

Stages of the process for the production of the Family Resources Survey 2009/10 based publications

Producing the Family Resources Survey dataset and its associated publications involves a number of concurrent processes, including work on a range of different years of data at the same time.

1. Pre-fieldwork

Summer to autumn 2008

All known FRS users (DWP and OGDs) are invited to put forward suggestions for changes to the FRS questionnaire for the next survey year.

For the 2009/10 survey, around 75 proposals were received, of which over 50 were accepted following discussion.

Winter 2008

Survey contractors¹ update the interviewer material and programme the changes to the FRS questionnaire.

2. Fieldwork

April 2009 to March 2010

Questionnaire in the field. The interviewers submit data captured from completed interviews to their home organisation daily. The home organisation editors then go through the questionnaires and make any corrections required, based on interviewer notes. Editing is completed within approximately a month of the final case being received.

In 2009/10, around 48,000 addresses were sampled, of which around 43,000 households were eligible, and of which around 25,000 were fully co-operating and used in analysis. A range of techniques are used by interviewers to encourage response. The average interview length in 2009/10 was around 1 hour and 24 minutes.

¹ The contract for FRS fieldwork for the survey in Great Britain has been re-tendered four times since the launch of the survey in 1992, most recently in 2010. A consortium made up of the Office of National Statistics and the National Centre for Social Research won that contract, for 2011-12 to 2014-15 inclusive. The fieldwork for the survey in Northern Ireland is managed by the Department for Social Development (DSDNI) and is currently carried out by the Northern Ireland Statistics and Research Agency.

3. Transfer of data and preparation

June 2009 to October 2010

As soon as each month's fieldwork and editing is completed, the survey contractors convert the data collected from the questionnaire into the datasets delivered to DWP. This involves converting all relevant monetary amounts to weekly equivalents (except those with missing, vague or lump sum period codes) and assigning Standard Occupational Classifications (SOC). A series of detailed validation checks is run on each dataset before delivery to DWP.

Delivery is on a monthly basis, with a six month re-supply in March 2010 and an annual re-supply in October 2010 to correct for errors that are identified. These re-supplies provide an opportunity to correct for errors or anomalies that, due to the complex nature of the data, were not picked up in earlier checks.

The full number of responses delivered as part of the FRS dataset is in excess of 15 million.

April 2010 to December 2010

The FRS Team prepares the test dataset in preparation for release to a Quality Assurance (QA) Group:

- Benefit editing - imputation of missing benefit amounts and dividing the combined benefit amounts in to the reported benefits
- Validation - period codes, credibility checks and outliers
- Derived variables - functions of constants and/or variables collected
- Imputation - of missing values
- Grossing - calculation of weights
- User documentation - updated to ensure all changes to the questionnaire and processing are reflected

The FRS team and the QA Group analyse the dataset and raise issues with the survey contractors for investigation. Errors found are corrected in the annual resupply.

While 99 per cent of responses are valid, missing income data need to be imputed and benefit editing needs to take place to make sure that total incomes can be calculated robustly. In total, around 120,000 values were imputed through a variety of methods.

4. Testing of dataset

Summer 2010

Six month test dataset released

Winter 2010

Two versions of 12 month test dataset released. The first version is based on the six-month re-supply and the monthly deliveries. It does not contain the re-issued cases and is normally released at the end of October. The second version, based on the annual resupply, contains approximately 600 previously unseen cases and is released by Christmas.

Summer 2010 to winter 2010: FRS testing taken forward

The test dataset access is restricted to the FRS QA group, who must commit to undertaking QA activity. A pack of QA material is sent to the QA group, and a number of meetings are held to discuss any issues raised by the QA group members.

The FRS Team maintain a list of all problems found with the dataset. The FRS team leader manages the list to ensure all problems are resolved.

FRS users who proposed changes that were made to the questionnaire are asked to check the dataset and confirm that the changes have fed through to the dataset as expected. Proposers are asked to respond to a series of specific questions about the changes. If proposers do not test the changes the variable may be removed from the final dataset

The FRS team leader looks at each change to the questionnaire documented in the OTV (One True Version), details the variables affected by each change, and compares the dataset with the previous year to ensure that all changes have fed through as expected. Any issues found are raised with the survey contractors.

For 2009/10, over 80 issues were raised as part of Quality Assurance checks. This includes issues raised by the FRS team, and issues raised by users. Issues considered at this point include unusual movements in the data, use of any external data sources and methodological changes. On occasion, if uncorrected, issues raised at this point in the process could have noticeable effects on the headline results.

As one example of this, the FRS team noticed that the total capital of the benefit unit variable, which is used extensively, was very sensitive to the interest rate assumptions. An investigation was carried out which resulted in more robust assumptions being made. We also make improvements here when possible, such as harmonising the way the FRS and Households Below Average Income (HBAI) methodologies edit unfeasibly high Council Tax Benefit amounts.

Summer 2010 to spring 2011: HBAI team checks

The Institute for Fiscal Studies (IFS) independently reproduce the HBAI dataset from the FRS dataset (to quality assure HBAI key statistics) which is checked at a case level with the HBAI dataset calculated by the DWP. This was brought in to ensure the statistics are independently quality assured. Methodological changes are also discussed and agreed with the IFS. In 2009/10, this included work to agree an improvement to the adjustment to top incomes using HMRC's Survey of Personal Incomes.

Any data issues discovered as part of this process are passed to the FRS team for investigation.

5. Production of reports

Mid February 2011 to May 2011

Work on the publication begins approximately mid-February when the FRS dataset has been provisionally finalised. It is provisional because often data issues are only seen when detailed analysis has been carried out. This also means that it is very much preferable to release the FRS and HBAI publications at the same time, as the high level of validation of the income statistics in the HBAI report will sometimes reveal issues with the underlying FRS dataset.

All changes to the publication are circulated to the relevant publication team (FRS, HBAI or Pensioners' Incomes (PI) Series) by the team's publication manager in advance of work beginning. A clear project plan detailing which team member is responsible for each chapter is produced.

Each chapter is independently re-run by another team member to ensure no mistakes have been made. Each chapter is quality assured by a policy facing analyst. A range of additional material is also produced to help understand the movements in the data.

6. Validation taken forward

Lead team	Process	Steps taken
FRS	FRS questionnaire consultation	FRS team works with proposer and the contractors to ensure the data will be collected as specified. These are signed off at a senior level.
	Initial validation	This is to check the dataset is as specified with all key variables populated and is structurally sound.
	Validation: Benefit editing	Uses a variety of internal and external sources including looking at year-on-year changes to check the plausibility of benefit amounts, to separate out combined amounts and to fill in

		missing values.
	Validation: Non-benefit editing	This looks in detail at the non-benefit variables, including an examination of extreme values, using rules to check the credibility of individual variables and their relationships with other variables and making sure relevant amount variables refer to a consistent period.
	Imputing missing values and deriving new variables	A variety of methods are used to fill in missing values, and further variables are derived, including through the use of externally sourced constants. These are cross-checked by another team member. Once this is done, both the benefit editing and credibility checks are repeated to ensure the imputed values make sense.
	Grossing	This scales up the data so it applies the whole population multi-purpose grossing factors which align the estimates to Government Office Region populations by age and sex. These are thoroughly checked by the FRS team and independently verified by the Institute for Fiscal Studies.
	Final FRS team validation	Full documentation of changes to the dataset is produced with comprehensive metadata and change specifications. This includes comparisons with past data, looking into statistically significant changes and producing a range of documentation about the dataset.
	Wider cross-Government validation	Proposers of new questions required to formally sign off that the new data is credible, and comparable to other sources. The FRS, HBAI, PI and Take Up teams also look at each change to ascertain its possible impact on their publication statistics.
	Issue resolution	All issues raised are carefully logged and resolved by either the contractors or the FRS team.
HBAI	Production of the Households Below Average Income (HBAI) dataset	All changes to the dataset production code are checked by a second team member.
	HBAI dataset validation with the Institute for Fiscal Studies (IFS)	This dataset agreed with IFS for each case of the FRS for each component of income is agreed with IFS, down to the last penny. Summary publication statistics are also agreed with IFS before finalising the HBAI dataset. IFS also examine changes to underlying FRS, and assess possible impact on HBAI. This validation includes the gross income

		variables which are used in the production of the Pensioners' Income series.
	Further validation of the HBAI dataset	<p>We have produced code to break down changes in components in income, by different client group across the income distribution. We do this for as many years as possible, to assess whether the changes are as we expected.</p> <p>We compare some income streams with similar variables on the FRS.</p> <p>Those users quality assuring the dataset are provided with their own dataset to check thoroughly.</p>
Take Up	Case checks	As the take-up report also relies on entitlement modelling performed by the Department's micro simulation model: the Policy Simulation Model, the take-up team undertake aggregate and case by case checks for a proportion of the sample which check income components for the FRS cases and ensure their entitlements have been modelled in accordance with the benefit regulations.
All	Publications	All publication tables have a second set of code to check that updates are correctly included. A staff member from a different team produces a third set of code that checks key estimates, new tables, plus some old tables. There is then senior sign off of the publication itself, as well as circulation to a core group of policy-facing analysts outside the production team who quality assure the statistics.

7. An assessment of the balance between user needs and timeliness

According to the United Nations' Canberra Group Handbook on Household Income Statistics², income statistics are "inevitably some of the most complex statistics produced by national and international organisations." This is reflected by the very extensive production and validation procedures described above. These procedures have been developed for three main reasons:

- Statistics derived from the Family Resources Survey are amongst the most high-profile produced by the Department for Work and Pensions and across Government, being used to monitor progress against the Child Poverty Act and two of the twelve Departmental Impact Indicators. The dataset is also used extensively for policy development and costings

² <http://www.unece.org/stats/groups/cgh.html>

across a range of Government Departments and for answering PQs and ad hoc queries.

- Because of the complexity of the analysis and the range of income sources and other variables used in the calculations, issues have arisen in earlier publications:
 - The 2006/07 Households Below Average Income and Family Resources Survey publications were delayed due to problems with FRS data on pension contributions.
 - The 2005/06 Households Below Average Income publication was re-issued due to problems with the grossing factors being applied to the dataset.

The role of the Institute for Fiscal Studies in signing off the Households Below Average Income dataset and results originally arose in part due to quality concerns with results.

- The datasets are used for a wide range of analyses beyond the published tables, and small groups of cases could affect these results. This necessitates a thorough examination of case-specific information. For many users, it is the dataset that is as important as the publications themselves.

As described above, the gap between fieldwork ending and the publication of the reports of approximately 14 months consists mostly of carrying out this validation, in addition to developing subsequent survey years' information in a continuous manner. This means that the team is working on a number of different workstreams simultaneously (see Annex A for further details). This gap is comparable with other household surveys.

There is also a need for validating the whole dataset given the large number of variables used in the derivation of the HBAI and PI datasets, so it is not possible to isolate the key variables and get headline results out earlier. It is also the case that the headline figures alone are not very useful without knowing either the changes in the percentage of a group in low income (for HBAI) or which income components are driving the change in pensioners' incomes.

We deem the risk of putting out inaccurate statistics, possibly breaching both UK and European legislation (with the possibility of large fines arising) necessitates a high level of validation. This, in combination with the complex nature of the calculations, is the primary reason for the gap between the end of the survey period and publication. Having a potentially inaccurate dataset for policy modelling may also mean we pursue a sub-optimal option or that costs or benefits are incorrectly assessed. It is therefore our view that the processes set out above do need to be carried forward, and that the current gap between fieldwork ending and the publication of results is the minimum possible without compromising on the high quality of the statistics, which is necessary for the uses to which the statistics are put.

We do however welcome feedback on this document, which can be sent to team.frs@dwpgsi.gov.uk or sent to:

*Julie Sullivan
Professional Services
Information, Governance and Security Directorate
2nd Floor Caxton House
Tothill Street
London
SW1H 9NA*

Annex A: Diagram showing periods covered by various processes described above

