

Initial Review of the FRS Weighting Scheme

Charles Lound & Peter Broad, Methodology Advisory Service, ONS

June 2013

Summary

The Department for Work and Pensions commissioned the Methodology Advisory Service of the Office for National Statistics to carry out an initial review of the weighting system used on the Family Resources Survey. This was to inform the department whether a more in-depth full review of the weighting system was required and to make recommendations for change or for further investigations.

In the review, carried out in the second quarter of 2013, we looked at the overall structure of the weighting system and in particular the design weighting. We looked at the quality of the control totals used in the calibration process, in the context of the main analysis of the survey. We also looked back at the main messages that emerged from the 2001 census-linked study of survey non-response to judge whether the observed patterns of nonresponse were likely to be captured by the existing process. Finally, we looked at the weighting systems used on other major UK and international surveys to see whether the method used on the FRS differs substantially from those.

The overarching conclusion to this review is that we could not find any problems with the existing weighting that suggest that a more substantial review and change to the weighting system be implemented for the 2013/14 survey year. We made a number of recommendations, some of which should be implemented in the course of the next round of weighting and others that should or could be addressed in the longer term.

1. Introduction

The Family Resources Survey (FRS) provides information on the living conditions and resources of people living in the UK. The fieldwork is carried out jointly by the ONS and NatCen in Great Britain and by NISRA in Northern Ireland. The sample in Great Britain is a standard two-stage stratified sample of addresses with Scotland sampled at a higher rate by sampling a disproportionate sample of primary sampling units. The Northern Ireland sample is a single-stage sample of addresses with a higher sampling rate than for the rest of the UK. For those sampled addresses with more than one household, a single household is sampled at the address. The achieved sample was around 25,000 households up to 2010-11, reducing to around 20,000 per year from April 2011.

The weighting process for the FRS, also known as the *grossing regime*, was last reviewed to coincide with the introduction of the 2001 census results. That review reduced the reliance on modelled population estimates which had made use of adjustments to determine a breakdown by *de facto* marital status that were not updated. In addition regional breakdowns were introduced for the first time. Following a comparison of worklessness estimates derived from the FRS and the LFS, the review included a recommendation to consider the future introduction of a control for the number of workless households, though that control has not been implemented.

Now the 2011 Census results have been published and are available for weighting the survey, it is a good time to consider any other changes to the weighting process. The Methodology Advisory Service (MAS) of ONS have been asked to carry out an initial review of the existing FRS weighting to help DWP decide whether to carry out a full weighting review.

This initial review does not include analysis of microdata or revisit issues around implementation, but does include the following:

- Comments on the design weights used
- Investigation of the quality of totals used in calibration process
- Comments on the effect of non-response from the 2001 census –linked study
- Comparison of the survey weighting with other surveys in the UK and elsewhere

2. Weighting strategy for FRS

A previous weighting review took place after the 2001 census and new weights were issued with the 2003/04 FRS data. The survey currently applies two stages of weighting; design weighting and calibration weighting. There is no sampled-based weighting step. Such a step is commonly implemented between the design weighting and calibration weighting steps to address observed patterns of differential nonresponse.

Design weighting

Currently the design weight comprises just a simple multi-household adjustment. If an interviewer finds more than one household at an address, only one household will be interviewed. This means such a household will have a lower probability of selection compared with those at addresses with just one household, and so are given a design weight equal to the number of households at the address. This is the only change made to the design weight; otherwise the design weight is simply 1.

The two-stage sampling addresses from the Postcode Address File (PAF) for the GB sample, with primary sampling units (PSUs) drawn with probability proportional to size and a fixed size sample of addresses from within each sampled PSU was designed to lead to an equal probability sample of addresses:

$$\begin{aligned} P(\text{address}) &= P(\text{address}|\text{PSU}) * P(\text{PSU}) \\ &= (\# \text{addresses per PSU}) / (\text{size of PSU}) * (\# \text{PSUs} * \text{size of PSU} / \text{size of PAF}) \\ &= (\# \text{addresses per PSU}) (\# \text{PSUs}) / \text{size of PAF} \end{aligned}$$

The number of addresses per PSU is constant throughout GB, but the number of PSUs in Scotland is approximately twice what it would be with a constant sampling fraction.

So, in the technical report Evans (2012) says that the probability of selection for an address in England is 1-in-648 chance (1-in-645 for Wales), whereas for Scotland it is 1-in-325. There has been no adjustment for this since introduction in 2004/05 Vekaria (2013). The chance of selection in Northern Ireland is 1-in-210 and there is also no reference to this being accounted for. It is not clear whether the scale adjustment factors for the population totals in the calibration step are linked to this, but this will not account for the different probabilities of selection in the different countries. We suggest that the design weight is

modified to account for these different probabilities of selection as well as the multi-household adjustment. The scale factors in the calibration step will be covered further in the next section.

Calibration weighting

Since the review in the early 2000s, the calibrations totals have been as shown in the following table:

Great Britain	Northern Ireland
Age group by sex by region (13x2x11=286 categories)	Age group by sex (12x2=24 categories)
Benefit Units <ul style="list-style-type: none"> with dependent children, England and Wales with dependent children, Scotland lone parents, male and female 	Benefit Units <ul style="list-style-type: none"> lone parent households
Council tax bands:	
Dwellings by <ul style="list-style-type: none"> Band A & Not valued separately Bands B Bands C and D Bands E to I 	
Tenure:	
Dwellings by <ul style="list-style-type: none"> Local Authority Private and HA renters Owner-occupied 	
Region:	Total households
Households by <ul style="list-style-type: none"> London Scotland Rest of Great Britain 	

A more detailed description of the calibration totals is given in appendix A.

When calibrating our survey weights to population totals, we keep the final weights w_k near to the original weights d_k using a distance function $G_k(w,d) = dG(w/d)/q_k$, where G is a convex, two times differentiable function. The q_k are constants, fixed for each case, which can be chosen to minimise the variance. These are typically given large values for ‘big’ units as they are related to the unit variance, as in ratio estimation. Bounds can be added, effectively selecting the form of the distance function, so that final weights are never allowed to depart too far from the original weights, although with narrow bounds the weighting algorithm cannot be guaranteed to converge.

In the FRS, the weights are formed using information at different levels: individual, benefit unit, household and dwelling. The resulting weights are actually formed at the household level, with the individual-level data first aggregated up to this level. The weights for individual-level analysis are the same as these household-level weights and are therefore the same for all members of the household. This process is important in our assessment of calibration totals because weighting in this way not only controls for the demographic

distribution of individuals, but also exerts some controls on the distributions of different household types, defined by the presence of different numbers of people from different ages and genders.

Currently the q_k weights, defined above, are all effectively set to one. For household level analysis, where measures like total income are related to household size, we may want to consider q_k weights related to the household size. However, this will not be optimal in all cases, for analyses where the outcome values are not related to household size or for individual-level analysis.

In the weighting process, the population totals used in the calibration are scaled down by a factor of 250 for Great Britain and 266 for Northern Ireland and, after calibration, the weights are scaled back up. It is not clear why this process is used. Standard practice is to use original population figures rather than scaled figures and since this is simpler and less prone to confusion, we recommend this standard practice is followed.

The calibration weights are calculated using the package CALMAR from the French National Statistical Institute. However, this package is no longer supported. GES is the standard tool for calibration used by the Office for National Statistics (ONS). Rahman (2009) compares the effect of CALMAR and GES on 2007-2008 FRS data and found negligible difference in the outcomes. Some extra manipulation was required to convert categorical variables into indicator variables for GES. Also the paper made some recommendations on scaling the input weights to population totals prior to calibration, which we understand have not yet been implemented.

Following that work the DWP decided not to transfer the weighting process to GES. This was decided on the basis that DWP wanted to retain the process in-house and that the costs of obtaining and maintaining GES were prohibitive. This represents a risk as Calmar is not well supported, knowledge on how to use it is very limited within the GSS and elsewhere in the UK, and the limited documentation available is largely in French. There is a further risk that the existing version of the macro cannot be guaranteed to work with future versions of SAS. So, while this process is not going to stop working overnight, it would make sense to plan to update in the reasonably near future.

The option to move to GES, with support from ONS, remains open. Otherwise, a recent project carried out by MAS for the Scottish Government investigated an *R* package called *ReGenesees* to weight their social surveys, via a simple interface with SAS (Broad, 2012). *ReGenesees* performed well and there were no material differences in outputs between this weighting package and other packages, so this might be an additional option available to DWP.

3. Quality of Control Totals

In this section, we consider the control totals used in the current calibration process. To help direct this analysis, we begin with a short section on the dimensions of statistical quality and how those apply to control totals used in calibration. We then consider the quality of each set of calibration controls currently in use. We conclude with some thoughts on alternative controls.

Aside from some analysis to check the details of derivation of some of the controls, we have not carried out any statistical analysis here. We suggest that the most revealing analysis that could be done quite quickly would be to compare for each variable in the calibration the design weighted distribution against the population controls. While we do not expect the two to be in perfect agreement prior to calibration – that is the task of calibration – this comparison might reveal some patterns that suggest that the two are inconsistent. For example, we comment below on the possibility that respondents may not report on tenure in a way consistent with the statistics on tenure, and the consequences of any misreporting may be evident from these simple comparison.

Quality dimensions

Relevance

Ideally the variables used in calibration controls are highly correlated with the outcome variables from the survey and associated with different levels of non-response. This will lead to the largest impact in terms of bias and variance reduction. As well as measuring the relevant attributes, this also means that the categories into which counts are divided are sufficiently detailed. For the purpose of estimates for analysis subgroups, or *domains*, these categories should be aligned where possible with those subgroups.

The most direct way to assess the relevance of the calibration controls with respect to variance reduction is to take the calibration into account when calculating standard errors. However, the current standard error calculations, which take into account stratification and multi-stage sampling, and the variance inflation due to weighting, do not capture the beneficial effects of the calibration weighting. This means that we would need to develop these routines to achieve an objective assessment of the impact of calibration.

Accuracy

Calibration totals are usually regarded as fixed quantities, but in practice may include both bias and variance. If the calibration totals are variable, the benefit in terms of increased precision will be reduced, and could even lead to an estimate with a larger variance than a non-calibrated estimate.

Timeliness and Punctuality

The calibration variables have to be available when the weights are required. Where the calibration controls for the survey period are not yet available, this may require a forward projection to be made, which may add to a reduced accuracy due to the prediction being made. This extra inaccuracy can be seen in the magnitude of subsequent revisions as forward projections are replaced by estimates.

If a total used in calibration were delivered later than scheduled, then it may mean that the survey processing has to be delayed or a modification to the weighting model has to be made. Therefore, it makes sense to ensure that totals will be available for future years as part of an ongoing process or agreement rather than provided on an ad hoc basis.

Accessibility and Clarity

As with punctuality, a discontinuation of a total used in the calibration would force a change to the weighting model that could lead to a discontinuity in the survey estimates.

A lack of clarity in the description of the processes used to produce calibration totals and the likely quality of those totals could lead to a less than ideal choice of calibration totals.

Comparability

Comparability of control totals over time is necessary for a consistent weighting model. Strict comparability over domains, such as regions, is less important as long as the weighting model is effective in all such domains.

Coherence

Where auxiliary data are collected in the survey process and used in calibration weighting, these must be consistent with those used in the production of the weighting totals. It is also important that the patterns in the calibration totals are consistent with the patterns in the survey data or the weights can be volatile or the calibration may not converge to a solution. Except where the inconsistency is obvious in the totals (e.g. two classifications summing to a different totals) such inconsistency can only be tested by running the weighting process.

Key outcome variables and subgroups used in FRS analysis

As stated under *relevance* above, for calibration to help with the precision of survey estimates, we need an association between the survey's outcome variables and the calibration variables. For estimates for subgroups, or domains, the calibration process has most benefit in terms of precision when the classifications into which the calibration totals are subdivided are aligned with these analysis domains.

There were fifty tables in the substantive chapters of the [FRS 2010/11 report](#) (excluding the methodology chapter)¹. The following tables summarise the outcome variables and domains used in the report's tables. In some cases it is not clear from the table description which is the outcome variable and which the domain, but we have been guided by the way the percentages are derived, taking the outcome variable to be the one for which the percentage add to 100%. Where a domain is cross classified, e.g. gender by age group, we have included both here.

¹ The report for the following year 2011/12, published after this work was completed, had two fewer chapters and so fewer tables. However, this distribution of variables remains a good indication of the most important.

Outcome variable	Tables
Types of state support received	2.7, 2.8, 2.9, 2.10, 2.11, 3.5, 5.4, 7.7
Type of savings and investments	4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8
Amount of savings and investments	4.9, 4.10, 4.11, 4.12
Employment status	5.1, 6.3, 7.1, 7.3
Pension participation	8.1, 8.2, 8.3, 8.4
Sources of income	2.1, 2.2, 2.3
Total weekly household income	2.4, 2.5, 2.6
Tenure	3.1, 3.2, 3.3
Person cared for	6.2, 6.7
Main source of household income	6.4, 6.6
Length of residency	3.4
Impairment type	5.2
Disability status	5.3
Number of hours providing care	6.1
Frequency of care received	6.5
Time since last in paid employment	7.2
Childcare costs	7.4
Percentage of total income	7.5
Economic status	7.6

Domain variable	Tables
Gender	4.7, 5.1, 5.2, 6.1, 6.2, 6.5, 6.6, 6.7, 7.1, 7.2, 7.3, 8.1
Economic status	4.6, 4.11, 7.4, 7.5, 7.7, 8.1, 8.2, 8.3, 8.4
Region and country	2.1, 2.5, 2.7, 3.1, 4.1, 8.4
Age group	4.7, 6.1, 6.5, 7.1, 7.2
Ethnic group	2.2, 2.6, 2.9, 3.3, 4.4
Age of head of household	2.3, 2.10, 3.4, 4.3
Household composition	2.4, 3.2, 4.2, 4.9
Family type	2.8, 4.5, 4.10, 5.3
Tenure	2.11, 3.4, 3.5
Total weekly household income	4.8, 4.12, 8.3
Working age/State Pension age (disabled)	5.1, 5.2
Carer status	6.2, 6.6
Disability status	5.3
Number of hours providing care	6.4
Length of time care provided	6.7
Standard Occupational status	7.3
Age of youngest child	7.6

Quality of existing calibration controls

In this section we review each set of calibration controls used in the existing weighting model, including a description of the sources and comments on quality using the framework described earlier.

Age group by sex by region

These are derived from the mid-year population estimates from ONS.

As gender, age group and region appear in the domain specification for many of the tables in the FRS report, there appears to be a very strong case for including these in the calibration. Similarly, we can see, for example, that the sources of income vary by age group (table 2.3) and distribution of total household income varies by region (2.5).

Population estimates are usually regarded as fixed, being based on census statistics amended by birth and deaths. However, the census data are themselves based partially on the census coverage survey and are therefore subject to sampling error. In addition, immigration is provided by information from the International Passenger Survey and is again subject to sampling error.

Mid-year estimates are produced one year after the reference date which means that the mid-year estimates are available in time to weight the survey.

Surveys such as the LFS make similar use of population counts in their calibration. However, most surveys make an adjustment to deduct the communal establishment population using region by sex by age group proportions from the most recent census.

Benefit units with dependent children

These data are provided as counts of families receiving child benefit, available separately by region. The data provided also give counts of families where one child is under one and counts of children broken down by age, including those under one, suggesting an adjustment is made to the counts for the youngest children.

The Child Benefit data are generally regarded as being accurate since the take-up rate for recent years² is around 96%. However, an error of 4% may be large in this context. Furthermore, there is a clear emerging risk here as the benefit has become limited to those earning below a threshold. The timing and accuracy of reporting of salary levels in the survey, together with the issue of cross-referencing both parents' salaries and the fact that there is some choice on opting out or paying back for those close to the threshold means that it would not be reliable to try and identify qualified recipients in the data.

Lone parents

Estimates of male and female lone parents are provided by direct analysis of the Labour Force Survey. These are derived from the LFS household file and limited to private households in Great Britain. The family types are specifically lone parent males and females with dependent children with the lone parent aged between 16 and 64, inclusive.

Since these estimates are derived from a survey, the sampling variance from that survey will reduce the benefit of calibration, compared with fixed totals.

We understand that the totals used here are not part of a standard publication. However, they are produced in DWP using a stable source, so it is reasonable to assume these data will remain available.

² <http://www.hmrc.gov.uk/statistics/fin-takeup-stats/cwtc-take-up.pdf>

Council Tax bands

Council Tax band does not feature directly in the analyses in the report, but it is reasonable to assume that measures like household income and savings and assets are associated with the value of the dwelling.

The counts are in four categories, covering: Council Tax Band A and Not valued separately; Band B, Bands C and D; and remaining Bands E to I.

The counts for England and Wales are taken from a [statistical release](#) from the Valuation Office Agency. The data include counts for regions and constituencies as well as national totals and are derived by tabulating data held in an administrative database.

Counts for Scotland are [published](#) by the Scottish Government. These are compiled from annual CTAXBASE returns of counts from local authorities.

It is notable that the dwelling counts are not limited to occupied dwellings. We recommend that evidence is sought from the housing surveys to establish whether occupancy rates are reasonably similar across Bands.

Those dwellings that are ‘not valued separately’ are not included in the administrative counts. The method for including these in the calibration is described in Vekaria (2013) as being ‘based on the FRS sample which varies quite widely from year to year’. It is not clear whether this is necessary – the counts do not need to cover all the sample (for example Northern Ireland is excluded from this) so it may be possible to implement this part of the calibration without the ‘not valued separately’ category. In fact, if the size of this group is derived from the current year’s sample then this omission of this category might result in the same weights.

One concern here is the reliability of people’s reports of their Council Tax Band. Since many will pay by direct debit they will have no need to consult the bill. Again, a simple comparison of design-weighted and population distributions should reveal any major distortions. In the longer term, linking the Council Tax Band to the address from administrative sources would improve consistency in recording.

Tenure

These totals of dwellings by tenure, from CLG (for England, Wales and Scotland), are divided into LA renters; private and Housing Association renters and owner occupiers.

The tables used (Tables 104, 106 and 107 on <https://www.gov.uk/government/statistical-data-sets/live-tables-on-dwelling-stock-including-vacants>) actually divide dwellings into:

- England: Owner-occupied; renting privately or with a job or business; rented from private registered providers; rented from local authorities; other public sector dwellings
- Wales and Scotland: Owner-occupied; renting privately or with a job or business; rented from Housing Associations; rented from local authorities; other public sector dwellings

However, it is clear from the labelling of the chart for England that ‘rented from private registered providers’ is synonymous with ‘rented from Housing Associations’

For England, the methods used to produce the counts are shown in the [Dwelling Stock Estimates](#) publication. Counts of dwellings are taken from the census and rolled forward using information on net annual changes to the housing stock. Information on Local Authority and other public sector dwellings is provided by returns from Local Authorities. The counts of dwellings provided by private registered providers are also collected by statistical returns from these providers.

The split between owner-occupied and private rented is determined using the Labour Force Survey and English Housing Survey (from 2003). This is done by estimating the number of private-rented dwellings and calculating the owner-occupied by subtraction. The number of occupied private-rented dwellings is calculated by smoothing LFS estimates of totals over two years (at the end of the series) or three years. This is then divided by an occupancy rate for this sector estimated from the English Housing Survey.

For Wales, the methods used to produce the counts are shown in the [Dwelling Stock Estimates First Release](#). The methods for counting Local Authority dwellings and Housing Association dwellings are similar to those used in England. For Wales, the division between private rented and owner-occupied is made by estimating the proportion of private renters from the Annual Population Survey and no adjustment is made for vacant dwellings.

The methods for Scotland are described in the [Housing Statistics for Scotland](#) tables. Estimates of vacant LA and HA stock are deducted from a total vacant estimates (source unknown) and the remainder providing an estimate of the vacant private stock which can be subtracted from the total private stock to give a total occupied private stock. This is then divided into private rented and owner occupied using proportions from the Scottish Housing Survey. In the tables on the CLG site, used to supply the calibration totals, the vacant dwelling counts are included in each tenure’s count.

Number of households by region

As well as calibrating the weights to the regional distribution of individuals, the FRS also employs a regional breakdown of the number of households, but into just three regions in Great Britain: London, Scotland and the rest of Great Britain.

In addition, the weighting for Northern Ireland includes a household total. The application of this total is different as NI is effectively weighted separately from the rest of the UK.

There is a clear case from the estimates shown earlier for including a regional household breakdown in the weighting as a regional analysis of household outcomes is included in five of the published tables and the survey estimates can be seen to vary by region.

These household totals come from the [CLG household projections](#), by country in Table 401 and for London in Table 406. The [methodology for producing the totals](#) varies by country and the base year varies according to data availability. They are all based on taking population projections and subtracting the communal establishment population, using fixed numbers or proportions taken from the census.

In England, to predict the number of households, the population data are combined with *household representative rates* which are the probabilities of people being the household representative, by age, sex and marital status. The number of households is estimated by multiplying these rates by the cell size and summing across all cells. The *household representative rates* are projected into the future using a trend model using data from censuses and the Labour Force Survey, with extensive smoothing to reduce sampling variation.

In Wales, the *household membership rate* for different household types is modelled by age and gender. These are then projected forward using a two-point exponential model. The resulting population projects by household type are divided by household size to give the predicted number of households.

In Scotland, the *headship rate* is modelled using a two-point exponential model fitted to 1991 and 2001 census data, producing separate estimates by age, local authority and household type. These rates are then aggregated across all individuals to estimate the number of households. The aggregate number of households is then constrained to a total estimated from council tax data when this is available.

In Northern Ireland, the *household membership rate*, which is the probability of being in each household type for each age group and sex, is calculated from the 1991 and 2001 censuses. These rates are then modelled for other years using a two-point exponential model. The rates are applied to the population to produce the distributions of people by household type and the results divided by the type-specific household size to produce an estimate of the number of households.

Overall, we can see that the methods rely largely on census data, with little use of survey data and then using the largest available source, the LFS, and smoothing over years.

It is unusual to see region applied only at a broad level, especially as the analysis table report down to individual regions. It would be possible in principle to introduce the detailed regional breakdown as well. There is a risk, when also using the regional breakdown for individuals that the resulting calibration would not converge or would produce much more variable weights, but since the household projections are based on the population projections, it seems likely that the two would be consistent together with the FRS data.

Characteristics not captured by the existing calibration controls

Under *relevance* in the earlier discussion of the quality of population controls for calibration, we stated that controls should be correlated by the outcome variables, different levels of response or define key domains. We have commented that the existing controls address this requirement, but are there other potential controls that could be introduced?

Some of the outcome variables are based on FRS questions where there is little available from population data, including those around sources and level of income, savings and investments, pensions, disability, and caring.

We noted that tenure as an outcome variable is very well addressed by the tenure control and that may be related to length of residency. Childcare costs are likely to be related to presence of children and whether a lone parent.

Some other outcomes like employment status, economic status might be covered by the inclusion of further controls derived from a larger survey, notably the LFS. However, some development work would be needed to check whether these measures are equivalent on the two surveys and whether the variance in the control totals is sufficiently small to improve the accuracy of the FRS estimates.

In principle, we could introduce controls on the types of state support received, using counts from DWP's administrative systems as controls. However, there is concern that measurement errors from people misreporting types of benefits could introduce errors here. Such calibration might be better achieved through data linkage to the address to ensure that the auxiliary data for the sample and population are measured consistently.

Looking at the domain variables used in the analysis tables, we see region and country, tenure, age group and gender are covered by existing calibration controls. Classifications based on the composition of families and households, including the age of head of household and age of youngest child are indirectly covered by the way households weights are influenced by the age and sex of everyone in the household and more directly by the dependent children and lone parent controls.

The remaining domains include economic status, which was discussed above. There is no explicit control for ethnic group. This could be included by bringing in controls from other surveys. For other domains variables like total weekly household income, disability status and caring are not directly addressed by the calibration controls, but are likely to have some association with controls such as age group and sex.

In addition to attempting to choose calibration controls that explain variation in outcomes and define domains, the calibration controls should also be related to patterns of non-response, particularly as the current weighting regime does not include a sample-based non-response weighting step. This is discussed later in section 5.

4. Other Surveys

Another important aspect in this initial review is to compare the weighting strategy in the FRS with other surveys, both from the UK and internationally. In the following sections, the weighting strategies for the Labour Force Survey (LFS), the Living Costs and Food survey (LCF) and the Wealth And Assets survey (WAS) from the UK will be outlined as well as a brief overview of the weighting strategies from some similar household income and living condition surveys from Australia, USA and Canada.

4a UK surveys

Labour Force Survey (LFS)

The Labour Force Survey (LFS) is a survey of households living at private addresses in the UK to provide information on the UK labour market. This information is used to develop, evaluate and report on labour market policies. It is the largest regular social survey in the UK. The sample in August 2011 consisted of around 40,000 responding households in Great Britain every quarter with around 1,600 households in Northern Ireland. This sample is split across 5 waves.

The weighting strategy for LFS focuses mainly on calibration, but does also have a design weight. The design weights are typically constant as the LFS sample design ensures an equal probability of selection. However there are three exceptions to this:

- A different sampling fraction is used in Northern Ireland, so therefore a different design weight is also used.
- Households comprising only those aged 75 and over are interviewed only in Wave 1; the design weight is adjusted for this.
- Only one household is sampled at addresses where there is more than one household, so the design weight is adjusted for this in a similar way to FRS.

The LFS does not include a sampled-based non-response step, so all adjustments for differential non-response is achieved through the calibration process. Three categorical variables are used in the LFS calibration, each partitioning the population in a different way:

- 433 Individual Local Authority Districts
- Cross classification of age by sex by Great Britain or Northern Ireland for ages 0-15, individually for years 16-24, and 25+. This results in 12*2*2 calibration groups, 12 age bands, 2 sexes and 2 countries.
- Sex by region by five-year age bands up to 80+. This gives 612 groups, 18 regions, 17 age bands and 2 sexes.

Overall, across all three variables, there are 1,089 calibration groups. The calibration is carried out using GES software from Statistics Canada.

The FRS makes a similar use of population data, but with far less detail. However, the FRS does draw in other data on households and dwellings and has a smaller sample size, so the use of less detail here is probably justified.

Living Costs and Food survey (LCF)

The Living Costs and Food Survey (LCF) collects information on spending patterns and the cost of living that reflects household budgets across the country. The primary uses are for spending patterns for the Consumer Price Indices, but also for food consumption and nutrition. In 2010, 5,116 households co-operated, resulting in a 49.6% response rate.

LCF is weighted in two stages, first for non-response based on factors created from the 2001 Census and second so it matches the population distribution by region age group and sex. The non-response factor reflects non-response classes based on the number of pensioners, region, number of cars and type of household. The population-based weighting uses 20 age & sex categories by the 12 regions. The age categories for males and females are different for over those aged over 30, with one more category for females (Hossack & Jarvis 2010).

Although the design is very similar to FRS, there is no design weight as there is currently no multi-household adjustment as with FRS and the sampling rate is the same across the sample.

Wealth and Assets Survey (WAS)

The Wealth and Assets Survey is a longitudinal household survey which gathers information on the economic well-being of households, including levels of savings and debt, saving for

retirement, how wealth is distributed among households and factors that affect financial planning. The WAS first commenced in July 2006 with a first wave of interviews at 30,595 households and 20,170 at the second wave (Black 2011).

Since the survey is longitudinal, there are two sets of waves: (i) a longitudinal and (ii) a pseudo cross-sectional weight. All weights first use the reciprocal of the selection probability, are adjusted for non-response (through attrition) and finally calibrated to population totals. In the calibration the 11 regions (excluding Northern Ireland) are used and there are 12 age categories for each sex, which unlike LCF, are the same for males and females.

The weights are used to account for the longitudinal aspect of WAS, which isn't necessary for FRS. This is an example of another survey which just uses age, sex and region as calibration totals for the cross-sectional first wave.

Understanding Society – the UK Household Longitudinal Study (UKHLS)

This is a longitudinal survey starting with approximately 40,000 achieved household interviews in the UK. Households are recruited at the first round of data collection and visited one year later to collect information on changes to their household and individual circumstances. Funding comes from the Economic and Social Research Council and from a number of government departments, including DWP. The overall purpose is to provide data on the long term effects of social and economic change on subjects such as health, work, education, income, family and social life.

The weighting strategy for the UKHLS is quite complicated as there are a number of survey instruments are used and the survey is longitudinal. The design weights adjust for a higher sampling fraction in Northern Ireland, for an ethnic minority boost sample and for dwellings with more than three households. There is also a non-response adjustment based on propensity to respond in wave 1 and subsequent waves. A post-stratification adjustment uses age, sex and region to meet ONS mid-year population estimates.

As with WAS, the longitudinal aspect of UKHLS is not applicable to FRS, but there is a non-response adjustment.

4b International surveys

Appendix B1 in the appendix gives a summary of some international living conditions and resource surveys to compare with the FRS weighting scheme. All involve some sort of non-response weight except the Survey of Income & Housing and Household Expenditure Survey from Australia, however there is evidence of using larger surveys as a source for calibration totals here. Calibration totals which are not age, sex or region are also more widely used when compared to UK surveys. For example Stats Canada uses wage and salary admin data and the U.S. Census bureau uses ethnicity from the Census in calibration. The FRS could ask respondents if they can link to HMRC wage and salary information. Around 80% give consent for this in Canada. Work from the Beyond 2011 project may result in this information being more freely available.

5. Nonresponse

The FRS was not included in the 2011 census-linked studies of survey non-response, but was included in 2001. The results from the study are available in a report on the FRS

website ([Freeth and Sowman, 2005](#)). The analysis for other surveys included in the 2011 study continues.

The census-linked study included univariate analyses of non-contact and refusal and logistic regression analyses of non-contact, refusal and total non-response. For a brief overview of the patterns of non-response, we can look at the results of the logistic regression analysis as this identifies those characteristics most strongly, and independently, associated with non-response. Here we show the variables listed in the report and indicate from the tables the categories with the lowest levels of response.

Non-contact significantly associated with:	Categories with high probability of non-contact
Area type (these are types of local authorities)	Metropolitan districts
Type of building occupied by the household,	Purpose built flat, converted house
Number of people in the household,	Smaller households
Age of the Household Reference Person	Lower age groups
Country of birth of the Household Reference Person.	Non-UK
Refusal significantly associated with:	Categories with a high probability of refusal
Government Office Region	North west, East Midlands, West Midlands, Eastern, London, South East
Housing tenure	Owned outright, buying with a mortgage
Age of the youngest dependent child	No children or dependent child aged 5-15
Highest qualification of the Household Reference Person or	Below degree level
Ethnic group of the Household Reference Person.	White
Total non-response significantly associated with:	Categories with a high probability of total non-response
Area type,	London
Housing tenure	Owned outright, buying with a mortgage
Age of the youngest dependent child	No children or dependent child aged 5-15
Marital status of the Household Reference Person,	Single (never married)
Highest qualification of the Household Reference Person	Below degree level
Economic activity of the Household Reference Person.	Self-employed

The patterns of non-contact and refusal are informative in showing how the total nonresponse is accumulated. If sufficient data were available for the set sample and contacted sample, separate weighting steps could be established for each of these processes. The total non-response shows us the net effect of these two steps in the response process.

If we look through the variables associated with non-response, we can see that the FRS already has weighting factors in place to cover variations in response by region (and London in particular) and by tenure. There is a calibration variable directly addressing the presence,

and this and the age of children present is also controlled for by the household-level calibration based on individual characteristics.

The marital status of the HRP is not directly controlled for by the existing calibration weighting. However, we note that the number of people in the household, which may be associated with the marital status of the HRP, is significantly associated with non-contact, and the number of people present will influence the weights through the household-level weighting.

The final two variables associated with total non-response: the qualification level and economic activity of the HRP do not appear to be captured by the current calibration controls. There is limited information available for the set sample that could be used here. We can think of two potential options for non-response weighting.

The first option is to make use of administrative data that can be attached at address level. This could either be to the set sample or the whole sampling frame. Indicators associated with economic status could include information about receipt of work-related benefits; information about self-assessment tax returns; or information about National Insurance status and payments. Any such approach would of course require access to such data and secure transit and storage. Also, care would need to be taken to ensure that the resulting weights do not inadvertently contribute to the identification of households within the data.

An alternative would be to create weights that attempt to capture these dimensions through association with the attributes of a local area. This could be introduced as a sampled-based non-response step between the design weighting and calibration step. One possibility is making use of the National Statistics Area classifications. These are currently being updated to take on the 2011 census data, but the variables used to form the clusters in 2001 included:

- *Demographic*: Age, Ethnicity, Country of Birth, Population density, Household Composition, Living Arrangements, Size/Family
- *Housing*: Tenure, Type and size, Quality/crowding
- *Socio-Economic*: Education, Socio-economic class, Ownership/commuting, Health and Care, Employment, Industry Sector

Note in particular that this included education level and employment status.

Such a sample-based nonresponse weighting step would take some time and development resource to evaluate in advance of implementation, so we consider that this may be a longer term objective.

6. Conclusion and recommendations

The overarching conclusion to this initial review is that we could not find any problems with the existing weighting that suggest that a more substantial review and change to the weighting system be implemented for the 2013/14 survey year.

We have made recommendations for changes that can be implemented for this year and for further investigations into particular aspects of the weighting system. We have prioritised these into the high priority recommendations that should be implemented immediately; medium priority recommendations that cover emerging issues or suggest improvements might be available and lower priority recommendations that are desirable to address in the longer term.

High priority

1. *Ensure that the design weighting includes the different rates of sampling in the England and Wales, Scotland and Northern Ireland, as well as the one-household-per-dwelling factor.*

This is the one recommendation that should be thought of as mandatory to complete before the next round of the survey is analysed, because the over-sampling of Scotland relative to the rest of Great Britain is part of the sample design. This can be checked by examining the weights from the previous year prior to calibration, looking at the mean weights for England and Wales, for Scotland and for Northern Ireland.

2. *Work with population totals in their original values, rather than artificially scaled.*

While this may not be leading to a quality deficit, using the unusual units of 250 population makes quality assurance less easy as the weights will not add to recognisable population totals. This is a simple change to make.

3. *Implement an initial scaling of the design weights to population totals prior to calibration.*

Again, a simple amendment to make, scaling the weights separately for England and Wales, Scotland and Northern Ireland by the ratio of the population total to the sum of the weights prior to calibration.

4. *Consider reporting quality on Council Tax band.*

In our discussion, we speculated that respondents would find it difficult to report on the council tax band for their property. A short-term approach to considering this is to check whether respondents are asked to consult the Council Tax Bill and whether the proportion of cases for which this happens is recorded.

5. *Consider re-categorising the tenure breakdown to avoid uncertainty around LA/HA split.*

In terms of process, there seems a good argument not to split the calibration totals between these sub-groups as tenants will often be unsure as to their landlords' category. As with council band, an initial check on whether documents are requested to be

consulted and whether they are consulted will inform this decision, as will the observed distributions before and after calibration.

6. *Produce design weighted distributions for each existing calibration variable for comparison with the population totals.*

Here we recommend that the weighted frequency distribution is produced for each calibration variable using the weight that is fed into the calibration process and the weight produced by calibration. These may reveal some distortions that are brought about by inconsistent recording of the calibration classes in the interview and control totals.

Medium priority

7. *Consider whether including vacant dwellings in the Council Tax calibration totals is likely to cause a distortion in the weights.*

We have recommended that we seek evidence for different occupancy rates by council tax band from the housing surveys. This investigation is also informed by a comparison of weighted distributions before and after calibration.

8. *Consider whether to exclude vacant dwellings from the tenure breakdowns.*

This decision is complicated by the fact that the current calibration uses fixed dwelling counts which are sub-divided using survey data. We could move over to an entirely survey-based source, but this would introduce more variation into the calibration controls.

9. *Consider options for replacing the controls for benefit units with dependent children for when the estimates are affected by the child benefit eligibility changes.*

The Child Benefit data is used indirectly to produce a count of benefit units with dependent children. Can another source be used to estimate this, or can this be replaced by a proxy measure that has a similar impact in terms of calibration.

10. *Calculate standard errors for the lone parent controls to help assess their impact on variance.*

This involves extending the existing analysis that replicated the lone parent counts to produce standard errors for these counts and then to assess the approximate impact of that variance on the variance of the survey estimates.

11. *Consider introducing more detail into the regional breakdown of households.*

As the calibration uses a breakdown of people by region, we consider this as only medium priority. It also requires re-running the calibration process with and without this set of controls.

Low priority

12. *Plan to modify existing standard error calculations to capture the possibly beneficial impact of calibration weighting.*

This involves verifying exactly what is being produced by the programs developed by ONS and, if necessary, making a recommendation for changing those programs. This is not crucial for the completion of the weighting review.

13. Consider including a variance weight (q_k) in the calibration calculation.

This is really a placeholder for something that would be desirable to investigate to improve the precision from calibration, but is not seen as a substantial technical deficiency.

14. Consider plans to transfer calibration to GES (short-term contingency) or, for example R (medium-term development) taking opportunities to benefit from other work in R within the DWP or elsewhere.

Again this is a placeholder to remark that this change could reduce risk, but would not introduce statistical improvements.

15. Consider longer term introduction of a sample-based non-response step to capture variations in response by socio-economic characteristics, possibly using linked administrative data or the output-area classification.

We identified from the 2001 study of census linked that there are dimensions relating to non-response but not represented in the calibration variables. These may be addressed using other variables available for the whole set sample.

7. References

- Black, O. "Wealth in Great Britain – Main Results from the Wealth and Assets Survey: 2008/10", <http://www.ons.gov.uk/ons/rel/was/wealth-in-great-britain-wave-2/2008-2010--part-1-/index.html>, December 2011
- Bosak, J., "Grossing factors for National Statistics from DWP Family Resources Survey", http://research.dwp.gov.uk/asd/frs/reports/new_grossing_regime.pdf, February 2005
- Broad, P., "Centralised Weighting Project for Scottish Government", *Available on request*, October 2012
- Evans, D., "Family Resources Survey", https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/222839/frs_2010_11_report.pdf, June 2012
- Hossack, P. & Jarvis, E. "Living Costs and Food Survey – Technical Report for survey year: January – December 2010", <http://www.ons.gov.uk/ons/guide-method/surveys/respondents/household/living-costs-and-food-survey/living-costs-and-food-survey-technical-report-for-survey-year-january-december-2010.pdf>, January 2012
- LFS team, "Labour Force Survey – User Guide: Volume 1 – LFS Background and Methodology 2011", <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/labour-market-statistics/volume-1---2011.pdf>, August 2012
- Lynn P, Kaminska, O (2010). "Weighting strategy for Understanding Society" <https://www.understandingsociety.ac.uk/research/publications/working-paper/understanding-society/2010-05.pdf>
- Rahman, J., "Comparing weights from CALMAR and GES on 2007-2008 Family Resources Survey data", Methodology Consultancy Report for DWP, April 2009
- Vekaria, R., "Issues to consider for a grossing review", *Internal DWP document*, February 2013

Appendix A: Detailed population controls for the FRS

GREAT BRITAIN:

AGE BY SEX BY

Male	0-9	Female	0-9
Male	10-19 (dependants)	Female	10-19 (dependants)
Male	16-24 (non-dependants)	Female	16-24 (non-dependants)
Male	25-29	Female	25-29
Male	30-34	Female	30-34
Male	35-39	Female	35-39
Male	40-44	Female	40-44
Male	45-49	Female	45-49
Male	50-59	Female	50-59
Male	60-64	Female	60-69
Male	65-74	Female	70-74
Male	75-79	Female	75-79
Male	80+	Female	80+

So there are a total of 286 Age by Sex by Region controls (13×2×11).

Source: ONS

BENEFIT UNITS

- With dependant children in Scotland
- With dependant children in England and Wales
- Male lone parents (with dependant children)
- Female lone parents (with dependant children)

Source a), b), c) & d): DWP estimates using data derived from ONS & HMRC

Source c) & d): LFS estimates

NORTHERN IRELAND:

AGE BY SEX BY

REGION

North East
North West & Merseyside
Yorkshire & Humberside
East Midlands
West Midlands
South East
South West
Eastern
London
Wales
Scotland

COUNCIL TAX BANDS

Band A and not valued separately

Band B

Band C and D

Band E to H (and I for Wales)

Source: Valuation Office, Scottish Government

TENURE

Local Authority renters

Private renters (including Housing Association)

Owners

Source: DCLG

REGION

London

Scotland

All nine remaining areas

Source: DCLG

Male	0-19 (dependants)	Female	0-19 (dependants)
Male	16-24 (non-dependants)	Female	16-24 (non-dependants)
Male	25-29	Female	25-29
Male	30-34	Female	30-34
Male	35-39	Female	35-39
Male	40-44	Female	40-44
Male	45-49	Female	45-49
Male	50-59	Female	50-59
Male	60-64	Female	60-64
Male	65-74	Female	65-74
Male	75-79	Female	75-79
Male	80+	Female	80+

So there are a total of 24 Age by Sex by Region controls (12×2).

Source: NISRA

BENEFIT UNITS

Lone parents

Source: DSDNI

HOUSEHOLDS

Source: DSDNI

Appendix B: Details on weighting in some international surveys

Table B1: Table of some International Living Conditions and Resource surveys to compare with FRS weighting scheme.

Country	Survey	Weighting Steps	Calibration totals & source	Source	Other aspects	Useful for FRS
Canada (Stats Canada)	Survey of Labour and Income Dynamics (SLID)	- Design weight - Non-response weight - Calibration	- Provincial level age/sex groups and household and family sizes (Census) - Wages & salaries (Derived from Canada Revenue Agency (CRA) admin data)	http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3889&lang=en&db=imdb&adm=8&dis=2#b7 (22nd April 2013)	- Cross sectional aspect for static measures - Longitudinal for transitions, durations and repeat occurrences - Selected from monthly LFS sample - 17,000 households a year	-Use of admin data from HMRC on wages & salaries -Ask permission to use their tax information (80% give consent)
Canada (Stats Canada)	Survey of Household Spending (SHS)	-Design weight -Non-response weight -Calibration	-Age (Census) - Number of households per size (Census) - Remuneration paid (CRA admin data)	http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3508&lang=en&db=imdb&adm=8&dis=2#a2 (22 nd April 2013)	-stratified, multi-stage sampling - 18,000 households a year	-Use of admin data from HMRC on wages & salaries -Ask permission to use their tax information (80% give consent) - If permission isn't given, they are just asked in the survey
Australia (ABS)	Survey of Income & Housing (SIH) and Household Expenditure Survey (HES)	-Design weight -Calibration	-State or territory by age/sex (Census) -State or territory by employment status (Census) - Capital city/balance of state -Composition of household, # of adults and whether children	http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/1D03ACBBD40275ADC A257A3900173E85/\$File/65030_1.pdf (22nd April 2013)	- 18,000 households a year - Of the 18,000 around 9,000 also interviewed for Household Expenditure Survey (HES)	-Has used estimates of one survey as benchmarks for another – scope here for some from LFS?

			<p>present</p> <ul style="list-style-type: none"> - Value of Government benefit cash transfers (Source unclear) <p>HES also calibrated to SIH estimates of:</p> <ul style="list-style-type: none"> -weekly household income -households by tenure 			
U.S.A (Census Bureau)	Survey of Income and Programme Participation (SIPP)	<ul style="list-style-type: none"> -Design weight -Nonresponse weight -Calibration 	<ul style="list-style-type: none"> -age -race -sex -Hispanic origin -family relationship -household type (source unclear) 	http://www.census.gov/sipp/weights.html (22nd April 2013)	<ul style="list-style-type: none"> -Rotating panel design -Non-response and attrition weight including the following characteristics: Census region, race, tenure, household size, poverty 	<ul style="list-style-type: none"> -Non-response adjustment - New calibration variables used, however these are not available for FRS.
U.S.A (Census Bureau)	American Community Survey (ACS)	<ul style="list-style-type: none"> -Design weight - Nonresponse weight -Calibration 	<ul style="list-style-type: none"> -subcounty areas (Census) -marital status (Census) -race/Hispanic origin by sex by age (Census) 	http://www.census.gov/acs/www/Downloads/survey_methodology/Chapter_11_RevisedDec2010.pdf (22nd April 2013)	<ul style="list-style-type: none"> -Large number of variables used in calibration -Appears that simple nonresponse factors are applied 	<ul style="list-style-type: none"> - Non-response adjustment - New calibration variables used, however these are not available for FRS.