACI Marketing Systems. It is primarily aimed at identifying consumer markets by using data for Enumeration Districts from the 1991 Census. Clusters of Census data items such as age, sex, marital status, occupation, education, home ownership, car ownership, family structure and ethnic group are used to derive 54 ACORN 'types' which are amalgamated into 17 'groups' which in turn can be classified into six 'categories.' The various ACORN codes are accompanied by brief descriptions and pen portraits of the sort of households included in a particular code.

Tables 3 and 4 summarise the results of the analysis, showing the ACORN area 'types' with response rates significantly below or above the average for 1995-96 and 1996-97 combined. The codes are numbered from one to 54 with the lower codes indicating the more prosperous areas. Table 3 shows that for the two years combined, eight of the 11 areas with a response rate significantly below average have ACORN codes from one to 27. The FRS appears to be under-representing wealthier households. Similarly, Table 4 shows that nine of the 14 areas with a response rate significantly above average are in the less prosperous half of the ACORN codes (a code of 28 or above). This suggests that the FRS is over-representing poorer households. However, as the tables show, the detailed ACORN 'types' do not differ by more than 10 per cent in either direction from the 1995-97 average of 69.5 per cent.

---

<sup>1</sup> Kate Foster (1996) 'A comparison of Census characteristics of respondents and non-respondents to the 1991 Family Expenditure Survey', Survey Methodology Bulletin, 38, OPCS

---

<sup>2</sup> CACI Limited 1994. All rights reserved. Source: Office for National Statistics and GRO(s) Crown Copyright 1991. All rights reserved.

**Table 3: ACORN<sup>1</sup> area 'Types' with a statistically significant lower than average response rate**

ACORN Types	Percentage response rate							
	Full (%)		Partial (%)		Refusal (%)		Non-Contact (%)	
	1995-96	1995-97	1995-96	1995-97	1995-96	1995-97	1995-96	1995-97
24 Partially gentrified multi-ethnic areas	60	60	3	3	29	30	8	7
21 Prosperous enclaves, highly qualified executives	61	61	4	3	23	25	13	11
47 Council estates with high unemployment	61	63	3	2	27	27	9	8
25 Converted flats & bedsits, single people	63	63	2	2	26	27	9	8
23 Affluent city centre areas, tenements & flats	64	63	0	0	26	27	11	9
9 Private flats, elderly people	65	65	0	0	29	30	5	5
1 Wealthy Suburbs, large detached houses	65	66	2	1	31	30	3	2
53 Multi ethnic, severe unemployment, lone parents	65	66	2	2	22	24	11	8
36 Home owning multi ethnic areas, young families	66	65	4	2	27	29	4	3
5 Mature, well-off suburbs	67	67	2	2	29	29	3	2
26 More established home owning areas	69	67	1	1	28	30	2	2

**Table 4: ACORN<sup>1</sup> area 'Types' with a statistically significant higher than average response rate**

ACORN Types	Percentage response rate							
	Full (%)		Partial (%)		Refusal (%)		Non-Contact (%)	
	1995-96	1995-97	1995-96	1995-97	1995-96	1995-97	1995-96	1995-97
6 Agricultural villages, home-based workers	78	76	1	1	20	22	1	3
12 Transient workforces, living at their place of work	76	77	0	0	23	20	1	3
50 Council areas, high unemployment, lone parents	75	75	3	2	21	21	2	2
35 Low rise estates, older workers, new home owners	75	74	1	1	22	23	2	2
43 Council areas, young families, many lone parents	75	73	1	1	21	21	4	4
39 Home owners, small council flats, single pensioners	75	72	1	1	22	25	3	3
40 Council areas, older people, health problems	75	71	1	1	22	26	3	2
51 Council flats, greatest hardship, many lone parents	74	75	0	0	19	19	7	6
42 Council areas, young families, some new home owners	73	74	2	2	22	22	3	2
27 Rural areas, mixed occupations	73	73	2	1	23	23	2	2
41 Better off council areas, new home owners	73	72	2	1	22	24	3	3
2 Villages with wealthy commuters	73	71	1	1	24	24	2	3
4 Affluent suburbs, older families	73	70	1	1	24	27	2	2
46 Council areas, residents with health problems	71	71	71	1	24	24	3	3

<sup>1</sup> CACI Limited 1994. All rights reserved.

Source: ONS and GRO(s) Crown Copyright 1991. All rights reserved.

**Table 5: Response and non-response rates by selected ACORN<sup>1</sup> area 'Types': FES 1995-97 and FRS 1995-97**

ACORN Types	Percentage response rate					
	Full (%)		Refusal (%)		Non-Contact (%)	
	FES	FRS	FES	FRS	FES	FRS
6 Agricultural villages, home-based workers	75	76	25	22	0	2
13 Home owning family areas	70	69	29	29	1	2
27 Rural areas, mixed occupations	72	73	27	28	1	2
32 Home owning areas, skilled workers	68	71	30	26	2	4
34 Mature home owning areas, skilled workers	67	68	31	30	1	2
41 Better-off Council, new home owners	69	72	30	25	1	3
44 Multi-occupied terraces, multi-ethnic	71	70	24	25	5	5
<b>Areas with below-average response (FES)</b>						
1 Wealthy suburbs, large detached houses	59	66	39	31	1	2
5 Mature, well-off suburbs	59	67	38	31	3	2
18 Furnished flats and bedsits, young singles	48	66	36	26	16	9
19 Apartments, young professionals	53	67	42	28	5	5
20 Gentrified multi-ethnic areas	55	55	55	26	8	5
21 Prosperous areas, highly qualified executives	51	61	38	28	12	11
47 Council estates, high unemployment	55	63	38	29	7	8
49 Council flats, very high unemployment, singles	58	72	33	20	10	7
53 Multi-ethnic, severe unemployment, lone parents	47	66	45	26	7	8

<sup>1</sup> CACI Limited 1994. All rights reserved.

Source: ONS and GRO(s) Crown Copyright 1991. All rights reserved.

The relationship between non-response and ACORN classification was also carried out for the Family Expenditure Survey (FES) combining data from two survey years (1995-96 and 1996-97). FRS results and FES results were compared. Table 5 summarises the analysis, showing that there was greater variation in FES response rates (47 to 75 per cent compared with the FRS range of 55 to 76 per cent) and in non-contact rates (zero per cent to 16 per cent compared with two to 11 per cent). In both cases the lowest response rate came from the 'Partially gentrified multi-ethnic areas'. The ACORN Type with the highest rate of response to the FES and FRS was the 'Agricultural villages, home based workers' group.

For the FES, the areas of higher response were characterised as family areas, having high proportions of skilled workers or home ownership, or as rural areas. These areas had both low non-contact and refusal rates - an exception being 'Multi-occupied terraces, multi-ethnic areas' which had a relatively high non-contact rate (five per cent) but was offset by a very low refusal rate (24 per

cent). Areas of low response can be explained by either high refusal (for multi-ethnic areas and those with young professionals) or high non-contact (for areas with flats or single people) rates.

### FRS non-response and Council Tax band

Comparisons were made between 1998-99 FRS data and administrative data on the number of households within each Council Tax band. The analysis showed that FRS proportions were similar to those obtained from examining administrative data.

There were slight differences between countries (see Table 6). FRS data in England had a higher proportion of Band D households, but otherwise closely followed the pattern obtained from administrative data. Compared to administrative data, the FRS Scottish sample had a lower proportion of Band A households and a higher proportion of Band D households; Wales had a lower proportion of Band E households.

**Table 6: Proportion of households in each Council Tax band by country: FRS data and administrative data**

Council Tax band	England		Wales		Scotland	
	Administrative data	1998-99 FRS	Administrative data	1998-99 FRS	Administrative data	1998-99 FRS
Band A	26%	25%	20%	21%	26%	23%
Band B	19%	19%	25%	26%	25%	24%
Band C	21%	21%	20%	20%	15%	17%
Band D	14%	16%	15%	14%	11%	14%
Band E	9%	9%	12%	10%	11%	11%
Band F	5%	4%	4%	4%	5%	5%
Band G	3%	3%	3%	2%	4%	4%
Band H	1%	1%	0%	-	0%	1%
Not valued separately	2%	2%	2%	2%	2%	1%
<b>Total households (base=100%)</b>	<b>20,540,000</b>	<b>19,634</b>	<b>1,186,000</b>	<b>1,287</b>	<b>2,170,000</b>	<b>1,992</b>

### Non-response form analysis

Direct information about the non-responding households is valuable, although by definition difficult to obtain. However some households who are not willing to take part in the full survey may be willing to provide some basic information by completing a non-response form.

For the 1997-98 FRS interviewers carried out a short non-response form to collect information on households not taking part in the main survey. The non-response form tested asked about household relationships, age, working status, certain benefits, tenure and motor vehicles. Analysis of the data has been carried out and a report produced by the survey contractors. Highlights from the report are given below. It should be noted that only 33 per cent of households who declined to take part in the main survey agreed to provide some answers to the non-response form. Thus, results from the non-response form do not cover the majority of non-responders to the main FRS questionnaire and so should be treated with caution:

- 49 per cent of these households had no adult in paid work;
- 21 per cent of these households had children, compared to 31 per cent of households who took part in the main survey;
- nine per cent of these households has someone receiving Income Support, compared to 13 per cent of households who took part in the main survey;

- 66 per cent of these households had access to a motor vehicle, compared to 70 per cent of households who took part in the main survey.

Because of the low response to the non-response form, an experiment was carried out using four different response forms during the 1998-99 survey year. For all forms the response rate was high (73 per cent to 80 per cent). The response form varied in the number of questions asked and the topics covered. The results from the experiment were used to produce a new non-response form for the 1999-2000 survey year.

### Comparisons with other surveys

Some of the information collected by the FRS is also available in other Government surveys and comparisons of results can be a useful method of validation.

For example, FRS results have been compared with the Family Expenditure Survey (FES) in the context of Households Below Average Income (HBAI) analysis. Appendix 9 of the 1996/7 HBAI report details comparisons between the FRS and the FES. The main findings were that lower equivalised income was recorded in the FRS, particularly for singles and couples without children and also lower investment income, particularly for pensioners.

Analysis of 1997-98 FRS data suggests that estimates of the International Labour Organisation (ILO) definition of economic status compare favourably with estimates produced on an equivalent basis from the Labour Force Survey (LFS). Both sources showed that 59 per cent of adults were classified as employed, four per cent were classified as unemployed and 37 per cent as economically inactive. Figures are given to the nearest per cent.

Analysis of the LFS was carried out in such a way as to be as consistent with FRS data as possible. This was achieved by averaging LFS data over four quarters from March 1997 to February 1998 and only considering those individuals aged 16 or over (excluding those aged 16 to 18 who were in full time education) living in private households in Great Britain.

### **Validation, editing conversion and imputation**

In addition to unit non-response, where a household does not participate, a problem inherent in all large surveys is item non-response. This occurs when a household agrees to give an interview, but either does not know the answer to certain questions or refuses to answer them (see the section on response in this chapter for further information). They are still classified as fully co-operating households because there is enough known data to be of good use to the analyst.

The fact that the FRS allows missing values in the data can create problems for users. It was therefore decided before the first full year's FRS data was released, that missing values should be imputed where appropriate. The policy has been that for variables which are components of key derived variables, such as gross household income and housing costs, and areas key to the work of the Department, such as benefit receipt, there should be no missing information in the final data.

In addition to imputation, prior to publication, FRS data must be put through several stages of validation and editing. This is to ensure that the final data presented to the public are as accurate as possible.

The stages in the validation, editing, conversion and imputation process are laid as follows.

### **Stage one - the interview**

As noted previously, one of the benefits of interviewing using CAPI is that in-built checks can be made at the interview stage. If answers are inconsistent or outside a certain range, the interviewer will receive a warning message instructing them to check that what the respondent is saying is correct. This method helps to check both respondents' responses and that interviewers do not make keying errors. There are checks to ensure that amounts are within a valid range and also cross-checks which make sure that an answer does not contradict a previous response. However, it is not possible to check all potential inconsistencies as this would slow down the program to an unacceptable degree, and there are also capacity constraints on text messages. Interviewers can override most checks if the answers are found to be accurate when confirmed with respondents.

### **Stage two - post-interview checks**

Once an interview has taken place, data are returned to the Office for National Statistics or the National Centre for Social Research. Here a certain amount of editing takes place, mostly based on any notes made by interviewers. Notes are made by the interviewer when a warning has been overridden. These may be, for example, where an amount is outside the specified range, but the respondent has documentation to prove it is correct. Office-based staff make edit decisions based on these notes. Other edits that take place at this stage are checking amounts of fixed rate benefits such as Child Benefit and, where possible, separating multiple benefit payments into their constituent parts.

### **Stage three - data conversion**

Before it can be validated further, FRS data must be converted from its CAPI format into SAS readable tables. Using DSS specifications SAS tables are created by the Office for National Statistics, each table recording different information from the questionnaire. Both the DSS and the Office for National Statistics then carry out validation checks on key input and output variables to ensure that the data have converted correctly to the new format. Checks include ensuring that the number of adults and children recorded is correct, and that records are internally consistent.

## *Stage four - pre-imputation cleaning*

In preparation for imputing missing values, data are made as clean as possible. This involves edits and checks of the following nature:

### Weekly amounts

In the FRS, all amounts received or paid are converted to a weekly value. To calculate this, respondents are usually asked firstly the amount last paid or received and then the length of time this covered. This is known as a period code. As part of the conversion process outlined in Stage three, period codes are used in conjunction with amount variables to give weekly totals for all receipts and payments. Some variables, such as interest on savings accounts, refer to the amount paid in the last year. These are also converted to a weekly amount.

Sometimes the period code relates to a lump sum or a one-off payment. In these cases the corresponding value does not automatically convert to a weekly amount. In order for the data to be consistent across the survey, edits are applied to convert lump sums and one-off payments to weekly amounts. In the same way, where period codes were previously recorded as 'don't know' or 'refused', these are imputed so that the corresponding amount can be converted to a weekly value in the final database.

### Zero amounts

In previous years it was possible for interviewers to enter zero amounts when it is inappropriate to do so, for example in response to a question on receipt of benefit, when in fact the amount should be entered as missing. This created problems at later stages of analysis. From 1997-98, zero amounts can no longer be entered without a warning message to the interviewer. Some interviewers tried to avoid this message by recording near-zero amounts. These are also examined.

### Outliers

Statistical reports of the data are produced to show individual cases where an amount was greater than five standard deviations away from the mean. For these cases

the individual record is examined and where necessary (if a value looked unrealistic) the case is edited. The outliers remaining in the database are verified as being true values by examining other relevant data. Compared to earlier years, the number of these types of edits that had to be carried out on the 1998-99 FRS was small because of range checks that have been put into the CAPI questionnaire.

### Credibility checks

Checks are carried out for the internal consistency of certain variables. For example it is ensured that there are no benefit units containing only one adult where the respondent states that they are married and their partner is in the household. Such cases are examined and edited where necessary.

### Stage five - imputation

The responses to some questions are much more likely to have missing values than others. For example, it is very unlikely that a respondent will refuse to give, or will not know, their age or marital status, whereas it is much more likely that they will not be able to provide detailed information on the exact amounts of interest received from an investment.

The two areas where missing values are a major problem are income from self employment and income from investments.

Data in the tables provided in this publication include imputed values. However, for some variables missing values remain.

Table 7 illustrates the extent of the problem of missing values. It should be noted that out of over 11 million set values in the FRS database, only 0.4 per cent were originally recorded as either 'don't know' or 'refused'. Out of 44,518 missing values, approximately 92 per cent were imputed.

A combination of methods of imputation were used for the 1998-99 FRS data. The main ones are summarised on the next page in the order in which they were used.

**Table 7: Summary of imputation in 1997-98 and 1998-99 FRS**

	1997-98		1998-99	
	Values	Percentage of values	Values	Percentage of values
<b>Responses</b>				
Expected number of responses	11,671,803	100	11,495,935	100
Valid responses	11,620,954	99.6	11,451,417	99.6
Missing values (don't know/refused)	50,849	0.4	44,518	0.4
<b>Treatment of missing values</b>				
Hotdeck	25,977	51	27,782	62
Neural (set to mean)	10,640	21	5,129	12
Other imputation method	8,161	16	6,359	14
Benefit editing	1,288	3	1,540	3
Left as Missing	4,783	9	3,708	8

## Closing down routes

As with any questionnaire, a typical feature of the FRS is the gatekeeper question positioned at the top of a block of further questions, at which a particular response will open up the block. If the gatekeeper question itself is answered as 'don't know' or 'refused', the block is skipped. This results in a potential problem.

A missing gatekeeper variable could be imputed such that a further series of answers would be expected. However, these answers will not appear because a whole new route has been opened. For example, if the amount of rent is missing for a record and has since been imputed, any further questions about rent would not have been asked. From the post-imputed database, it will appear that these questions should have been asked because a value is there for rent.

This is why, where appropriate, the decision was taken that with manual imputations a route should be closed down. In most cases, gatekeeper variables are of the 'yes/no' type. These would be imputed to 'no', assuming that if a respondent does not know whether an item is received or paid, then it is not.

## Hotdecking

A hotdecking program is used to impute those variables that were to be used for segmenting on the neural network system and also those for which the neural network system was known to be unsuitable, following attempts in previous years. Hotdecking essentially looks at characteristics within a record containing the missing value to be imputed and matches it up to another record with similar characteristics

for which the variable is not missing. It then takes the known variable and copies it to the missing case. For example, for imputing the amount included in rent for services, classes of Council Tax band, number of bedrooms and Standard Statistical Region are used to search for a case with a similar record. This method ensures that imputed solutions are realistic, and gives a wide range of solutions maintaining variability in the data.

## Algorithms

Algorithms are used to impute missing values for certain variables, for example variables relating to education grants and to Council Tax. The algorithms range from very simple calculations to more sophisticated models based on observed relationships within the data and individual characteristics such as age and sex.

## Neural Networks

Following a pilot study by the organisation Recognition Systems Limited (RSL) prior to processing of 1993-94 FRS data, it was agreed that the DSS would use neural networks as a method of imputing missing values.

Neural networks are information processing systems that learn by example, recognising patterns in data. Their main advantage over standard statistical techniques is that they can extract and model non-linear relationships without assuming any particular underlying distribution.

The neural network system is not an appropriate technique to use for imputation of all the variables on the FRS which have missing values. Those for which it is most suitable have emerged through running the system over successive

**Table 8: Extent of imputation**

Variable	Actual number missing	Proportion of final values imputed	Method of Imputation
Amount of National insurance lump sum	303	65%	ALGORITHM
Full rent before HB/rent rebate	17	53%	HOTDECK
Amount of 3rd refund for HH expenditure	1	50%	ALGORITHM
Number of weeks had income from a Trust	3	50%	ALGORITHM
1st loan repayment incl interest/capital	13	48%	ALGORITHM
How much income tax deducted last time	87	40%	HOTDECK
Amount of profit before tax	116	39%	HOTDECK
Higher/lower rate of SMP	9	38%	ALGORITHM
Amount incl in rent for water/sewerage	706	37%	HOTDECK
Amount from absent partner paid directly	19	34%	NEURAL NETWORKS
2nd loan repayment incl interest/capital	2	33%	ALGORITHM
Amount included in last wage for SMP	24	30%	ALGORITHM
Amount: Insurance part of repayment	974	29%	HOTDECK
Amount received from absent partner	11	29%	NEURAL NETWORKS
Amount of net profit or loss	614	25%	ALGORITHM+HOTDECK

years. Where feasible other methods of imputation are completed before the neural system is run, so as complete a dataset as possible can be utilised. Each year it is necessary to update various control tables to ensure consistency with the latest survey variables. The tables instruct the neural system on how to segment cases when deriving the neural solutions (for example on age, sex, etc). Since the neural system was first used significant improvements have been made in the segmentation of variables used.

Once the neural network has been run there were still a number of cases where it was unable to find a neural solution for a variable and provided a mean by class as an alternative. This is usually because there was insufficient data to build the necessary models. All output from the RSL imputation system is checked and only used if a realistic value had been imputed and if there was a reasonable spread of data. After examination of the data, some of the solutions are found to be inappropriate and these are then edited. For example, where a large number of cases had been imputed to a single mean the output is not used.

### 'Mop-up' imputation

This is achieved by running a general validation report of all variables and looking at those cases where missing values were still present. At this stage, variables are looked at on a case-by-case basis to decide what to impute.

Credibility checks are then re-run to ensure that imputation

had not resulted in any inconsistencies in the data, and edits were applied where necessary.

All imputations, by each of the methods above, are applied to the unimputed data set via a transaction database. This ensures that it is always possible to reproduce the original data.

Table 8 lists the variables for 1998-99 where more than 25 per cent of final values were imputed.

### Points to note with imputed data

Although a great deal of time has been spent on imputing missing values, it should be remembered that they represent only a very small proportion of the dataset as a whole. However, the following points should be noted:

- as mentioned above, in certain situations, imputed values will be followed by 'skipped' values. It was decided that it was better to impute the top of a route only and not to impute large amounts of data. There are a small proportion of imputations for which it was not appropriate to close down a route. These cases are followed by 'skipped' responses (where a value might otherwise be expected).
- imputation will have a greater effect in distorting the distribution of original data for variables that have a higher proportion of non-response, as proportions of imputed data will be higher.

## *Stage six - Benefit validation*

Information on Social Security Benefits received is one of the key areas of the FRS and it is very important that this section is thoroughly validated and cleaned.

It is not appropriate to use the imputation methods outlined above for benefits data so instead a separate procedure of validation and editing is used. The following types of validation were carried out for 1998-99 FRS data:

### **Missings**

For cases where a respondent had answered 'yes' to whether they were in receipt of a particular benefit, but had not given the amount received, an imputation decision was made depending on the benefit. For benefits such as Income Support, where the rate would vary greatly depending on the situation of the respondent, individual benefit assessments were carried out. However, for benefits such as Retirement Pension, where fewer rates apply, a more general program could be written.

### **Near zero amounts**

Where benefit amounts were recorded as near zero, the case was examined individually and an edit decision was made.

### **Multiple benefits**

Any remaining combined benefit amounts (for example where Retirement Pension is paid with Income Support) not split at the office editing stage were edited by carrying out benefit entitlement assessments on individual cases, while preserving the reported total.

### **Attendance Allowance**

It has been noted in previous years that the FRS under-reports receipt of Attendance Allowance (AA). In the past receipt of Retirement Pension (RP) was investigated to assess whether the amounts might include AA. If the amount of RP received was above a certain threshold, AA cases were created for these respondents. From 1997-98, extra questions were asked of RP recipients on whether the amount of RP they stated that they received included AA or Disability Living Allowance (Care component) or Disability Living Allowance (Mobility component). Thus there was no longer any need to create AA records. An assessment was made on whether AA recipients were receiving higher rate or lower rate AA based on the amount they received for their RP.

## **Validation reports**

Computer programs were run to carry out a final check for benefit entitlement and to output any cases that looked unreasonable. All cases detected as a result of this validation exercise were individually checked and edited where necessary.

### **Quality of benefits data**

As part of the data validation process, comparisons are made between the FRS and other data sources. Table 9 shows a comparison of FRS benefit recipients compared to administrative data. The table shows both FRS sample data and grossed up sample estimates (see explanation of grossing factors in next section). Despite much time and effort being spent on benefit validation, there are still areas where there are known problems with the FRS data. The FRS under reports receipt for most of the benefits. The discrepancy between FRS and administrative data is particularly pronounced for Attendance Allowance, Severe Disability Allowance and Jobseeker's Allowance.

Users should note that some of the discrepancies in the two sources of data may be due to the fact that it is not always possible to compare like with like. For this year's report changes have been made to try and eliminate some of the differences between the two data sources. For example, the denominator for the administrative and the FRS data in Table 9 is the same and the administrative data figures for Retirement Pension and Widow's Benefit have been adjusted to remove those resident overseas. However, the fact that the FRS only interviews members of private households whereas administrative benefit systems (apart from Income Support) do not distinguish between people in private households and those in institutions remains a problem in comparing the two sources. For most benefits, only a very small minority of recipients will be in institutions, but this will have a greater effect on Attendance Allowance comparisons.

**Table 9: Individuals in receipt of Social Security benefits**

						Individuals	
1998-99 FRS			1998-99 FRS		DSS administrative data		
Grossed			Sample				
Benefit received	Number	Percentage	Number	Percentage	Number	Percentage	
Family Credit <sup>1</sup>	673,000	1.5	725	1.8	782,000	1.8	
Income Support <sup>2</sup>	3,144,000	7.2	3,139	7.7	3,546,000	8.1	
Housing Benefit <sup>1</sup>	4,321,000	9.9	4,359	10.7	4,414,000	10.1	
Council Tax Benefit <sup>3</sup>	5,503,000	12.6	5,558	13.7	5,258,000	12.0	
Jobseeker's Allowance <sup>1</sup>	1,008,000	2.3	833	2.1	1,175,000	2.7	
Retirement Pension <sup>4</sup>	9,497,000	21.7	9,615	23.7	10,061,000	23.0	
Widows Benefit <sup>5</sup>	233,000	0.5	253	0.6	251,000	0.6	
Incapacity Benefit <sup>6</sup>	1,715,000	3.9	1,548	3.8	1,557,000	3.6	
Severe Disablement Benefit <sup>6</sup>	246,000	0.6	221	0.5	370,000	0.8	
Attendance Allowance <sup>7</sup>	886,000	2.0	884	2.2	1,242,000	2.8	
Invalid Care Allowance <sup>8</sup>	382,000	0.9	354	0.9	373,000	0.9	
Disability Living Allowance (Care Component) <sup>9</sup>	1,077,000	2.5	990	2.4	1,213,000	2.8	
Disability Living Allowance (Mobility Component) <sup>9</sup>	1,299,000	3.0	1,214	3.0	1,645,000	3.8	
Child Benefit <sup>10</sup>	6,873,000	15.7	7,176	17.7	7,054,000	16.1	
<b>Total adults<sup>11</sup></b>	<b>43,788,000</b>	<b>100.0</b>	<b>40,586</b>	<b>100.0</b>	<b>43,788,000</b>	<b>100.0</b>	

<sup>1</sup> Administrative data average over 1998-99.

<sup>2</sup> Administrative data average over 1998-99, data adjusted to remove those in institutional accommodation for Income Support.

<sup>3</sup> CTB data available for the first person in the household only for FRS data, administrative data average over 1998-99.

<sup>4</sup> Administrative data at March 1999, figure excludes recipients resident overseas.

<sup>5</sup> Widows Pension and Widowed Mother's Allowance, figure excludes recipients resident overseas, administrative data as at March 1999.

<sup>6</sup> Administrative data as at 28 February 1999.

<sup>7</sup> Administrative data as at 31 August 1998.

<sup>8</sup> Administrative data as at 31 September 1998.

<sup>9</sup> Includes those receiving both care and mobility components, figure excludes recipients under 16, administrative data at 31 August 1998.

<sup>10</sup> Administrative data as at 31 May 1999.

<sup>11</sup> FRS figure used as a base for both comparisons and excludes those aged 16-18 in full time nonadvanced education.

## Estimation methodology

The 1998-99 FRS publication presents tabulations where the percentages refer to sample estimates grossed up to apply to the whole population.

Grossing up is the term usually given to the process of applying factors to sample data so that they yield estimates for the overall population. The simplest grossing system would be a single factor, the uniform grossing factor, which could be calculated as the number of households in the

population divided by the number in the achieved sample. However, surveys are normally grossed by a more complex set of grossing factors, which attempt to correct for differential non-response at the same time as they scale up sample estimates.

The system used to calculate grossing factors for the FRS divides the sample into different groups and the grossing factors are the ratio of population estimates to sample counts for those groups. The groups are designed to reflect differences in response rates among different types of

households. They have also been chosen with the aims of DSS analyses in mind. The population estimates are based on control variables, with values derived from external data sources.

The control variables and their sources are listed below. The FRS grossing system controls for variables at both household level and benefit unit level. A grossed count of the number of owner occupying households would thus tie in with the DETR figure, whilst the grossed number of single men under 35 would be consistent with the Office for National Statistics estimate. Some adjustments have been made to the original control total data sources so that definitions match those in the FRS, eg an adjustment has been made to the demographic data to exclude people not resident in private households.

In order to reconcile control variables at different levels and estimate their joint population, software provided by the French national statistics institute INSEE has been used. This software works by iterating towards a solution and options within it that give the solution which minimises the range of grossing factors have been used. This should maximise the potential precision of grossed estimates; if a few cases are associated with very small or very large grossing factors, grossed estimates will have relatively wide confidence intervals.

Careful consideration has been given to the combination of control totals and the way age ranges, Council Tax bands and so on, have been grouped together. The aim has been to strike a balance so that the grossing system will provide, where possible, accurate estimates in different dimensions without significantly increasing variances. Further details of how the FRS grossing system was developed are available in a DSS Analytical Note.

## Reliability of estimates

All survey estimates have a sampling error attached to them, calculated from the variability of the observations in the sample. From this, a margin of error (confidence interval) is derived. It is this confidence interval (rather than the estimate itself) which is used to make statements about the likely 'true' value in the population; specifically, to state the probability that the true value will be found between the upper and lower limits of the confidence interval. In general, a confidence interval of twice the standard error is used to state, with 95 per cent confidence, that the true value falls within that interval. A small margin of error will result in a narrow interval, and hence a more precise estimate of where the true value lies.

The calculation of sampling errors (and thus confidence intervals) is based on an assumption of a simple random sampling method, but in practice this is almost never used with large general population surveys, due to its inefficiencies with regard to cost and time. The sample for the FRS, as described earlier, is selected using a stratified multi-stage design, based on addresses clustered into postal sectors. The sampling error estimate is therefore not simply based on the variability among all units in the sample (whether households or individuals) but must also take into account the variability within and between postal sectors. If a sample characteristic is distributed differently by postal sector (ie is clustered) this produces a greater overall variance than would occur in a simple random sample of the same size. In other words, the complex (actual) sampling error is greater than the (assumed) simple random sampling error.

### Control variables used to generate grossing factors

Variable	Groupings	Source of data
Age/sex/marital status	Single men: <35, 35-59, 60+ Single women: <35, 35-64, 65+ Couples: <65, 65+	Office for National Statistics, GAD
Lone parents	Male, female	DSS estimates
Families	No. of couples with children	DSS estimates
Tenure type	LA renters, private renters, owner occupiers	DETR estimates
Council Tax Band	A, B, C-D, E-H	DETR estimates
Region	London, other	DETR estimates

The size of the actual standard error relative to the simple random sampling error is represented by the design factor (DEFT) which is calculated as the ratio of the two. Where the standard errors are the same, the DEFT is one, implying that there is no loss of precision associated with the use of a clustered sample design. In most cases, the DEFT will be greater than one, implying that the estimates based on the clustered sample are less precise than those for a simple random sample of the same size.

Tables 10 to 19 provide standard errors and design factors for a selection of variables from the 1998-99 FRS. In common with other tabulations the percentages and sampling errors incorporate weighting factors which are designed to compensate for non-response. An example of how to interpret them follows:

*Example: Table 10: Standard errors for household composition*

Table 10 shows that 10.8 per cent of households were composed of one female adult over pension age. The standard error is 0.2. This can be interpreted in the following manner:

It can be estimated with 95 per cent confidence that the true percentage of households composed of one female adult over pension age is:

$$10.8 \pm 2(0.2) = 10.8 \pm 0.4$$

ie if sampling error is the sole source of error, the percentage of the population composed of one female adult over pension age is between 10.4 and 11.2 per cent, with 95 per cent confidence.

The design factor for this variable was 0.98. This implies that the effect of using a clustered sample rather than a simple random sample results in a gain of precision of two per cent on standard errors. Similarly, a design factor of 1.02 would have denoted a loss of precision of two per cent.

The sampling errors shown are likely to be slightly larger than the true sampling errors because the software used for the calculation does not take into account the improvement in precision due to post stratification.

In addition to sampling errors consideration should also be given to non-sampling errors. As is clear from the above discussion, the sampling errors generally arise through the process of random sampling and the influence of chance. Non-sampling errors arise from the introduction of some systematic bias in the sample as compared to the population it is supposed to represent. Besides response biases, considered above, there are several potential sources of such bias such as inappropriate definition of the population, misleading questions, data input errors or data handling problems - in fact any factor that might lead to the survey results systematically misrepresenting the population. There is no simple control or measurement for such non-sampling errors although the risk can be minimised through careful application of the appropriate survey techniques from questionnaire and sample design through to analysis of results.

**Table 10: Standard errors for household composition**

Household composition	Percentage of all households	Standard error	Design factor
<b>Households without children</b>			
One male adult			
over pension age	3.2	0.10	0.86
under pension age	10.0	0.26	1.31
One female adult			
over pension age	10.8	0.20	0.98
under pension age	4.8	0.16	1.13
Two adults			
both over pension age	9.4	0.20	1.04
one over pension age	4.9	0.15	1.05
both under pension age	18.6	0.23	0.89
Three or more adults	9.3	0.24	1.25
<b>Households with children</b>			
One adult			
one child	2.7	0.10	0.93
two children	2.2	0.09	0.93
three children or more	1.0	0.06	0.91
Two adults			
one child	6.6	0.16	0.98
two children	8.5	0.17	0.92
three children or more	4.0	0.13	1.00
Three or more adults			
one child	2.6	0.12	1.14
two children	1.0	0.07	1.06
three children or more	0.5	0.05	1.07
<b>Total households without children</b>	71.0	0.32	1.07
<b>Total households with children</b>	29.0	0.32	1.07
<b>Total households</b>	<b>100.0</b>		

**Table 11: Standard errors for Council Tax band**

Council Tax band	Percentage of all households	Standard error	Design factor
Band A	25.2	0.40	1.39
Band B	20.0	0.34	1.29
Band C	19.5	0.31	1.18
Band D	15.4	0.28	1.17
Band E	9.3	0.23	1.20
Band F	4.6	0.16	1.16
Band G	3.2	0.15	1.29
Band H	0.7	0.08	1.45
Not valued separately	2.1	0.19	2.01
<b>Total households</b>	<b>100.0</b>		

**Table 12: Standard errors for gross weekly household income**

Gross weekly household income	Percentage of all households	Standard error	Design factor
Wages and salaries	57.2	0.37	1.13
Self employment income	11.1	0.23	1.11
Investments	67.9	0.40	1.30
State Retirement Pension plus any IS	29.8	0.32	1.06
Other pensions	26.1	0.32	1.10
Social Security disability benefits	16.0	0.26	1.07
Other Social Security benefits	48.3	0.36	1.09
Other sources	15.7	0.29	1.21
<b>Total households</b>	<b>100.0</b>		

**Table 13: Standard errors for sources of income**

Source of income	Percentage of average gross weekly household income	Standard error	Design factor
All sources of income	100	0.00	
Wages and salaries	63	0.48	1.39
Self employment income	9	0.34	1.17
Investments	3	0.12	1.40
State Retirement Pension plus any IS	6	0.10	1.19
Other pensions	7	0.18	1.29
Social Security disability benefits	3	0.06	1.28
Other Social Security benefits	6	0.11	1.29
Other sources	2	0.11	1.64
<b>Total households</b>	<b>22,913</b>		

**Table 14: Standard errors for benefit receipt**

<b>Benefit units by benefit receipt</b>	<b>Percentage of all benefit units</b>	<b>Standard error</b>	<b>Design factor</b>
Family Credit	2.3	0.09	0.99
Income Support	10.5	0.20	1.08
Housing Benefit	14.5	0.24	1.13
Council Tax Benefit	18.5	0.26	1.11
Retirement Pension	24.0	0.29	1.12
Widow's Benefit	0.8	0.05	0.93
Jobseeker's Allowance	3.4	0.13	1.19
Incapacity Benefit	5.6	0.15	1.08
Severe Disablement Allowance	0.8	0.06	1.11
Attendance Allowance	2.9	0.10	0.98
Invalid Care Allowance	1.3	0.07	1.02
Disability Living Allowance (care component)	3.9	0.12	1.02
Disability Living Allowance (mobility component)	4.4	0.13	1.05
Industrial Injuries Disablement Benefit	0.8	0.05	0.93
War Disablement Pension	0.6	0.05	1.07
Child Benefit	23.1	0.27	1.06
On any income related benefit	23.7	0.30	1.17
On any non-income related benefit	54.7	0.39	1.29
On any benefit	59.3	0.39	1.31
No benefits	40.7	0.39	1.31
<b>Total benefit units</b>	<b>27,292</b>		

**Table 15: Standard errors for tenure and accommodation type**

Tenure and type of accommodation	Percentage of all households	Standard error	Design factor
<b>Rented accommodation</b>			
<b>Rented from:</b>			
Council	17.0	0.33	1.33
Housing Association	5.0	0.22	1.53
All social sector rented tenants	22.0	0.34	1.24
<b>Rented privately</b>			
Unfurnished	5.2	0.18	1.23
Furnished	5.3	0.25	1.69
All rented privately	10.5	0.28	1.38
<b>Accommodation</b>			
House or bungalow			
Detached	1.3	0.08	1.07
Semi-detached	7.4	0.21	1.21
Terraced	9.8	0.24	1.22
All houses and bungalows	18.6	0.30	1.17
Flat or maisonette			
Purpose built	10.5	0.27	1.33
Non-purpose built	2.6	0.16	1.52
All flats or maisonettes	13.1	0.33	1.48
Other accommodation	0.8	0.13	2.21
<b>Owner occupiers</b>			
<b>Tenure</b>			
Buying with mortgage	40.5	0.38	1.17
Owned outright	27.0	0.34	1.16
All owners	67.5	0.39	1.26
<b>Accommodation</b>			
House or bungalow			
Detached	19.4	0.35	1.34
Semi-detached	23.3	0.38	1.36
Terraced	18.7	0.37	1.44
All houses and bungalows	61.4	0.43	1.34
Flat or maisonette			
Purpose built	4.3	0.17	1.27
Non-purpose built	1.3	0.09	1.20
All flats or maisonettes	5.6	0.19	1.25
Other accommodation	0.4	0.05	1.20
<b>Total households</b>	<b>100.0</b>		

**Table 16: Standard errors for weekly housing costs**

Weekly housing costs	Percentage of all households	Standard error	Design factor
Under £20 a week	32.6	0.36	1.16
£20 but under £40 a week	16.9	0.30	1.21
£40 but under £60 a week	23.2	0.34	1.22
£60 but under £80 a week	11.9	0.23	1.08
£80 but under £100 a week	6.4	0.18	1.11
£100 but under £150 a week	6.0	0.17	1.08
£150 a week or more	3.0	0.14	1.24
<b>Total households</b>	<b>100.0</b>		

**Table 17: Standard errors for types of account held**

Type of account	Percentage of all individuals	Standard error	Design factor
Current account	64.0	0.30	1.45
Post Office account	6.5	0.15	1.42
TESSA	8.3	0.16	1.35
Other bank/building society	48.6	0.38	1.77
Gilts	0.6	0.04	1.20
Unit trusts	3.3	0.10	1.30
Stocks and shares	16.5	0.22	1.38
National Savings Bonds	4.6	0.11	1.22
Save As You Earn	0.8	0.04	1.04
Premium Bonds	15.7	0.22	1.41
PEPs	8.7	0.17	1.40
Any type of account	79.1	0.29	1.66
No accounts	20.9	0.29	1.66
<b>Total individuals</b>	<b>100.0</b>		

**Table 18: Standard errors for employment status**

Adults by employment status	Percentage of all adults	Standard error	Design factor
Employee			
full time	39.6	0.31	1.28
part time	11.6	0.17	1.07
Self employed			
full time	5.6	0.14	1.23
part time	1.7	0.07	1.09
All in employment			
full time	45.3	0.31	1.25
part time	13.3	0.18	1.07
ILO unemployed	3.8	0.12	1.26
Retired	22.2	0.27	1.31
Student	1.7	0.15	2.34
Looking after family home	4.8	0.1	0.94
Permanently sick or disabled	5.7	0.14	1.22
Temporarily sick or disabled	0.4	0.03	0.96
Other inactive	2.8	0.09	1.10
<b>Total adults</b>	<b>100.0</b>		

**Table 19: Standard errors for ethnic group of adults**

Ethnic group	Percentage of all adults	Standard error	Design factor
White	94.2	0.28	2.41
Black	1.7	0.11	1.71
Indian	1.6	0.16	2.57
Pakistani/Bangladeshi	1.1	0.13	2.51
Other	1.4	0.10	1.71
<b>Total adults</b>	<b>100.0</b>		

## GSS Harmonisation Project

In 1995, a number of Government Departments came together to discuss the best way of making the results of major official surveys comparable. This led to the Harmonisation Project, whereby the inputs – the questions and related interviewer instructions and edit checks - used in the major surveys, and outputs – the concepts for analysis and publication - from them, is identical in as many of the surveys as possible. A list of the current harmonised questions may be found at <http://www.statistics.gov.uk/harmony/harmonfp.asp>.

Different surveys have different purposes and hence cover topics in different depths. Harmonised questions are designed to provide the recommended minimum information to allow common classifications and facilitate the analysis of data from different surveys in combination. Not all surveys will include questions on all topics or in every year, but the recommendation is that, where a topic is covered, harmonised questions should be included wherever possible.

Some surveys will require further detail on topics than can be obtained from the harmonised questions alone. It will normally be the case that such surveys already ask for that detail. The harmonised questions have been designed so that the surveys which ask for more detail can either derive them, without asking them directly, or combine them with the further detail, without adding to the length of interview.

The FRS has actually set the standard for a number of harmonised questions, but many more on the FRS are being changed to fall in line with Government surveys such as the Survey of English Housing, the Family Expenditure Survey and the Labour Force Survey.

The table below summarises the status of harmonised outputs in this and previous FRS reports.

Topic	Outputs	Inputs (questions)
Accommodation type	1995-96	1996-97
Age last birthday	1997-98	Already harmonised, although full household grid only introduced in 1996-97; date of birth for those under 20 years from 1996-97.
Consumer Durables	N/A	1997-98. However, these are rotated off the FRS biannually. Therefore revised harmonised questions will not be implemented until 2001-02.
Economic Activity (ILO)	1996-97	1996-97
Educational Attainment	N/A	Trial questions piloted in 1998-99.
Ethnic origin	1994-95	1996-97
Gender	Already harmonised.	Already harmonised.
Geography - use of GORs	1995-96, although North West and Merseyside combined since 1996-97.	
Health	No tables use this classification, see Table 19 for further information.	1996-97
Household income	1996-97	Detailed questions harmonised with FES from 1996-97.
Household Reference Person	N/A	Piloted in 1999-2000, but will not affect outputs until 2001-02 as the likely impact of the change will have to be quantified.
Household response unit	Already harmonised.	Already harmonised.
Housing costs and benefits (Council Tax, Rent, Housing Benefit and Mortgages)	Output bands harmonised 1995-96 for rents. Differences in method of calculation of mortgage interest.	1997-98, also includes additional questions.
Industry, occupation, socio-economic classifications	Already harmonised.	Already harmonised. In questions on job description, present tense used for those currently in work.

Topic	Outputs	Inputs (questions)
Legal marital status	No tables use this breakdown.	1996-97
Length of residence	1996-97	1996-97 (original proposal) revised 1996-97 to include 20 year or more cut off.
Length of time since last in paid work	1996-97	1996-97
Living arrangements	1994-95	1997-98
Reference period	Already harmonised.	
Social Security benefits	More detailed breakdown used.  Jobseeker's Allowance (contributions based) included with non-means tested benefits category.  Definition of disability benefits is different from harmonised classification, which includes Invalid Care Allowance (ICA). ICA is paid to the carer not the severely disabled person being cared for.	Already harmonised.
Tenure	1994-95 (although minor differences in treatment of cases where accommodation goes with the job of someone in the household).	1996-97
Time in present job	No tables use this classification.	1996-97
Usual hours in main job	1996-97	1996-97
Vehicle ownership/continuous use	No tables use this classification, although tables showing consumer durables include light vans in the car category.	1996-97

Harmonised output categories for marital status, ethnic group, Social Security benefits, tenure, accommodation type, housing cost bands and length of residency have been included in previous publications although in some cases these were based on non-harmonised questions. From 1997-98 many more harmonised questions were included in the survey particularly in the area of housing costs and living arrangements.

Another area where the FRS output categories differ from harmonisation proposals is disability. The harmonised question for dealing with the health of individuals is:

*Do you have any long-standing illness, disability or infirmity? By long-standing I mean anything that has troubled you over a period of time or that is likely to affect you over a period of time?*

On the FRS, individuals who answer 'yes' to this question are also asked if this illness or disability limits their activities

in any way and if they are registered disabled. In addition all adults below pension age are asked if they are restricted in the work they can do by some illness, injury or disability. Responses to all these questions are used to define sick or disabled adults for use in the household composition and economic status classifications (ILO definitions are based on those who have stated they cannot work or look for work because of sickness/disability). Table 20 shows health status by age group using the different definitions. As can be seen, the additional query about whether activities are limited by the illness or disability does affect the classification. The final column shows also that up to state retirement age there are some individuals who consider themselves restricted in the work they can do but have not categorised themselves as disabled.

**Table 20: Long Standing Limiting Illness**

Percentage of all adults			
Age group	Long Standing Illness	Long Standing Limiting Illness	Long Standing Limiting Illness or Restricted in Work
16 to 24	13	7	7
25 to 34	16	9	10
35 to 44	22	14	14
45 to 54	32	20	20
55 to 59	43	29	30
60 to 64	52	35	36
65 to 74	58	39	39
75 to 84	67	51	51
85 or over	77	67	67
<b>All adults</b>	<b>33</b>	<b>22</b>	<b>22</b>
<b>Total adults</b>	<b>41,800</b>	<b>41,800</b>	<b>41,800</b>

### Differences due to survey design features

Despite the use of harmonised inputs, there will inevitably be differences in the outputs of the varied surveys as a result of differences between the design of the surveys (and due to sampling variability). Primary among these is sample size; where the direct results of a survey are grossed up to estimate a total population, assumptions have to be made of the relevance of identified features to that population and it is reasonable to expect a wider margin of error from a sample of, say, 5,000 than from one of 50,000.

Relevant survey design features include:

- question wording and context effects;
- definition differences;
- non-response bias;
- geographical coverage;
- sampled population;

- mode effect;
- acceptance of proxy information;
- treatment of multi-households;
- unit of analysis;
- field practices;
- item non-response;
- time period;
- organisational effect.

Further details are provided in Amanda White and Sarah McCreith's 'An initial look at harmonised survey data', Survey Methodology Bulletin 43 (July 1998).

The GSS is continuing to try to quantify these likely differences, and advise users of official statistics of where such differences are likely to arise, the reasons for them, and the potential size of any.