# United Kingdom 1994-99: Survey Information

This document is based upon the Family Resources Survey reports (see www.dwp.gov.uk/asd/frs).

**Table of contents:**



## A. General characteristics

Official name of the survey/data source:
Family Resources Survey (FRS)

Administrative Unit responsible for the survey:


Department for Work and Pensions
Analytical Services Directorate - ASD Income Analysis 1
The Adelphi - 1-11 John Adam Street – London WC2N 6HT
United Kingdom
e-mail: Team.FRS@dwp.gsi.gov.uk
Phone: +44- 20 7962 8991


The Family Resources Survey [FRS] is a continuous survey, which was launched in October 1992 by the then Department of Social Security, DSS [now the Department for Work and Pensions, DWP]. The FRS is commissioned by the DWP to meet their specific data requirements and has been carried out jointly by the Office for National Statistics [ONS] (Social Survey Division) and the National Centre for Social Research (then SCPR) since 1992. Data from past survey years can be obtained for academic use from the UK Data Archive. To view the original questionnaires, please look at http://qb.soc.surrey.ac.uk/surveys/frs/frsintro.htm#one.

General design

The survey is organized on a quarterly basis and runs over the fiscal year. Its primary function is to collect information on the resources of households, that is, income received from all sources including wages and salaries, State supports, private (occupational and personal) pension schemes and investments. It is crucial for all DWP uses of the survey that this information on income is as accurate as possible. The FRS is also a valuable

source of information for analyses of the nature of the support given by the Government to individuals, and in particular, the types and combinations of state supports that households and benefit units receive. It also puts state supports in the context of other sources of income received.

The main advantages of the FRS (compared to the Family Expenditure Survey, FES and the General Household Survey, GHS, for example) are the larger sample size, the ability to identify benefit units and the inclusion of more detailed questions relating to social security benefits.

Changes in the 1999-2000 survey:

- The most important change in this year's survey is the rotation of questions: certain questions are asked every other year in order to reduce the length of the questionnaire and the burden on respondents; for 1999-2000 questions on NHS treatment and travel to work have been rotated off.
- Questions on children's savings have been removed, while supplementary questions on Individual Savings Accounts (ISA) and accounts capable of accepting Automated Credit Transfers (ACT) were included.
- Some of the care questions have been changed.
- Since October 1999, Family Credit - the benefit for working parents on low income – has been phased out, and replaced by the new Working Families' Tax Credit, administered by the Inland Revenue rather than the DSS; the same is true for Disability Working Allowance, which has been replaced by Disabled Persons' Tax Credit; from October 2000 they can be received through the wage packet, as a tax adjustment, rather than as a cash benefit; however, for the first year the Inland Revenue continued to pay people the benefit as before, by cheque or direct debit; in the survey thus these are both combined with the previous benefit.
- There has been a definitional change to bring a person's employment status in line with the classical ILO definition: individuals between the state pension age and the age of 70 who are looking for work are now asked whether or not they are able to start in the next two week (previously they were classified as inactive by default).

## B. Population, sampling size and sampling methods

Population

The Family Resources Survey sample aims to cover private households in Great Britain. Like some other household surveys, the area to the North of the Caledonian Canal and all of the Scottish Highlands and Islands 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 26 regions and also by three other variables derived from the 1991 Census of Population. Stratifying ensures that the 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 or 13. Within each of these six bands, the PSUs are ranked by the adult economic activity rate and formed into three further bands, resulting in 18 bands for each region. These are then ranked according to the proportion of unemployed men. 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 below summarises the stratifiers.

| | |
|---|---|
| Regions | 19 in England - (Metropolitan/non-Metropolitan/4 in London)<br>2 in Wales<br>5 in Scotland |
| Socio-economic groups | 1 Employers/Managers in large establishments<br>2 Employers/Managers in small establishments<br>3 Professional workers (self employed)<br>4 Professional workers (employees)<br>5 Non-manual ancillary workers, foremen and supervisors<br>13 Farmers (employers and managers) |
| Economic activity rate | |
| Male unemployment rate | |

Within each PSU a sample of addresses is selected. In 1999-2000, 25 addresses were selected per PSU. This means that nationally there was approximately a one in 571 chance of an address being selected.

Each year, one half of the PSUs are retained from the previous year's sample, but with new addresses chosen; 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.

Response and Sample size

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 (i.e. excluding the assets section of the questionnaire). Proxy

interviews are accepted only under restricted circumstances. In 1999-2000, for those households classed as fully co-operating, proxy responses were obtained for 15 per cent of adults. 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.

The table below summarises the household response for the 1999-2000 FRS.

| | Number of Households | Percentage of effective sample |
|---|---|---|
| Sampled addresses | 42,472 | |
| Ineligible addresses | 4,602 | |
| Effective sample (eligible households) | 37,870 | 100% |
| of which: fully co-operating households | 24,988 | 66% |
| partially co-operating households | 432 | 1% |
| refusals | 10,819 | 29% |
| households with no contact | 1,613 | 4% |

The original sample chosen for 1999-2000 consisted of 42,472 addresses. However, 4,602 were then found to be ineligible because they were not defined as private households or were empty households. This left an effective sample of 37,870 households. Of these, 24,988 fully co-operated (66 per cent), 432 only partially co-operated (one per cent) and 10,819 refused to proceed with the interview (29 per cent). The interviewer was unable to make contact with 1,613 households (4 per cent).
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 an 'invasion of privacy' (17 per cent); followed by 14 per cent who said they 'couldn't be bothered' and 14 per cent who 'don't believe in surveys'. Concerns about confidentiality were only raised by 4 per cent of households. 6 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 11 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 23,970 households.

The table below shows response rates broken down by Government Office Region for the 1999-2000 FRS.

| **Government Office Region** | **Percentage of households** |
|---|---|
| North East | 72 |
| North West and Merseyside | 67 |
| Yorkshire and the Humber | 67 |
| East Midlands | 66 |
| West Midlands | 64 |
| Eastern | 67 |
| London | 60 |
| South East | 66 |
| South West | 68 |
| England | 66 |

| | |
|---|---|
| Wales | 71 |
| Scotland | 65 |
| **Great Britain** | **66** |

Response rates are calculated as the number of fully co-operating households divided by the number of eligible households, all multiplied by 100.
The overall response rate for the FRS for the year 1999-2000 was 66 per cent.
The region with the highest response rate was the North East, where 72 per cent of all households selected responded fully, followed closely by Wales (71 per cent). The region with the lowest response rate was London where only 60 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.

Weighting procedure: grossing up

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 DWP 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 DLTR 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. e.g. an adjustment has been made to the demographic data to exclude people not resident in private households.

| **Control variables used to generate grossing factors** | | |
|---|---|---|
| Variable | Groupings | Source of data |
| Age/sex/marital status | Single men: <35, 35-59, 60+<br>Single women: <35, 35-64, 65+<br>Couples: <65, 65+ | Office for National Statistics, Government Actuaries Department |

| | | |
|---|---|---|
| Lone parents | Male, female | DWP estimates |
| Families | No. of couples with children | DWP estimates |
| Tenure type | LA renters, private renters, owner occupiers | DTLR estimates |
| Council Tax Band | A, B, C-D, E-H | DTLR estimates |
| Region | London, other | DTLR estimates |

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.

## C. Data collection and acquisition

Data Collection methods
The Consortium of Social Survey Division (SSD) of the Office for National Statistics (ONS) and the National Centre for Social Research (formerly SCPR) have been conducting fieldwork for the FRS since 1992. In September 1996 and September 2000 the FRS contract was tendered as part of the good practice in government programme. The consortium was successful both times and currently holds the contract until 2004 with the option of a two year extension.
Interviews are carried out jointly on behalf of the DWP by interviewers from ONS 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 ONS and also a slit sample test conducted jointly by the ONS 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 1999-2000, 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 length is around one hour and 20 minutes, but the time will vary according to the size of household and its circumstances. The most common length of interview in 1999-2000 was recorder at 60 minutes.

The questionnaire itself is divided into three parts:

- Household Questionnaire: 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.
- Benefit Unit questionnaire: Second, the individual schedule, which is addressed to each adult in turn and asks questions about employment, benefits, pensions, investments and other income.
- Assets Questionnaire: A final section goes on to ask the value of investments for relevant respondents.

To contain the length of the overall questionnaire, and to reduce the respondent burden of an overlong interview, FRS users have agreed to rotate off blocks of questions. 'Rotated' sections of the questionnaire will be asked every other year, rather than every year.

Rotated off for 1999-2000 are "NHS treatment" and "Travel to Work". Rotated off for 2000-01 will be "Vehicle ownership" and "Household Durables", and in 2001-02 again "NHS treatment" and "Travel to work".

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.

When answering questions on income from employment, employees consulted a payslip in 54 per cent of cases in 1999-2000. However, it should be noted that in 8 per cent of cases they simply did not have a payslip to consult.

In recording data on benefit and payable tax credit receipt, some form of documentation (an order book, a letter from the DSS or Benefits Agency or a bank statement) was consulted for 67 per cent of all benefits received.
The questionnaire records consultation of documentation for questions relating to Council Tax. 48 per cent of households consulted a Council Tax bill or statement in answering questions on their Council Tax payments.
In addition, self-employed respondents are asked if they have documentation when they provide information about the profit or loss of their business. Of the 75 per cent of self employed respondents who had prepared business accounts, 31 per cent were able to refer to such documentation.

## D. Definition of the survey units

There are three levels in the survey: the household, the benefit unit and the individual.
The definition of a household used in the FRS is a single person or group of people living at the same address who either share one meal a day or share the living accommodation. So, for example, a group of students with a shared living room ewould be counted as a single household even if they did not eat together, but a group of bedsits at the same address would not.
A household will consist of one or more benefit units, which in turn consists of a number of individuals (adults and children). "Benefit unit" is a standard DSS term, which relates to the tighter family definition of "a single adult or couple living as married and any dependent children". A dependent child is aged under 16 or under 19 if still in full time non-advanced education. So, for example, a man and wife living with their young children and an elderly parent would be one household but two benefit units. It should be noted that "benefit unit" is used throughout the publication as a description of groups of individuals regardless of whether they are in receipt of any Social Security benefits.

## E. Contents

- Tenure and Address Information
- Details of Rent or Mortgages paid
- Council Tax and other Property Charges
- Insurance Policies
- Ownership of Household Durables
- Ownership of Vehicles
- Welfare Services used
- Childcare arrangements
- Any help given or received by members of the household
- Health and Ability to work
- Education, Grants and Loans
- Employment Status and Current Situation
- Job Description
- Traveling to Work
- Pensions

- State and Other Benefits
- Income from Other Sources
- Savings and Investments
- Assets

## F. Quality of data

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 1999-2000 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 . 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.

A study based on 1995-96 and 1996-97 data compared response rates in relation to the ACORN code of a household. ACORN is a geo-demographic classification system developed by CACI Marketing Systems which classifies areas into 54 types, using Census data items such as age, sex, marital status, occupation, education, home ownership, car ownership, family structure and ethnic group. This suggested that wealthier households are less likely to respond to the survey than poorer households, but none of the 54 types had response rates more than 10% from the average.

*FRS non-response and Council Tax band:* Comparisons were made between 1999-2000 FRS data and administrative data on the number of households within each Council Tax band. The results showed that FRS proportions were similar to those obtained from examining administrative data.

*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. In 1999-2000, 72% of households which

refused to take part in the full survey answered questions from the non-response form. In addition, interviewers filled in a short questionnaire on all refusing and non-contactable households, based on observation or on information gleaned from neighbours. Key findings were:

- non-responding households were more likely to live in flats or maisonettes than respondents;
- refusing households were less likely than responding households to contain three or more adults and were less likely to contain any children. They were also significantly more likely than responding households to contain members aged 60 or over;
- there was no significant difference in ethnic origin between refusal and respondant households;
- refusing households were significantly more likely than responding households to have at least one member working.

*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. 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.

Comparisons of the FRS and other surveys with the 2001 census will be carried out when the information is available.

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 total 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 out below.

1. *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. This 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.

2. *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.

3. *Stage three: data conversion*

Before it can be validated further, FRS data must be converted from its CAPI format into SAS readable tables. Using DWP specifications SAS tables are created by the Office for National Statistics, each table recording information from different parts of the questionnaire. Both the DWP 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.

4. *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, 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 most 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 three 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 now have to be carried out are 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.

5. *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.

The table below illustrates the extent of the problem of missing values for the 1999-2000 FRS.

| | Values | Percentage of values |
|---|---|---|
| **Responses** | | |
| Expected number of responses | 11,938,060 | 100 |
| Valid responses | 11,880,641 | 99.5 |
| Missing values (don’t know/refused) | 57,419 | 0.5 |
| **Treatment of missing values** | | |
| Hotdeck | 41,220 | 72 |
| Bulk Edits | 9,638 | 17 |

| | | |
|---|---|---|
| Other imputation method | 582 | 1 |
| Benefit editing | 1,701 | 3 |
| Left as Missing | 4,278 | 7 |

It should be noted that out of over 11 million set values in the FRS database, only 0.5 per cent were originally recorded as either 'don't know' or 'refused'. Out of 57,419 missing values, approximately 93 per cent were imputed.
A combination of methods of imputation were used for the 1999-2000 FRS data. The main ones are summarised below in the order in which they were used.

*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* - 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.

*'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.

*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.

6. *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 1999-2000 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. To deal with this, 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). An assessment was then 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.

*7. Stage seven: Derived Variables*

Derived variables (DVs) are customised variables in the FRS datasets, derived using information collected in the survey and some from auxiliary sources such as the Department for Transport, Local Government and the Regions' deprivation indices. They are created at the data users' requests, as the main purposes of the derived variables are to make it easier for the users to carry out analysis and to ensure consistent definitions are used in all FRS analyses. For example, INDINC is a DV that sums all components of income to give an individual's total income - information on respondents' income from various sources is collected in the survey. As new information is collected in the survey, the relevant DVs are updated if necessary. Any unnecessary DVs are removed from the dataset.

Information on Travel to Work costs, National Health Service treatment, Consumer durables and vehicle ownership is collected every other year. Therefore, the DVs using this information are only created in the year for which the information is available.

There are over 200 DVs in the 199-2000 dataset.

Quality of Benefits Data

As part of the data validation process, comparisons are made between the FRS and other data sources.

The table below shows a comparison of 1999-2000 FRS benefit recipients compared to administrative data.

| | **1999-00 FRS** | | **1999-00 FRS** | | **DSS administrative data** | |
|---|---|---|---|---|---|---|
| | **Grossed** | | **Sample** | | | |
| **Benefits received** | **Number** | **%** | **Number** | **%** | **Number** | **%** |
| Family Credit/Working Families' Tax Credit | 677,000 | 1.5 | 802 | 1.8 | 784,000 | 1.8 |
| Income Support | 3,181,000 | 7.3 | 3,391 | 7.7 | 3,547,000 | 8.1 |
| Househing Benefit | 4,244,000 | 9.7 | 4,585 | 10.4 | 4,243,000 | 9.7 |
| Council Tax Benefit | 5,370,000 | 12.3 | 5,802 | 13.2 | 5,083,000 | 11.6 |
| Jobseekers' Allowance | 971,000 | 2.2 | 843 | 1.9 | 1,097,000 | 2.5 |
| Retirement Pension | 9,619,000 | 22.0 | 10,596 | 24.0 | 10,131,000 | 23.1 |
| Widow's Benefit | 247,000 | 0.6 | 287 | 0.7 | 247,000 | 0.6 |
| Incapacity Benefit | 1,492,000 | 3.4 | 1,479 | 3.4 | 1,538,000 | 3.5 |
| Severe Disablement Benefit | 241,000 | 0.6 | 223 | 0.5 | 370,000 | 0.8 |
| Attendance Allowance | 834,000 | 1.9 | 890 | 2.0 | 1,267,000 | 2.9 |
| Invalid Care Allowance | 378,000 | 0.9 | 374 | 0.8 | 375,000 | 0.9 |
| Disability Living Allowance (Care component) | 1,205,000 | 2.8 | 1,199 | 2.7 | 1,483,000 | 3.4 |
| Disability Living Allowance (Mobility component) | 1,401,000 | 3.2 | 1,413 | 3.2 | 1,808,000 | 4.1 |
| Child Benefit | 6,873,000 | 15.7 | 7,920 | 18.0 | 7,090,000 | 16.2 |
| **Total adults** | **43,786,251** | **100** | **44,107** | **100** | **43,786,251** | **100** |

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. Adjustments are made to try and eliminate some of the differences between the two sources. For example, the denominator for the administrative and the FRS data in the table above 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.

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. For example, if a sample characteristic is distributed differently by postal sector (i.e. 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.

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. Similarly a DEFT less than one implies the estimate is more precise than would be obtained from a simple random sample.

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.

## G. Uses of FRS

The FRS is used widely across the Department. The main uses are:

The Department's Policy Simulation Model (PSM), used extensively by DWP analysts for policy evaluation and costing of policy options. FRS responses are uprated to current prices, benefits and earnings levels and calibrated to DWP Departmental Report forecasts of benefit caseload. Using FRS data has made it possible to model some aspects of the benefit system which could not be done previously,: eg income related benefits' severe disability premia and allowances for child care costs. In addition to their use in formal modelling, FRS data play a vital role in the analysis of patterns of benefit receipt for policy monitoring and evaluation and benefit forecasting. Examples are the extent of multiple benefit receipt and the distribution of individual benefits.

Households Below Average Income (HBAI) methodology. The income measure used is based on weekly net (disposable) equivalised household income (ie income adjusted for household size and composition by means of equivalence scales). The HBAI data set also forms the basis of the Pensioners' Income Series, the Department's analysis of trends in components and levels of pensioners' incomes and the Individual Income Series.

Individual Income series published by the Women and Equality Unit. The Individual Income series provides estimates of the gross, net and disposable incomes of men and women, whether living as couples or single persons.

Estimates of take-up of income related benefits. Figures are based on a combination of administrative and survey data. The FRS provides information about people's circumstances, which is used to estimate numbers of people who are not claiming benefits to which they appear to be entitled.

The FRS has also been used as a sampling frame for follow up studies to look at particular groups. The most recent example is a follow-up survey of pensioners entitled to, but not claiming, Minimum Income Guarantee, which is currently underway. The survey remit was to look at the circumstances of this group and establish why they were not claiming benefits to which they appeared to be entitled. The largest example is the Disability Survey, which re-interviewed over 7,000 disabled respondents who appeared in the FRS between July 1996 and March 1997. The survey provides a detailed picture of type and severity of disability, extra needs and participation in leisure activities of the disabled. Merged with FRS information, findings were used to measure and analyse receipt of disability benefits and gather information to enable more accurate forecasting of expenditure.

Although primary users of FRS data remain within the DWP, the survey is increasingly being used outside the Department. The data set is provided to other government departments on request. It is also accessed by researchers and analysts outside government through the Data Archive at Essex University.